Seminar

Data Properties as Economic Goods and Data Markets

Yuri Demchenko

22 May 2019, Kiev

Outline

- Background
- Overview of existing research, activities and available resources on defining economic value of data and future data markets
  - Research and philosophy
  - FAIR data principles and their adoption in research and industry
  - Data Factories and other concepts to enable Data Market
  - Industrial Data Spaces Architecture and Data Sovereignty
- STREAM properties of data as economic goods
- Technology for Data Markets overview
  - Cloud Computing and Virtual Private Cloud (VPC)
  - Blockchain and data monetisation (by Datapace)
- Open Data Market Architecture and component

H2020 call ICT-13-2018-2019: Supporting the emergence of data markets and the data economy

- **Innovation Actions** for setting up and operating platforms for secure and controlled sharing of "closed data" (proprietary and/or personal data).
  - The actions should address the **necessary technical, organisational, legal and commercial aspects** of data sharing/brokerage/trading, and
  - Build on existing computing platforms.
- The lack of trusted and secure platforms and privacy-aware analytics methods for secure sharing of personal data and proprietary/commercial/ industrial data hampers the creation of a data market and data economy by limiting data sharing mostly to open data.
- ***Industrial data platforms*** shall enable and facilitate trusted and secure sharing and trading of proprietary/commercial data assets
  - with automated and robust controls on compliance (including automated contracting) of legal rights and fair remuneration of data owners.
- Strong facilitator of research and innovation on priority topics in Europe

Gaps: Data is becoming an economic goods but no facility to unleash their full market potential

- Data use as Economical model
  - Data use and re-use
  - Data localization
  - Vendor lock-in (apps and data)
  - Legal uncertainty (non-personal data, cross-border, GDPR, provenance)
- Data property as economic goods is not researched and not defined
  - **Data is more than oil of the future economy --- Water?**
- There is no common vision and model how to trade data while retain data ownership (and sovereignty)
  - **GDPR** provided common rules but there is not clear technology alignment
  - New **ePrivacy legislation** will make data management rules even stricter
  - The new Data Market model needs to be developed and adopted
- Use of modern Cloud Computing and Big Data technologies and infrastructure is inevitable
  - There is not well developed security and trust model for storing and processing sensitive/proprietary data on cloud
- There is no (or limited) coordination between industry and academia/research to develop new market model and mechanisms

Research and Technology Overview

- On the way to define data properties as economic goods
  - FAIR principles and STREAM data properties
  - Information value research, economics of “superstars”
- Modern Data Market architecture and components
  - Cloud and Big Data technologies enabled
- Industrial Data Space (IDS) Architecture and activities
- Data related standards as basis for Data Exchange design
  - RDA outputs: PID, Data Factories, Data Type Registries, Repository certification, etc.
  - NIST, OASIS, BDVA, IIC in US
  - Industry best practices on Data Management and Governance

Data Properties as Economic Goods

STREAM data principles for industrial and commoditised data

- [S] Sovereign
- [T] Trusted
- [R] Reusable
- [E] Exchangeable
- [A] Actionable
- [M] Measurable

Leverages FAIR principles for research data
Findable – Accessible – Interoperable - Reusable

- Other data properties: Important ***to commoditise*** data
  - **Quality, Value, Auditability/Trackability, Branding, Authenticity**, as well as original FAI(R) properties Findability, Accessibility, Interoperability.
  - **Data ownership and IPR**: Special features that must be managed in all data transfer and tracked along all data transformation.
  - **Not-Rivalry:** data is not depleted because of sharing and exchange

- Data as Economic Goods: Definitions, Properties, Challenges, Enabling Technologies for Future Data Markets, by Yuri Demchenko, Wouter Los, Cees de Laat. Position paper. To appear in ITU Journal: ICT Discoveries, Special Issue “Data for Goods”, 23 November 2018
  https://www.itu.int/en/journal/002/Documents/ITU2018-12.pdf
  https://www.itu.int/en/journal/002/Pages/default.aspx

Data Markets and IoT

- IoT is considered as a key use case and a facilitator for Data Markets
  - Potentially many consumers for centrally or locally operated IoT infrastructure
- IoT (sensors) infrastructure often created by community/federal project and will produce open or community value added services
  - IoT data can be exchanged and traded
  - **Sensor networks are core of many Research Infrastructures**
- IoT is supported and powered by Edge/Fog computing infrastructure
  - Special cloud services by major cloud providers
- Numerous EU and industry studies and roadmap for IoT Data

Data Market initiatives around Europe – Smart cities

- Amsterdam Data Exchange – calling for pilot projects, Joint meeting with EC and JRC
- Data Market Austria @ ICT2018 4-6 Dec 2018, Vienna
- Data Exchange pilot by Copenhagen city, Barcelona city, etc.

Focus

- Use of open city/municipal data: transport, energy, traffic, etc.
- Smart cities: ecology, energy, water, etc

Need for Usable Data Pricing Model

- Data production and maintenance cost money
  - In particular, sensor and IoT network
- Data Exchange infrastructure cost money
- Buying data from professional data producers (or markets/brokers) will save money and allow focusing on application aspects
- Need for smart data contracts and properties embedding
  - Ensuring policy enforcement

Data pricing model - Origin and related papers

- Daniel Moody, Peter Walsh, Measuring The Value Of Information: An Asset Valuation Approach, 1999
  - 7 Laws of information
  - Few followers to map to data properties

- J.Heckman, E.Peters, N.G.Kurup, E.Boehmer, M.Davaloo, A Pricing Model for Data Markets, iConference 2015 Proceedings

- A. Muschalle, F.Stahl, A.Loser, G.Vassen, PricingvApproaches for Data Markets, Proc. International workshop on business intelligence for the real-time enterprise, 2012

7 Laws of Information by Daniel Moody and Peter Walsh (1999)

"Measuring the value of information: An asset valuation approach" by Daniel Moody and Peter Walsh (1999)

- First Law: Information is (infinitely) shareable
- Second Law: The value of information increases with use
- Third Law: Information is perishable
- Fourth Law: The value of information increases with accuracy
- Fifth Law: The value of information increases when combined with other information
- Sixth Law: More is not necessarily better
- Seventh Law: Information is not depletable

Data Pricing Factors (J.Heckman, 2015) - Attributes selection

- Value based parameters (value of data to the consumer)
  - The value of data in terms of saving time, efforts or money
  - ROI for customer
  - Risk exposure
  - Data exclusivity
  - Level of ownership (ownership transfer)
- Qualitative parameters (attributes or meta-attributes of the datasets)
  - Age of data
  - Credibility
  - Accuracy of data elements
  - Quality
  - Format and structure
- Fixed and marginal costs parameters (directly measurable costs)
  - Data collection, storage, maintenance
  - Delivery cadence

Philosophy research on Information and Data

- Pyramid: Data – Information – Knowledge
  - Many research and publications but time to revisit
- The philosophy of information by Luciano Floridi. Oxford University Press (2011)
  - 18 Open Problems
- Floridi's "Open Problems in Philosophy of Information", Ten Years Later, by Gordana Dodig Crnkovic (Sweden), Wolfgang Hofkirchner (Austria)
- An application of the dynamic knowledge creation model in Big Data, By Jestine Philip (2018)
  - SECI process including social/community phase/cycle
  - Leverage 'Ba' concept of "time-space-nexus" from Japanese philosophy to include the shared context for knowledge creation

Data Markets and Economics of “superstars”

- **Amazon, Microsoft, Google, Facebook, Apple**, IBM, Baidu/Alibaba?
- Big technology companies use exclusiveness of information and data to get (sometimes unfair) market leadership and dominance
  - Discussed at the last IMF 2017 conference (paper by A.Korinek)
  - Possibly regulation is needed to Data Market democratisation
- Data value increases with more (different) data collected

