ITU Regional Forum on Emergent Technologies
Tunis - Tunisia, 23-24 April 2019

# Big Data trends: application domains, technologies and standardization

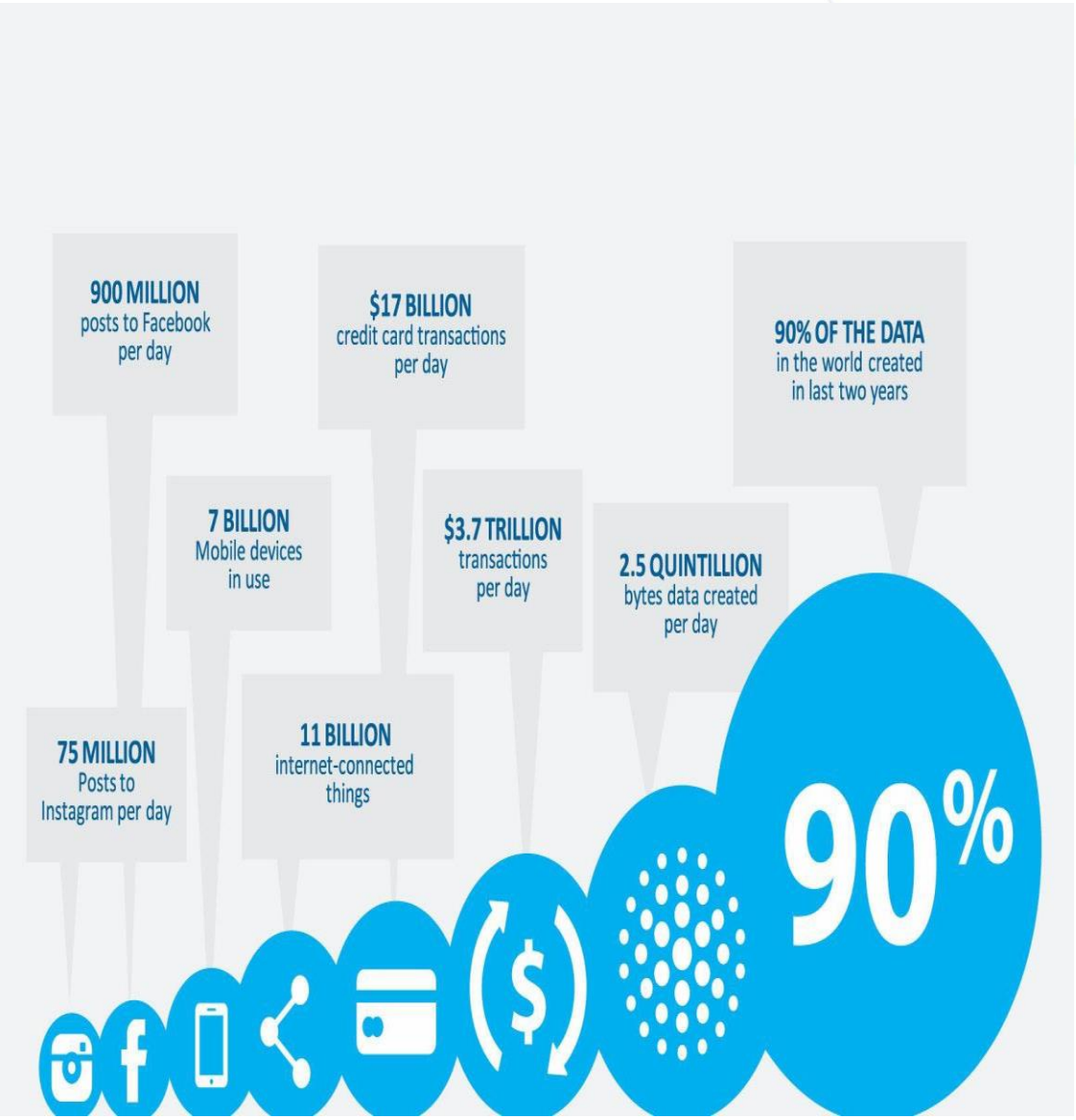
Presented by:
Marco Carugi, ITU expert
marco.carugi@gmail.com

