TC DATA
The Trail to
Open Data Infrastructures

Presented by: Diego López

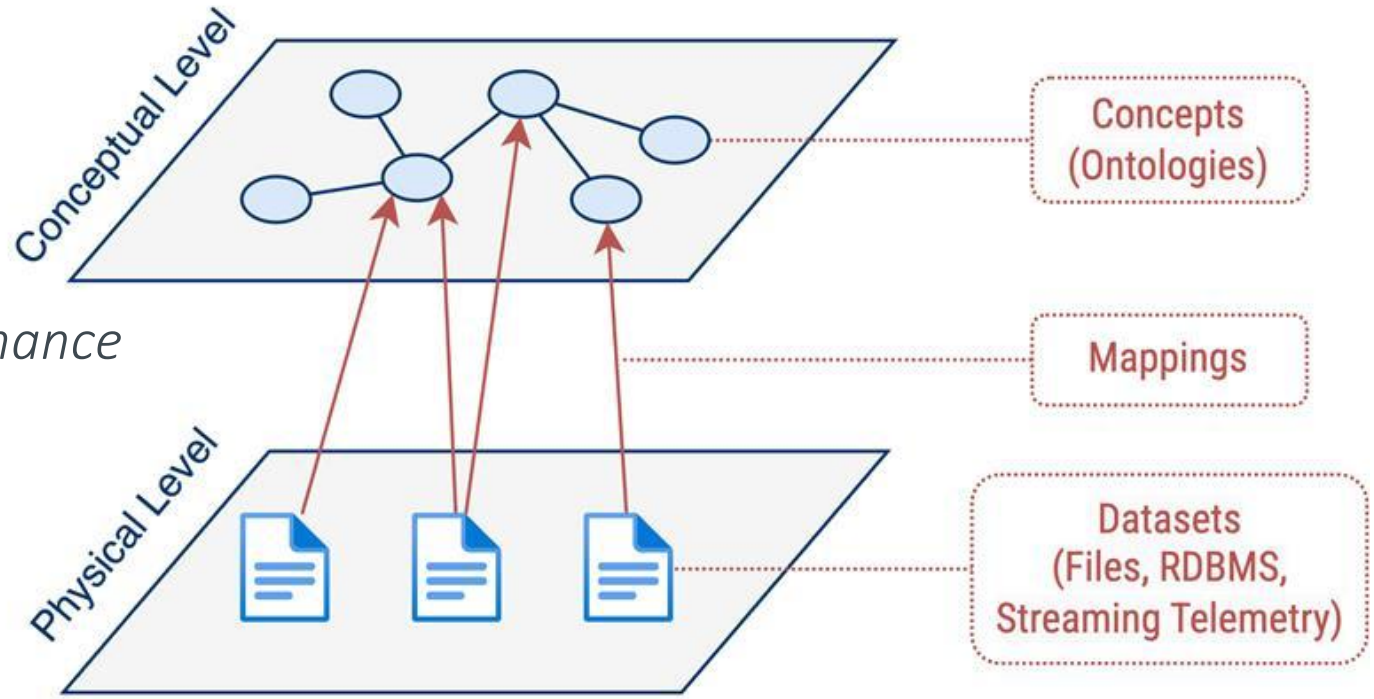
2nd April 2025

Where We Come from

- Data exchange is the essential functionality of any *data* network
  - The requirements for such data exchanges evolve rapidly as their applications evolve
- We are witnessing an exponential acceleration of the consolidation of agent applications
  - Acting autonomously according to intent
  - Being able to collaborate and select appropriate data
  - Nowadays, we often use the term AI to refer to them
- Data has been identified as the lifeblood for AI
  - The key aspect to facilitate AI interoperability
  - More if it becomes distributed/federated/…
- Data is central in the ETSI Technology Radar