- Anton Korinek, Ding Xuan Ng, The Macroeconomics of Superstars, November 2017 [online] https://www.imf.org/~/media/Files/Conferences/2017-stats-forum/session-3-korinek.ashx
- Managing Our Hub Economy, by Marco Iansiti, Karim R. Lakhani, Harvard Business Review, September 2017 [online] https://hbr.org/2017/09/managing-our-hub-economy

Data Tao

- From records to logs and monitoring to raw data as economic goods
  - Meaning of data as representation of the process
  - Value of data as guide for action – Actionable data
- Data – Information – Knowledge – Action – Production – Development (social) – (Life?)
  - How deep link to data should be maintained?
  - Data well – The deeper data we can maintain the greater transformation we can do
- Data value transformation
  - Data transformed are the same or new data?
- Data commoditization
- Data monetization
  - Requires quality measure and price model

Data Market Architecture components

Data Exchange – Data Connectors – Catalog - Brokers - Trust

- Architecture and conceptual model of the Data Market space, including **technological, organisational, legal and commercial aspects**
- **Data Exchange(s)** as the main component for Data Market actors interaction and data exchange
- **Data Connectors** to enable sovereign end-to-end data provider and consumer connection
- **DataHubs** to support for generic services for data suppliers such as caching, streaming, containerised delivery
- **Federated Access Control and Trust Management** infrastructure to access and operate the Data Market
- Federated hybrid cloud based Big Data infrastructure to support data storage, processing and exchange in a secure and trusted way
- Support for on-demand connectivity and bandwidth provisioning between data handling services/hosts in the data lifecycle
- Gateway based and computationally enforcement of market policies and rules

Combining Data and Algorithms

- With the complexity of modern data collected and generated by human activity and IoT
  - Metadata and schema are not enough (passive knowledge)
  - Algorithms and API are required (active knowledge)
  - Blueprint for use and deployment (ready for integration)
- Moving data to compute vs compute to data
  - Actually Hadoop Distributed File System (HDFS)

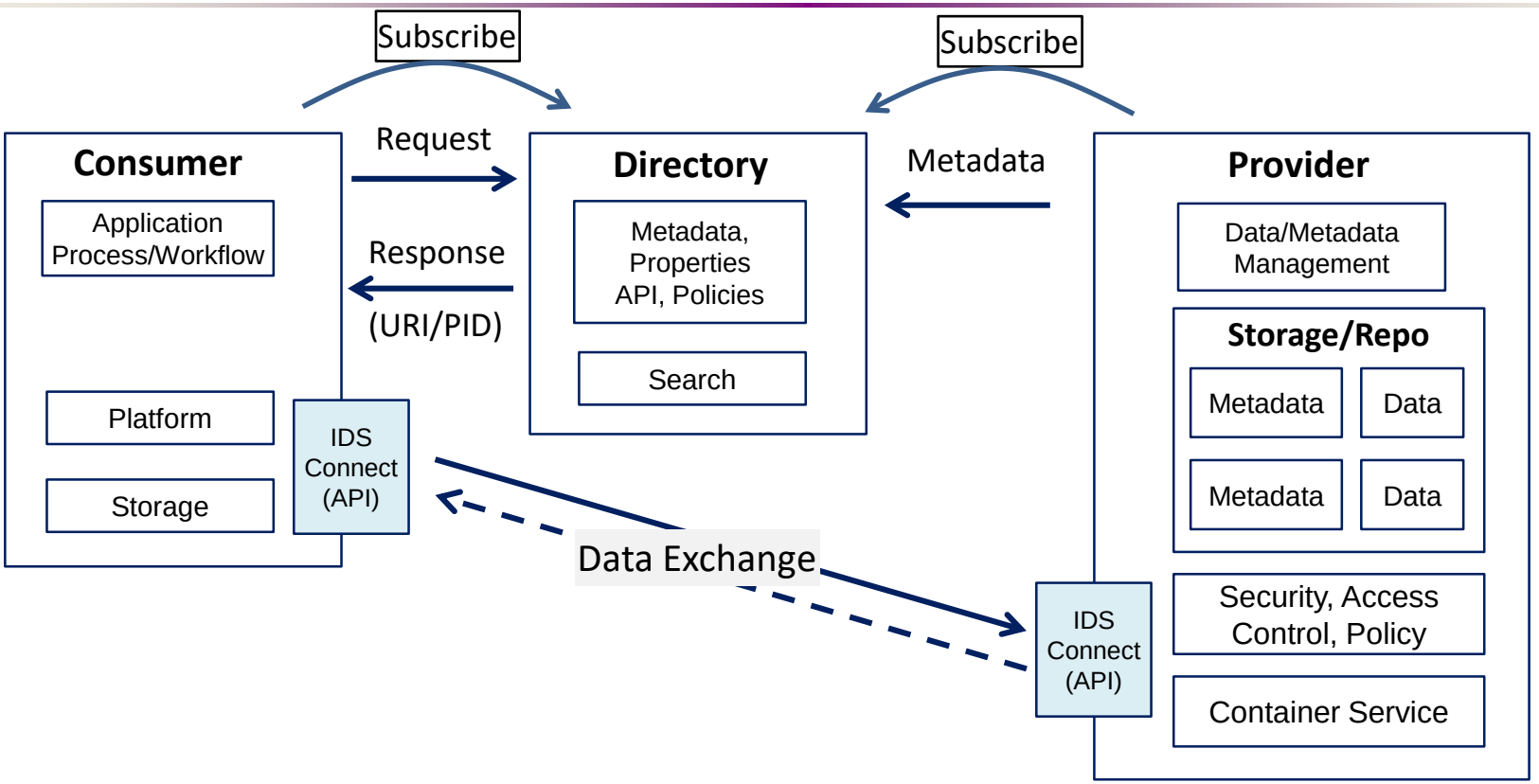
Holistic Approach to Data Markets Definition and Architecture

- **Main functional components and actors/roles in Data Markets**
  - Consumer/User/Data target
  - Producer/Provider/Data Source/Data Owner
  - Directory/Catalogue
  - Broker (Provider side and Consumer side)
  - Infrastructure (Data Exchange, Storage/Cache/Data Lake, Data Analytics)
  - Data connector (containerized API leveraging IDS Connector architecture)