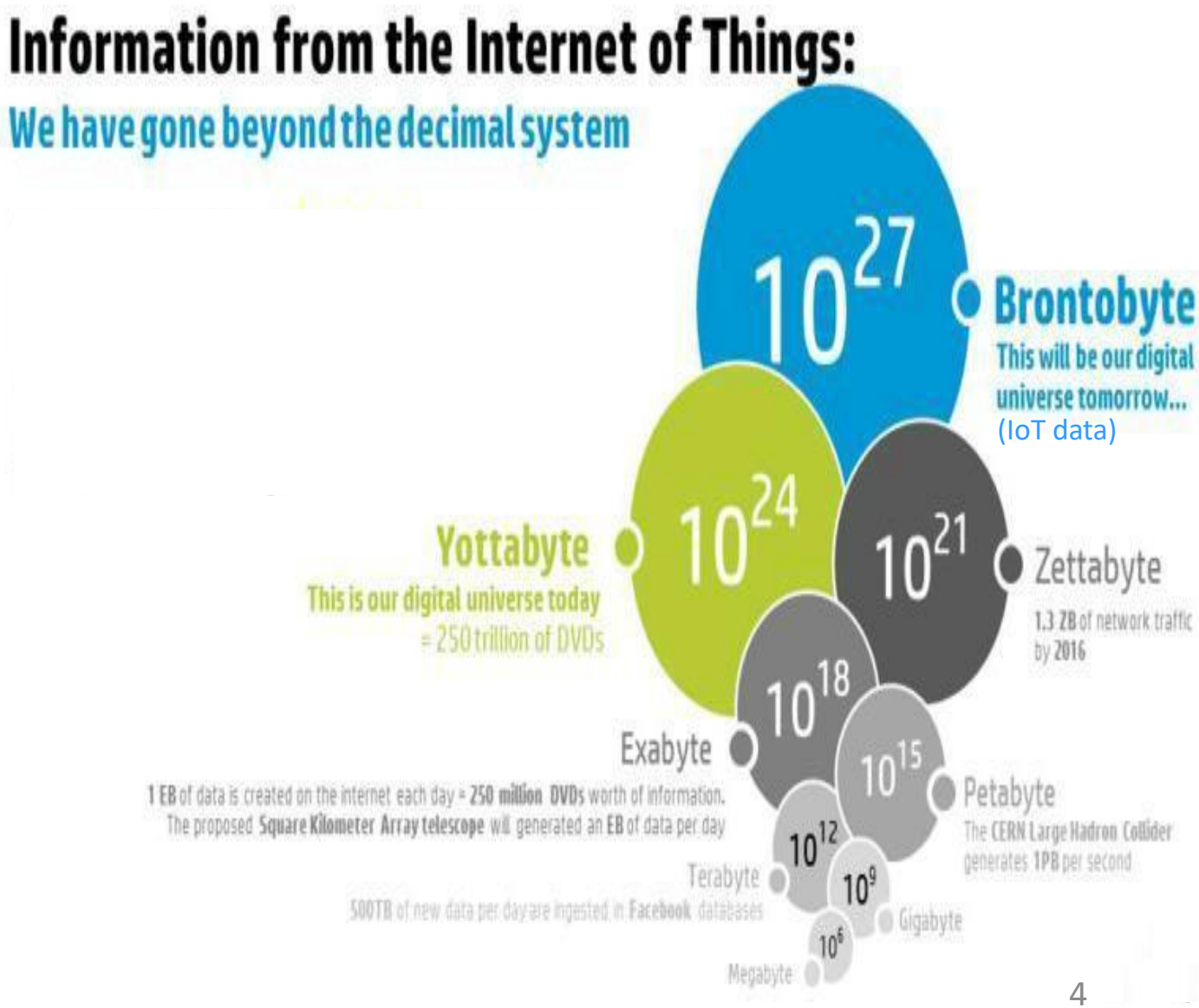
# Outline

- Introduction to Big Data, Digital Transformation and Data-driven Economy
- Application Domains of Big Data - few examples
- Big Data Technologies - highlights
- As backup information: few highlights on Big Data Standardization