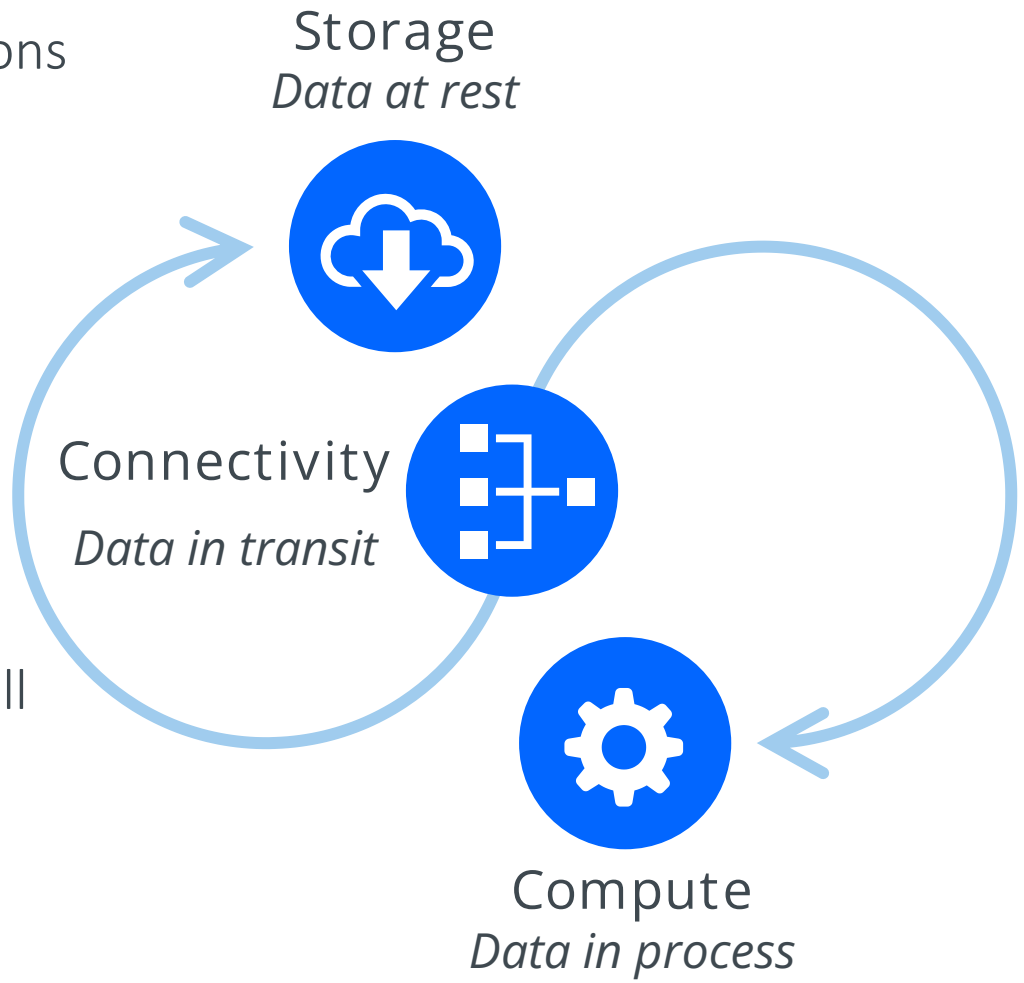
What We Want to Achieve

- Support the autonomous use of data by agent applications
  - And by any other data-driven technology
- Aligned with FAIR principles
  - **F**indable
  - **A**ccessible
  - **I**nteroperable
  - **R**eusable
- According to an appropriate *data governance*
  - Data access control
  - Data consistency
  - Data privacy preservation
- Include the associated policies
  - And support their enforcement

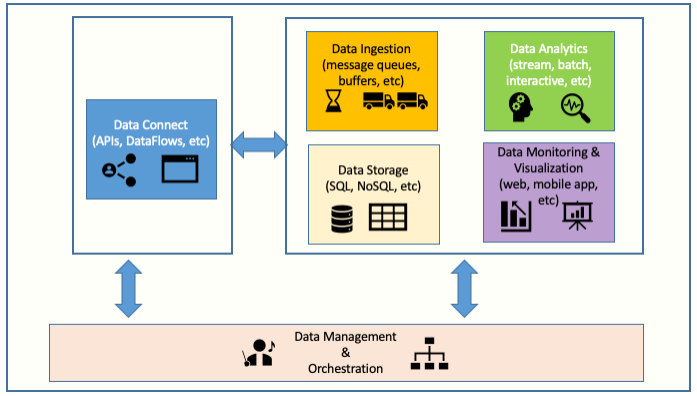


How Many Dimensions under Consideration

- Data management implies three independent dimensions
- *Connectivity*: data transported through space
  - Data in transit
- *Storage*: data transported through time
  - Data at rest
- *Compute*: data acted upon
  - Data in process
- Current applications require to focus on data flows in all three dimensions
  - Whatever the stakeholder: Telcos, cloud providers, application developers…