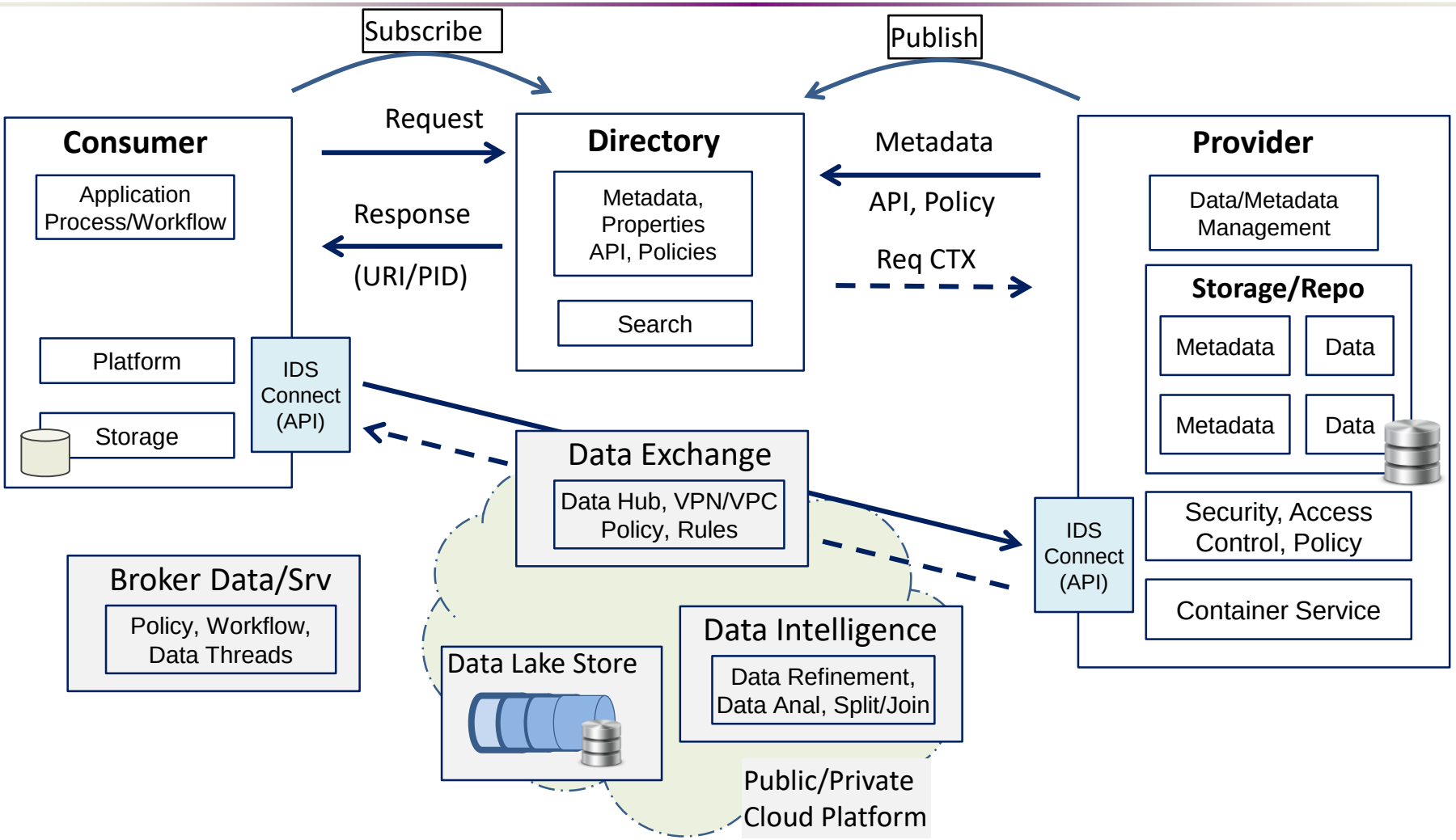
- **Interaction models (supported by API)**
  - Consumer – Provider: 1 x n, n x 1, n x m
  - Data exchange model:
    - Consumer – Directory – Provider
    - Consumer – Directory/Broker – Provider
    - Consumer – Directory – Data Exchange – Provider
    - Consumer – Directory/Broker – Data Exchange + Data Hub (Storage/Cache, Data Analytics) – Provider
- **Business model:**
  - Internal: integrated with the organisational workflow
  - External/open: Data markets in a public space
  - Hybrid: Internal/external + Federated/distributed/multi-partner + Community marketplace (MIDIH)

- Data Containers and/or Application Containers
- Data Lifecycle – Data Workflow – Digital Threads

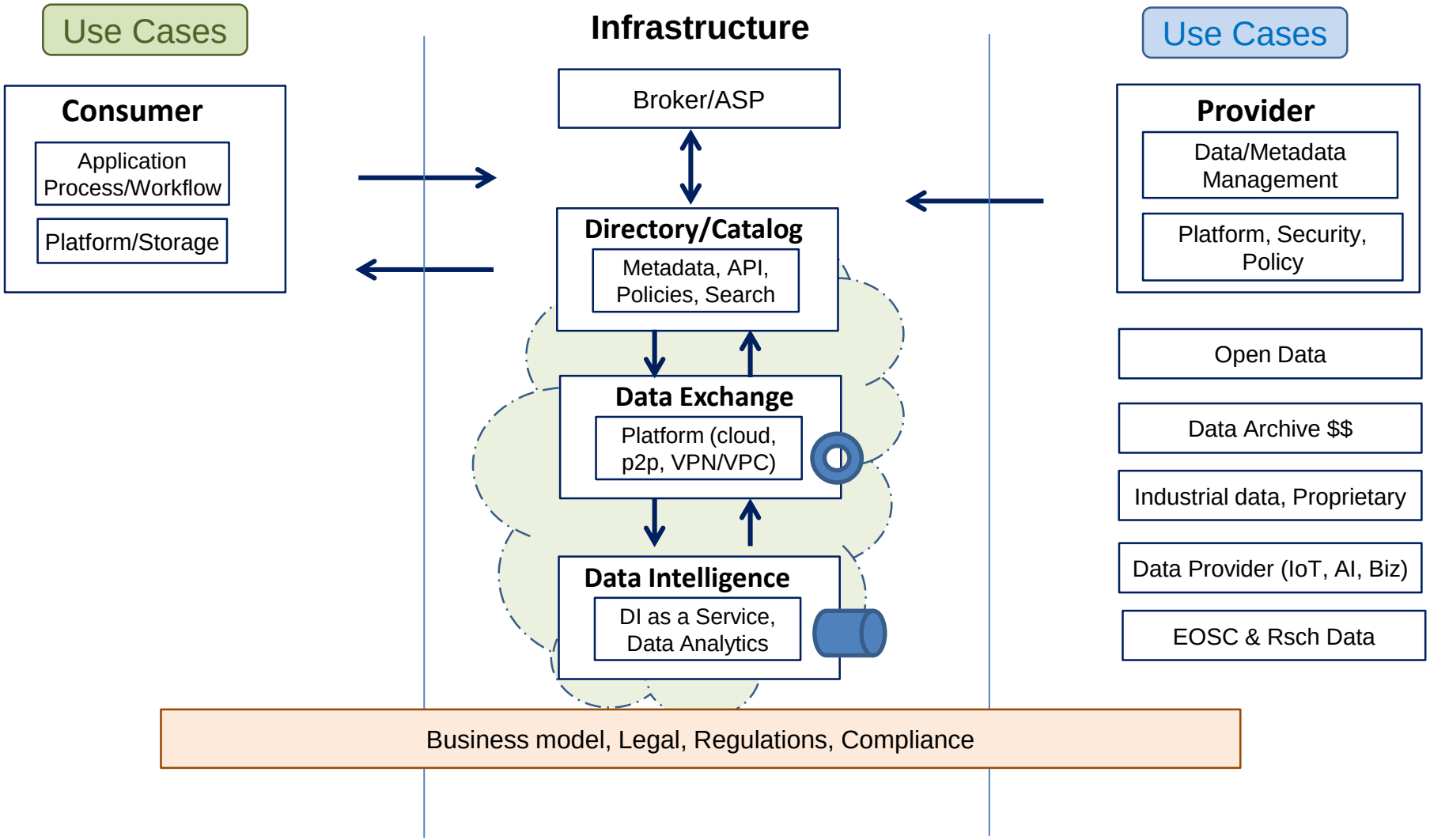
Data Market models:
(1) Consumer – Directory – Provider