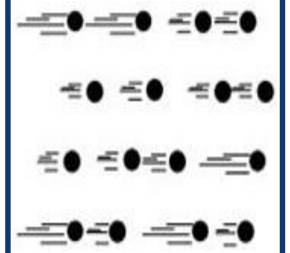
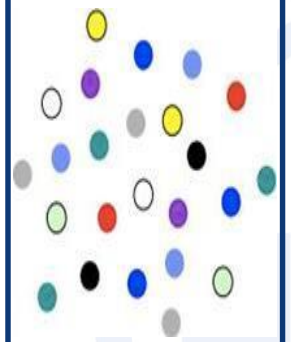
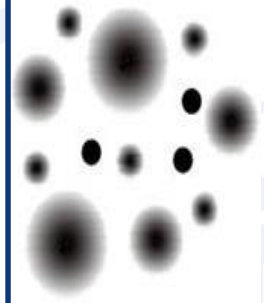
# Introduction to Big Data, Digital Transformation and Data-driven Economy

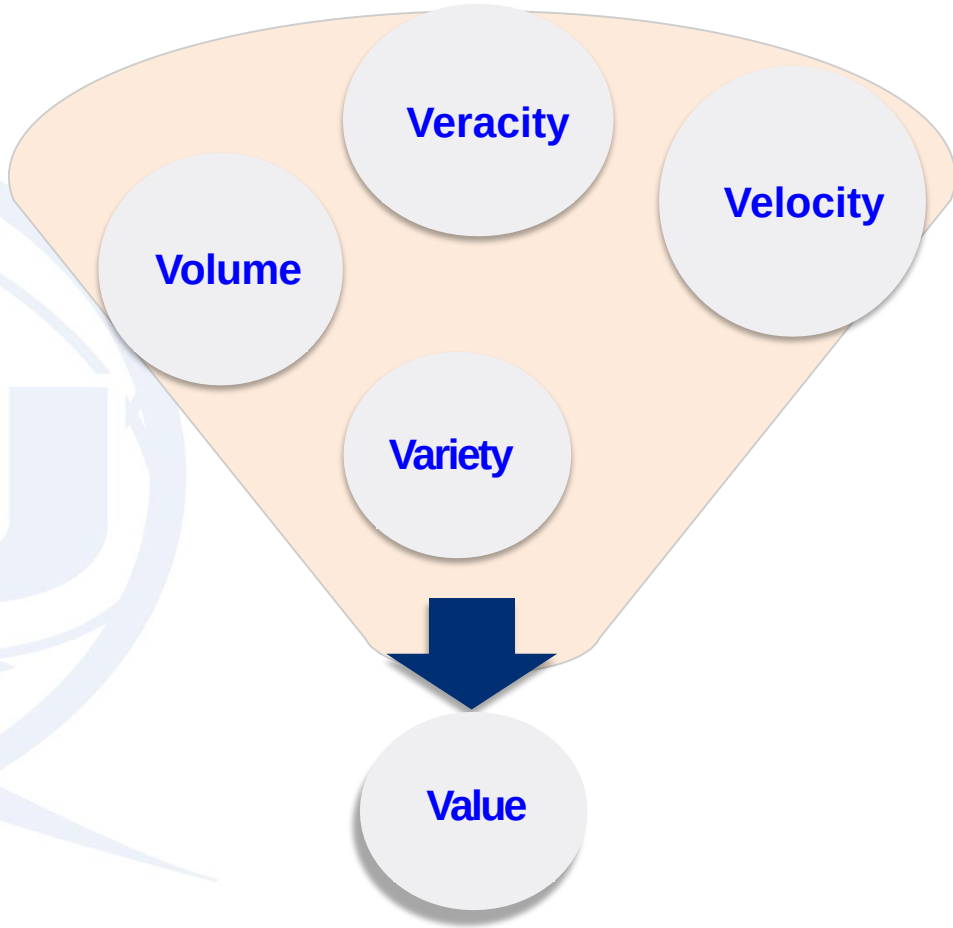
# Data everywhere: how «Big» is Big Data