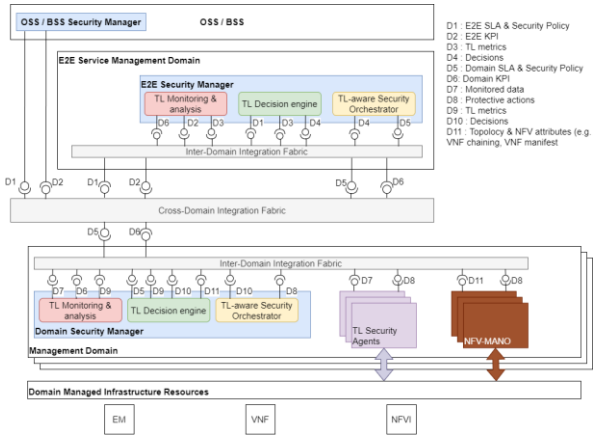


Applicable to Any Aspect of the Lifecycle



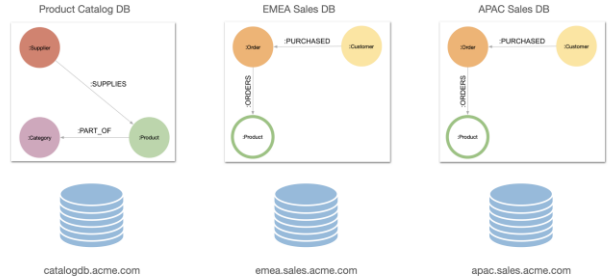
Input flows

- Deal with heterogeneity and access conditions
- Usable: Adaptation (formats, scales…)
- Sufficient: Topology (sources, aggregators…)
- Safe: Provenance (origin, timestamps…)
- Steady: Continuity (pace, availability…)



Output flows

- Similar nature and needs, whatever the target
- Control in closed loops
- Recommendations
- Visualization
- …



And models themselves

- Model metadata
- Training (meta)data
- Usage evidences
- . . .

How We Got Here

A call to action on data technologies was presented to the ETSI Board

- Requesting a decisive ETSI positioning on data solutions
  - Leveraging the results of existing fruitful initiatives
  - Consolidating open data solutions
  - Fostering innovation in base technologies
  - Exploring different application domains
  - Enhancing privacy and security
  - Solidifying ETSI's presence in the data sphere

TC DATA was approved by the end of January

- And we are currently working in making it take off

The Scope

- Develop deliverables to support the deployment and operation of distributed data solutions
  - Connectivity: data in transit
  - Storage: data at rest
  - Compute: data in process
- Address European policy and regulatory requirements and engage with other regulatory bodies
  - Ensuring relevant global, regional, and national requirements
- Provide input on technical aspects of the ETSI responses to governmental requests on data solutions
  - Special emphasis on the European Data Act and to the data-related aspects of the European AI Act
- Collaborate with open-source initiatives relevant for the data domain standardization
  - Reference implementation and interoperability testing
- Cooperate with other European and international standards organizations
  - Avoid duplication of work and promote harmonization

The Proposed Activities

- Providing a centre of expertise in the area of data infrastructures, services and applications
  - In coordination with other ETSI activities
- Developing technical standards to support data interoperability and semantic interoperability
- Maintaining and evolving specifications related to data solutions and published by other ETSI TGs
  - A list of initial agreements and interests
    - TC SmartM2M and TC ESI
    - ISG CDM, ISF CIM, and ISG PDL
- Supporting the development and the maintenance of semantic and data models
  - SAREF (including the SAREF open portal)
  - NGSI-LD
- Supporting the transposition in ETSI of the outputs of oneM2M
- Supporting the maintenance and evolution of relevant industry data standards

The Security Aspects

Leverage and extend existing work

- Permissioned Distributed Ledgers, Smart Contracts and Oracles
  - Consolidating ISG PDL results and the associated European standardization effort
- Safe AI applications
  - Data aspects in connection with TC SAI
- Identity data
  - In direct collaboration with TC ESI
  - eIDAS regulations and alignment of ETSI and ISO standards regarding e-signatures
  - AdES LOTL program for mutual trust, with participation of EU and Japan

Explore new fields

- Data provenance and transparency mechanisms
- Fine-grained AAA and explainability

Come Join the Drill!

The notion that “Data are the New Oil” has been around for quite some time. Generally credited to mathematician Clive Humby: [2006]
Data, like oil, are not useful in its raw state. They need to be refined, processed and turned into something useful.

Their value lies in their potential