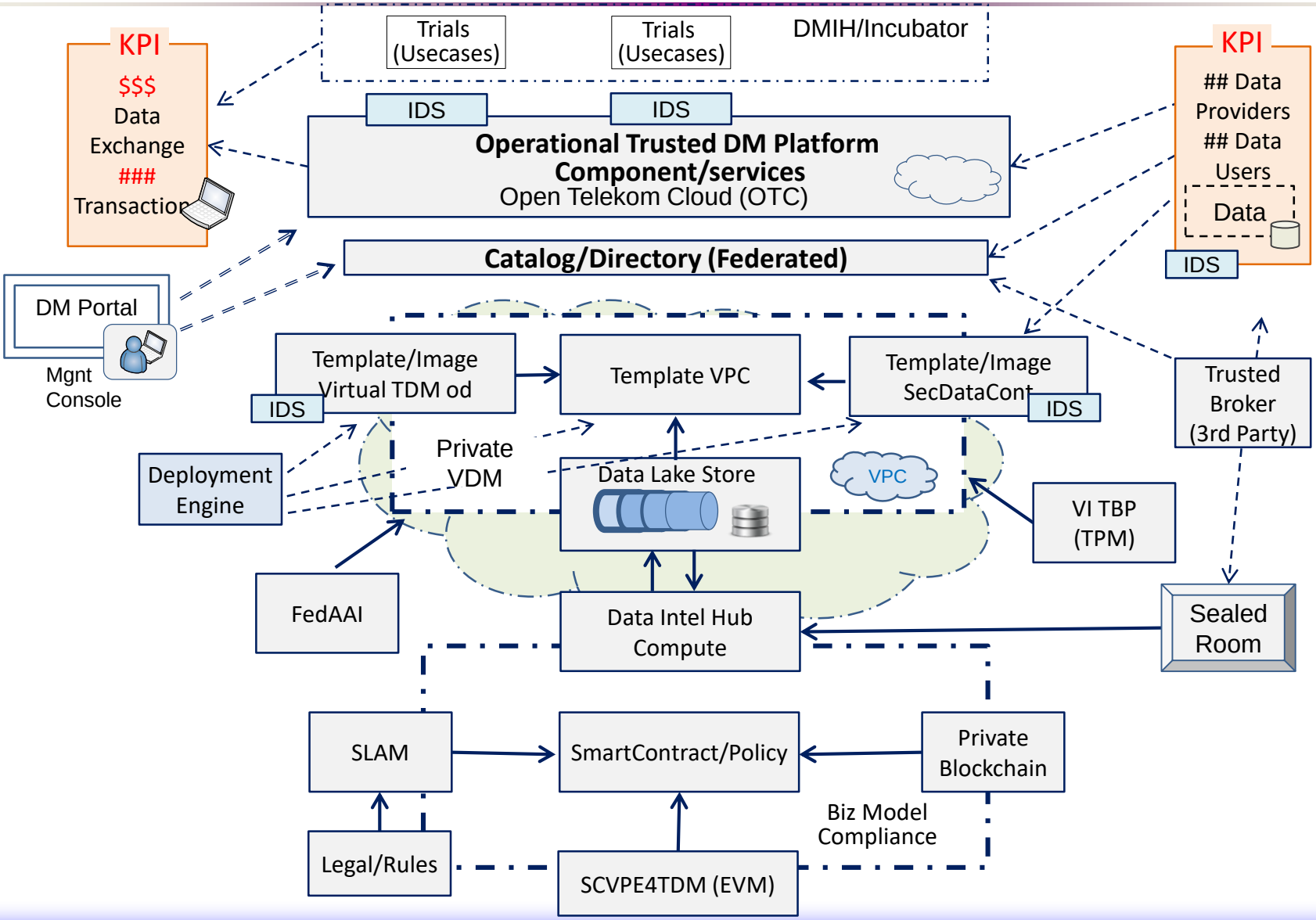
Other Data Market models: Extending Services with Broker, Data Exchange, Data Hub, Data Analytics



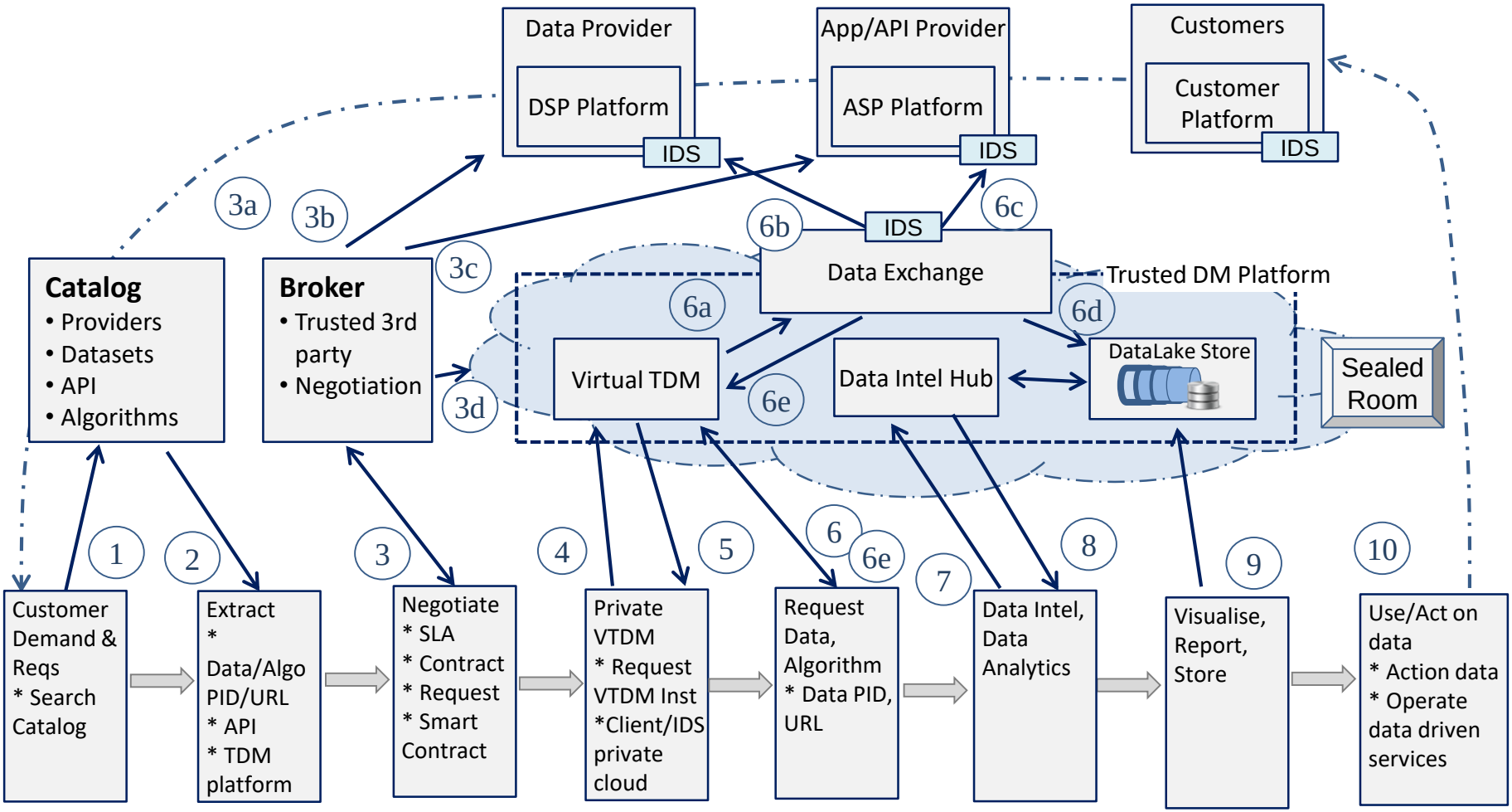
Data Market Ecosystem – Main Roles



Trusted Data Market composable components



Data Market Workflow: Provisioning Virtual Private DM Infrastructure



Data Exchange protocol - Components

- PID, OID, Data Registries, Data Factories – *All RDA output*
- Metadata and data types registry, data annotation and data discovery mechanisms
- Directory and registry (to enable publish - subscribe)
- API and schema
  - Combining data and algorithm
- Provenance, Auditing
- Blockchain for data exchange tracking/provenance/lineage and validation
- Trust Management and Trust establishment Protocol
- Policy and rules construction (+ enforcement)

- Build on top of reliable and proven by practice Internet protocols
  - ***Data exchange protocols defined as upper layer protocols***
- Leveraging Virtual Networks and Virtual Privat Clouds
- Leveraging IDS architecture