# Big Data characteristics

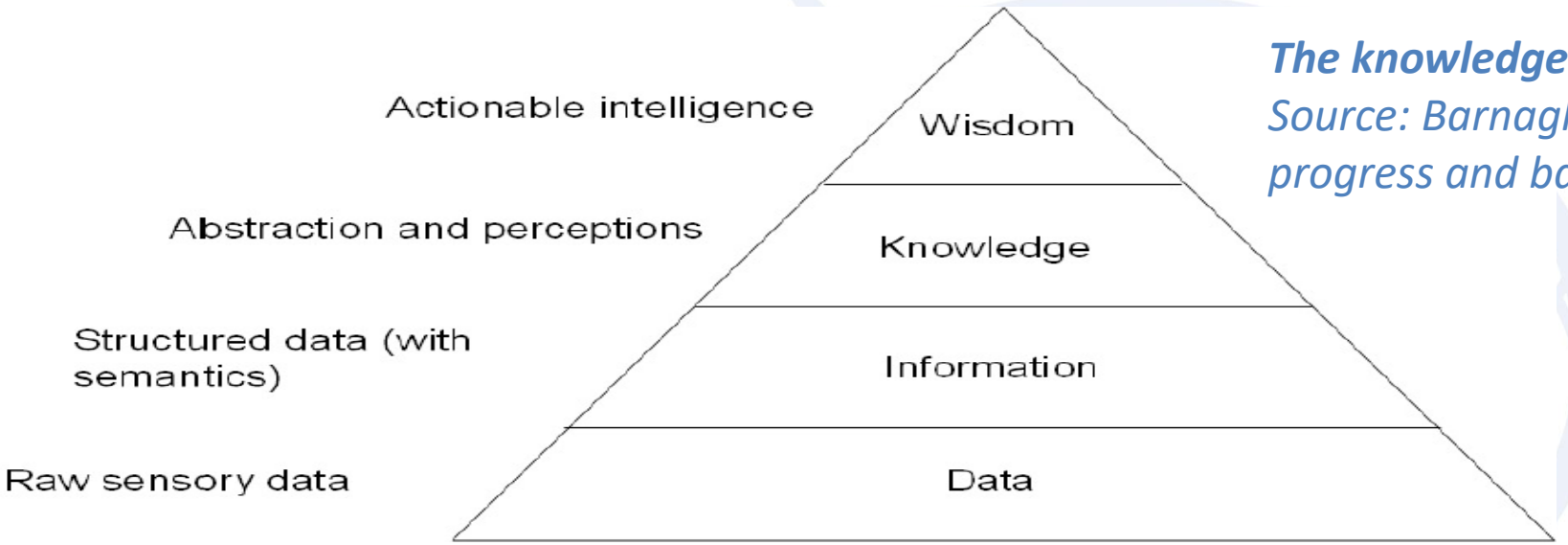
| Volume | Velocity | Variety | Veracity | Value |
|---|---|---|---|---|
|  |  |  |  |  |
| **Data at Rest** | **Data in Motion** | **Data in Many Forms** | **Data in Doubt** | **Data into Money** |
| Terabytes to exabytes of existing data to process | Streaming data, requiring mseconds to respond | Structured, unstructured, text, multimedia, … | Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, manipulation | Business models can be associated to the data |




**big data** [ITU-T Y.3600]: A paradigm for enabling the collection, storage, management, analysis and visualization, potentially under real-time constraints, of extensive datasets with heterogeneous characteristics.

NOTE – Examples of dataset characteristics include high-volume, high-velocity, high-variety, etc.