Existing Standards and Mechanism to Identify and Manage Data Exchange

- Industrial Data Space Architecture
  - Focus: Data Exchange while preserving Sovereignty
- NIST - Big Data Architecture (Interoperability) Framework
- Industry 4.0
- PID - RDA Persistent Identifier
- OID – Object Identifier
- OData - OASIS Open Data protocol 4.0
- Sensor Data
- Industrial Internet Consortium (IIC)
  - Industrial Internet Reference Architecture (IIRA)
  - Industrial Internet Security Framework (IISF)

References

- Data as Economic Goods: Definitions, Properties, Challenges, Enabling Technologies for Future Data Markets, by Yuri Demchenko, Wouter Los, Cees de Laat. Position paper. To appear in ITU Journal: ICT Discoveries, Special Issue “Data for Goods”, 23 November 2018 https://www.itu.int/en/journal/002/Documents/ITU2018-12.pdf https://www.itu.int/en/journal/002/Pages/default.aspx
- Research Data Alliance (RDA) BoF on Data as Economic Goods https://www.rd-alliance.org/bof-data-properties-economic-goods-rda-12th-plenary-meeting
- RDA12 Poster "Bringing Data to Market: Data Properties as economic goods" http://www.uazone.org/demch/posters/rda12poster29-Data-economic-goods-markets-v01.pdf
- FAIR Data Initiative - https://www.dtls.nl/fair-data/
- Industrial Data Space Reference Architecture, Version 2.0 https://www.internationaldataspaces.org/wp-content/uploads/2018/04/InternationalDataSpacesAssociation_ReferenzArchitecture2.0.pdf
- DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition), DAMA International, July 2017.

Technical Addendum

- Data types revisited
- Big Data Infrastructure and Data Markets
- Existing standards
- International Data Space Architecture overview

Data types (not exhaustive & not authoritative)

- Personal data
- Non-personal data
- Personal operational data
- Industrial data
- IP data
- Metadata
- Research data
- Open data
- Public data
- Social Media data
- Network operational/log data

Non-personal data

- Open data
- Process control data
- Process recording/log data
- City data
- Surveillance data?

Personal data

- Personally identifiable data (P.I.D.)
  - Can be not only directly containing Name, DoB, Address
  - GDPR requires searching and identification of all P.I.D. in data collected by companies
- Control over personal data use required by GDPR
  - Ownership
  - Transformation and usage
- Privacy metrics
  - Essential for technology evaluation and design
- Privacy by Design (PbD)

Combining Data and Algorithms

- With the complexity of modern data collected and generated by human activity and IoT
  - Metadata and schema are not enough (passive knowledge)
  - Algorithms and API are required (active knowledge)
  - Blueprint for use and deployment (ready for integration)
- Moving data to compute vs compute to data
  - Actually Hadoop Distributed File System (HDFS)

Modern data architecture vs Data Market

Characteristics of modern data architecture

1. Customer-centric
2. Automated
3. Smart
4. Adaptable, Agile
5. Cloud based
6. Elastic
7. Collaborative
8. Governed
9. Secure, Trusted

Characteristics of emerging data markets

1. Customer-centric
2. Automated
3. Smart
4. Regional/sectoral specialised
5. Cloud powered/integrated
6. Collaborative
7. Governed
8. Secure, Trusted
9. Auditable
10. Transparent
11. Commoditised/Monetised
12. Combining data and algorithms (as part of containers)

Data Catalog/Directory properties and functions

1. Cataloging data sets
2. Cataloging data operations
3. Metadata Catalog
4. Searching
5. Recommendations and relationships
6. Data sets evaluation
7. Quality
8. Data Curation
9. Data access (metadata)
10. Usage metadata
11. Lineage/Provenance
12. Integration and interoperability
13. Deployment (cloud, on-premises, hybrid)
14. Services
15. Data visualisation
16. Security
17. Compliance
18. Socialisation
19. Pricing
20. Vendor roadmap

Reflecting STREAM data properties as economic goods

Additional information

- Garry Berg-Cross, Ontolog

Data has economic and social value (by Garry Berg-Cross, Ontolog)

Research data has social value - things we want, such as

- improvements in agriculture
- social wellbeing in fields such as
- education healthcare, public safety and security

People and group differ in how they evaluate these

Social value comprises benefits for single users as well as larger societal/community benefits such as employment growth, productivity, and consumer surplus

Measuring Data's Economic value (by Garry Berg-Cross, Ontolog)

Some are macro:

- organization's increase in profit,
- business growth, and/or
- competitive advantage resulting from big data adoption
- Economic value often comprises monetary benefits that are appropriated by organizations.
- Some value is avoiding expense such as building walls to keep out rising ocean levels.

What is (digital)Data? What type of object is it? Symbolic? (by Garry Berg-Cross, Ontolog)

- Data is a collection of datum.
  - ***Data plays a role as a potential information bearer to a cognitive agent.***
  - ***Data is a collection of propositions representing an agent's understanding of some entity or state.***
- This gives a concrete and persistent status to some (potential) information, but it is by itself (as a string of bits, for example). Without context such as the assumed formatting/ representation information or reference systems it is without a single meaning.
- Thus, to extract the information intended in collections of data, an interpretation is necessary that assigns meaning to it using elements of this context such as assumed representation.

(Research) Data plays a **resource role**

- Data, especially digital forms seems a type of “resource.”
- It may be an institutional object, such as money, drivers’ licenses, & borders, with functions because of their social roles rather than immediate physical properties.
- Is it created from a raw form to be hosted, mined and used?
- Even raw research data has some assumed fitness for a purpose
- There are aboutness qualities to data than make it valuable for human operations, but hard to value objectively.

Technology Overview

Data Markets and Data Infrastructure – Overview Related Activities

- IDS – International Data Space
- BDVA – Big Data Value Association
- RDA – Research Data Alliance
- NIST – National Institute for Standardisation and Technology, USA
- OASIS – Advancing Open Standards for Information Society

RDA: PID, Data Types Registry, Data Factory

- Traded as one of key RDA deliverables
- Already implemented
- Data Factories:
  - Work continues
  - Attempt to benefit from Internet experience using analogy with Internet Protocol and IP addresses

OASIS OData 4.0 Standards for an Open, Programmable Web

- OASIS Approves OData 4.0 Standards for an Open, Programmable Web: *Axway, BlackBerry, CA Technologies, Citrix, IBM, Microsoft, Progress Software, Red Hat, SAP, SDL, and Others Enhance Open Data Protocol*; 17 Mar 2014
  - ISO/IEC JTC 1 Approves OASIS OData Standard for Open Data Exchange, 23 Feb 2017
- Current use and importance for ODM
  - Exchange of information about trade operation and goods movement
  - Can be extended to data along with the STREAM data properties

Industrial Data Space architecture

- Industrial Data Space architecture Association
- Industrial Data Space architecture
- Use cases

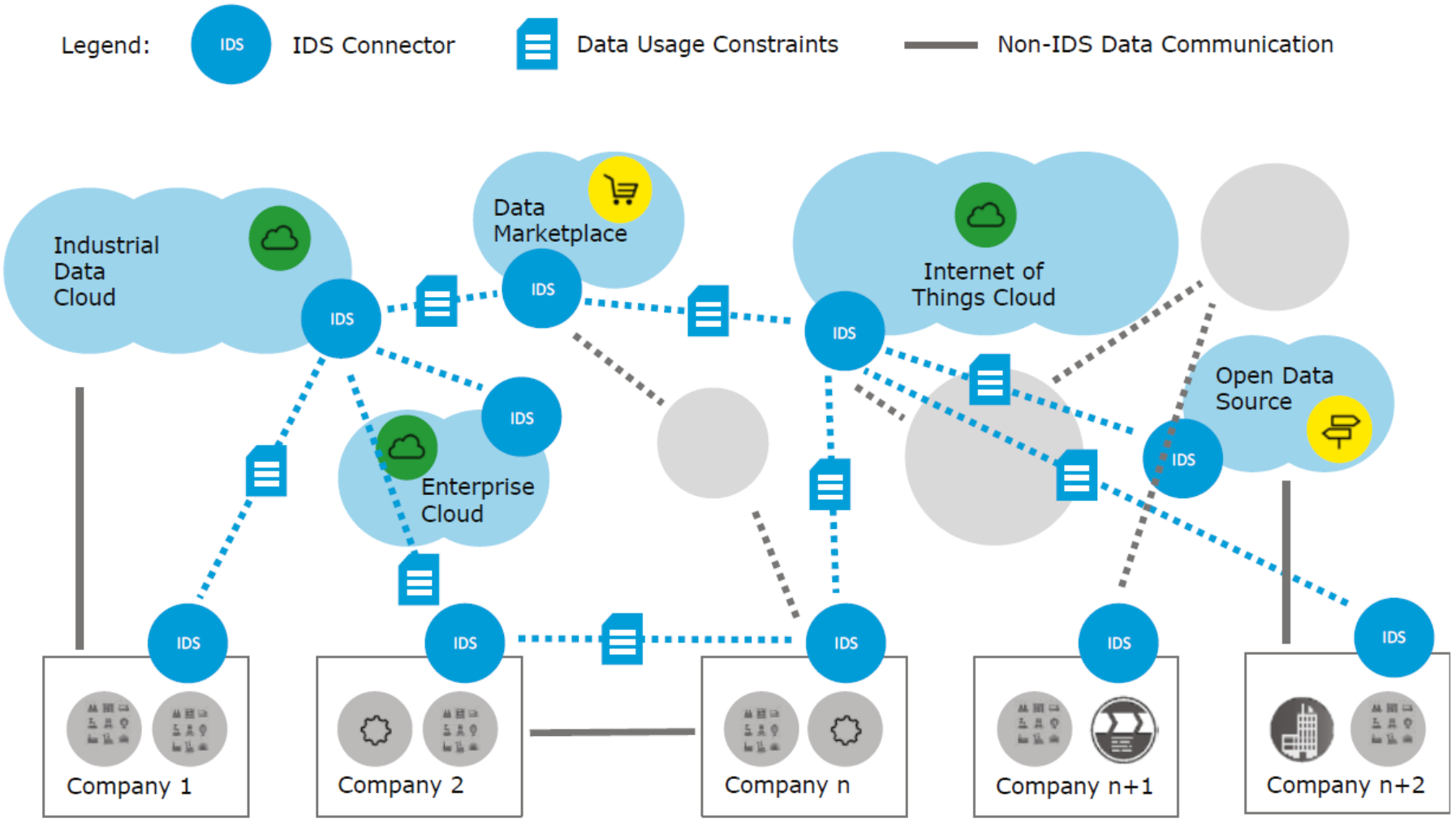
International Data Space Association

- Started 2016 as Industrial Data Space initiative (supported by German project)
- Re-defined as International Data Space Association (IDSA)
  - Published International Data Space Architecture Version 2.0 (2018)
  - Whitepaper and use cases
- Associated H2020 projects
  - Boost4.0 – Big Data for Factories (20 Mln (100 Mln private), 3yrs, 50 partners, 16 countries)
  - MIDIH – Manufacturing Industry Digital Innovation Hub (22 partners, 12 countries)
    - Services: technological, business, skills building
    - Open calls
  - Close cooperation with FIWARE Foundation (cloud like infrastructure resulted from Future Internet program)
    - Positions itself against IoT and Open-Data solutions in the areas of smart cities, Industry 4.0 and agriculture
- Ongoing active outreach campaign
  - Appearance at FIWARE Global Summit 8-9 May 2018 in Porto
  - Serial of webinars in Sept – Oct 2018

International Data Space Use cases

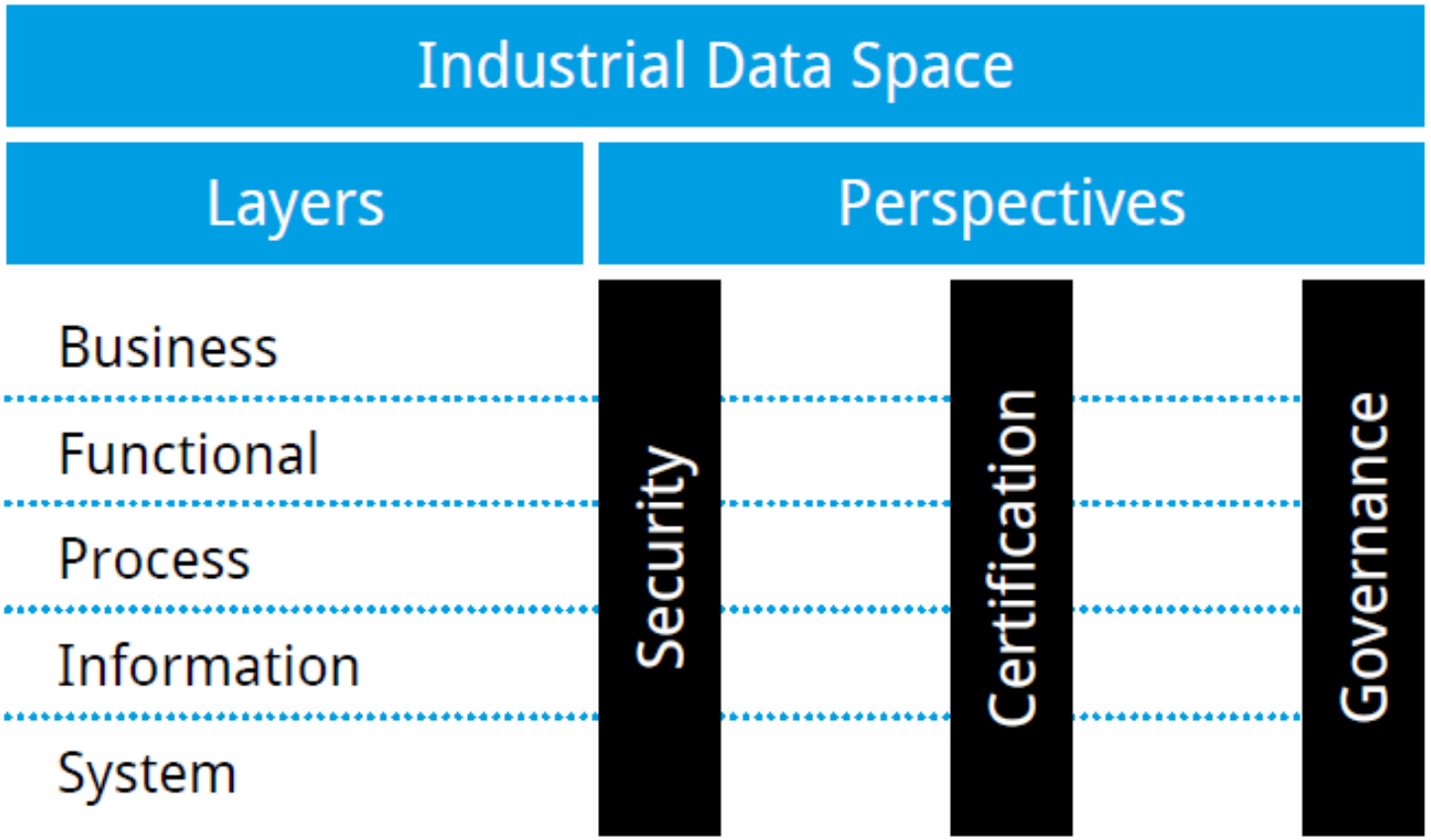
- Total 10+ in the published document (2017)
- UC#01: Advaneo: Advises European railway companies on standardisation and digitalisation
  - Broker for a secure data exchange on a Virtual Data Market place
  - Broker service arranges the secure exchange of data between provider and user via IDS structures and ***metadata*** exchange
- UC#05 DataAhead: Renewable Data Management – Readiness for multi-stakeholder
  - Open source components
- UC#09 Fraunhofer: INSTEAD – Information Sharing to Advance Antibiotic Discovery
  - IDS Connector (Internal and External), Broker
- UC#10 Nicos AG: Identity Provider in the Environment of IDS
  - IDS Connector, Identity Provider
- UC#12 SETLOG: Predicting Lead Time (transport arrival for textile industry)
- UC#14 Telekom: Data Intelligence Hub
  - IDS Connector, Broker, App Store
- UC#** Atos: Broker based design of Supply Chain
  - Connect all supply network, optimise transportation order, enable tenders

IDS and Cloud



- IDS Connector is the main functional component
- No specifically defined infrastructure

General Structure of IDS Architecture

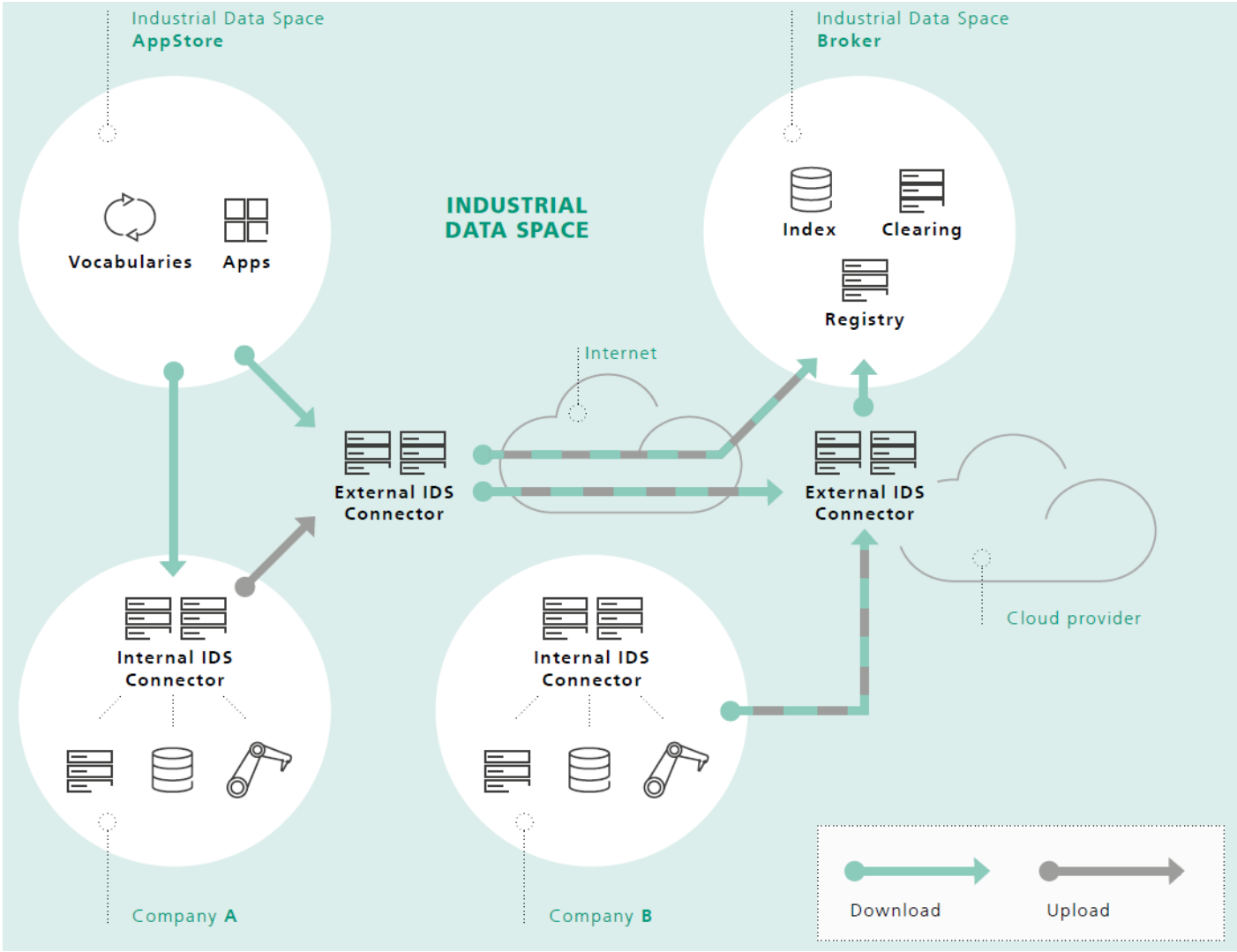


- Specification defines functionalities by layers
- Details are sufficient to define processes, functional components and API

International Data Space Architecture Version 2.0

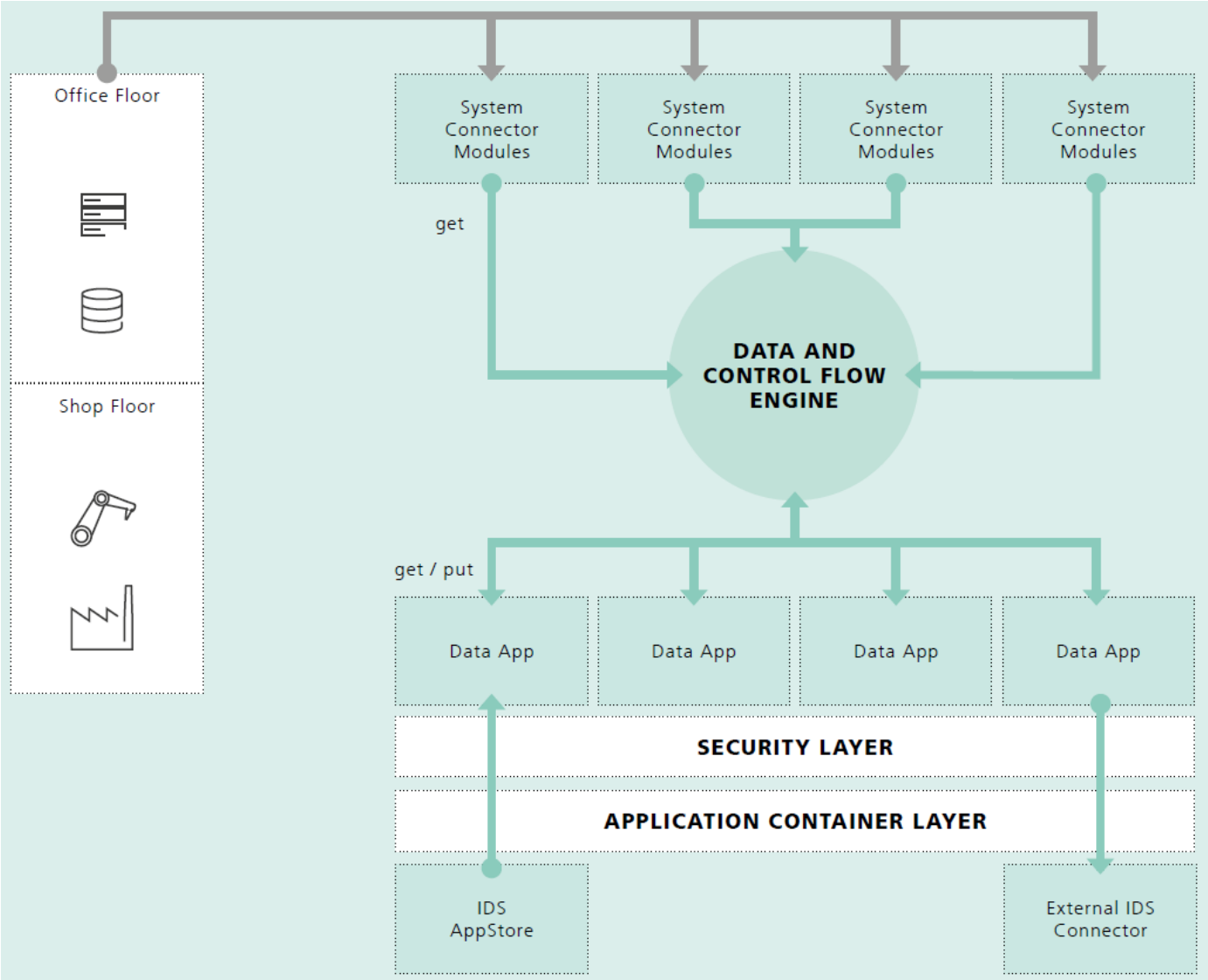
Industrial Data Space AppStore	Basic Data Services Provisioning	Data Service Management and Use	Vocabulary Management	Software Curation
	Data Provenance Reporting Data Transformation Data Curation Data Anonymization	Data Service Publication Data Service Search Data Service Request Data Service Subscription	Vocabulary Creation Collaborative Vocabulary Maintenance Vocabulary/Schema Matching Knowledge Database Management	Software Quality and Security Testing
Industrial Data Space Broker	Data Source Management	Data Source Search	Data Exchange Agreement	Data Exchange Monitoring
	Data Source Publication Data Source Maintenance Version Controlling	Key Word Search Taxonomy Search Multi-criteria Search	»One Click« Agreement Data Source Subscription	Transaction Accounting Data Exchange Cleaning Data Usage Reporting
Industrial Data Space Connector	Data Exchange Execution	Data Preprocessing Software Injection	Remote Software Execution	
	Data Request from Certified Endpoint Usage Information Maintenance (Expiration etc.) Data Mapping (from Source to Target Schema) Secure Data Transmission between Trusted Endpoints	Preprocessing Software Deployment and Execution at Trusted Endpoint	Data Compliance Monitoring (Usage Restriction etc.) Remote Attestation Endpoint Authentication	

IDS Architecture Software Components



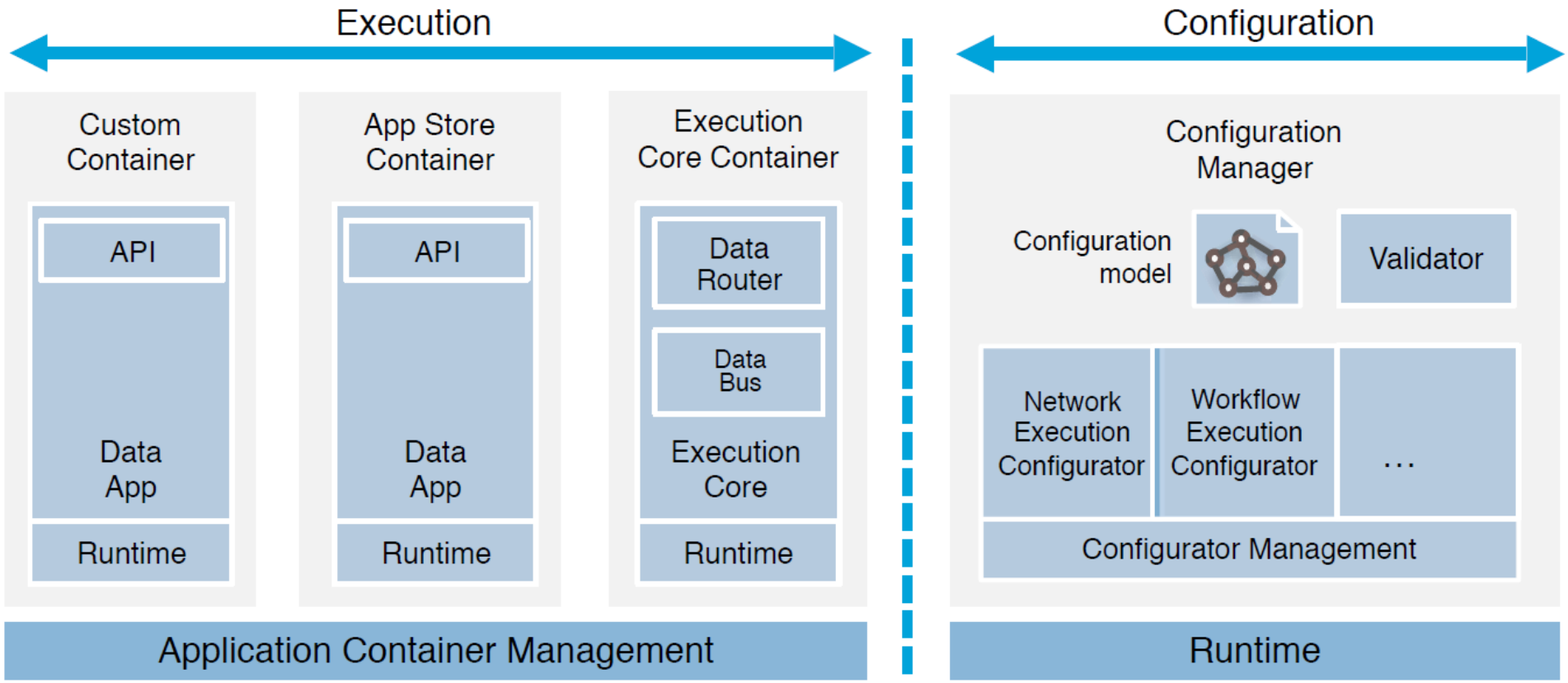
- Currently no special shared infrastructure components are defined
- Issues
  - Interoperability
  - Compatibility
  - Apps development
  - Monitoring/Auditing
- Required Open Infrastructure components
  - Data Exchange
  - Directory
  - Data cache/delivery network
  - Trust broker?

IDS Internal Data Connector



- Simple data relay/ forwarding service
- Issues
  - Does not define standard API
- Required features
  - Protocol for working with the IDS and Data Market Infrastructure
  - Integration with Enterprise IT/Data Warehouse infra

Reference Architecture Data Connector



- Execution and configuration
- Application container

Security Perspective: IDS Security Architecture

- General Security Principles
- Key Security Concepts
  - Communication Security, using IPSec
  - PKI and Trust management
  - Trusted Platform
  - Isolation and remote execution
- Connector security profiles
  - Aspects and local/remote enforcement
- Access Control
  - RBAC model and XACML
  - PEP are integrated into IDS connector
- Data usage control

IDS Certification Perspective

- Applied at different layers
- Roles in the certification process
  - Certification body, Evaluation, Applicant
- Targets of certification

IDS Governance Perspective

- Governance aspects on different architecture layers
- Data as an economic goods
- Data Ownership
- Data Sovereignty
- Data Quality
- Data Provenance

Leveraging IDS – Infrastructure

- Extend Trust beyond PKI
  - Trusted Computing Base (TCB)
    - Remote Attestation, Bootstrapping
    - Trusted remote execution (and storage)
    - Intel TXT and SGX technology for trusted VM/container deployment
  - Federation and Federated clouds
- Use modern Big Data Infrastructure and tools
  - Storage and Computing for Big Data
  - Cloud infrastructure: Hybrid model allowing outsourcing workload and temporal storage from private cloud to public cloud
- Data (management) infrastructure
  - Data is infrastructure itself
  - Registries, Data Factories, PID