# Data Transformation: from Raw Data to Actionable Intelligence

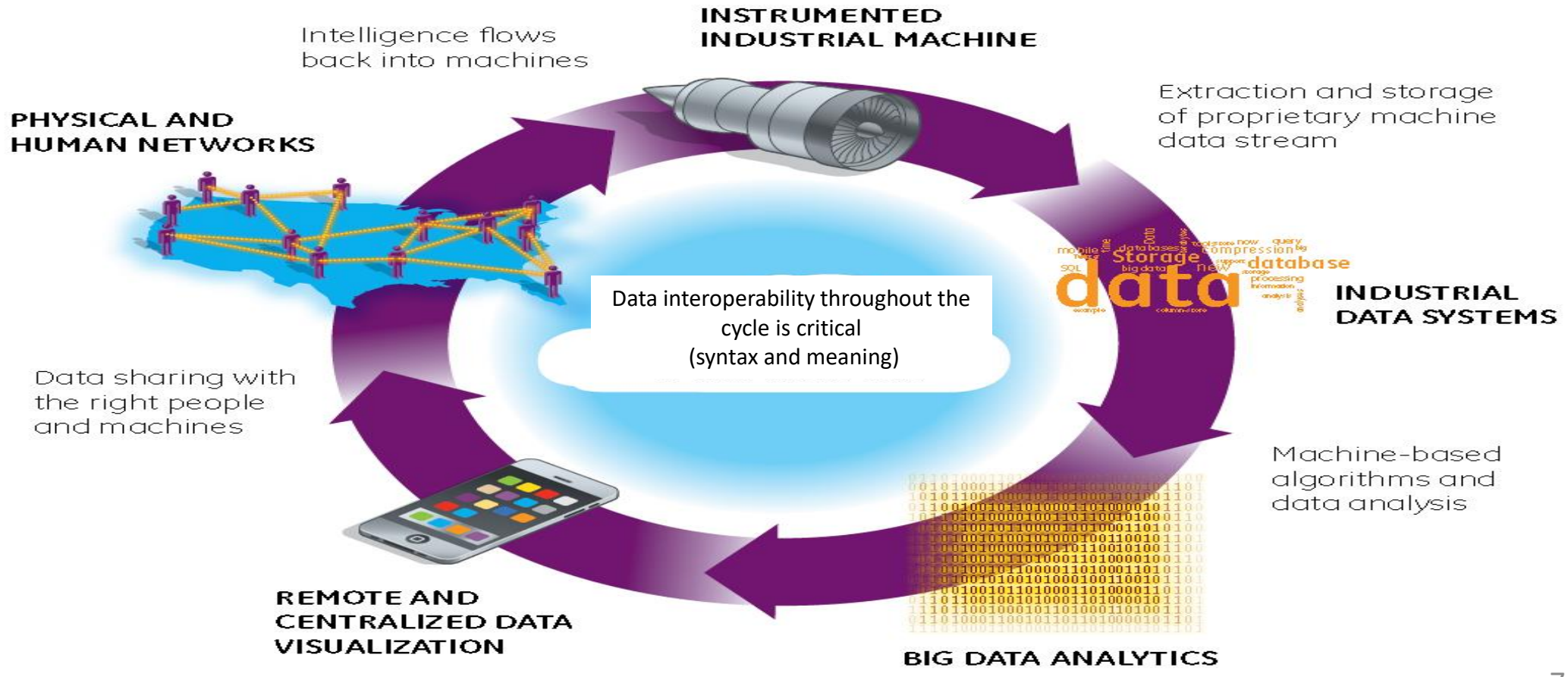



***The knowledge hierarchy applied in data processing***
*Source: Barnaghi and al., "Semantics for the IoT: early progress and back to the future" (IJSWIS, 2012)*

- Raw data are generated - as an example, by things (and more) in IoT
- Additional information enables creation of structured metadata (first step of data enrichment)
- Abstraction and perceptions give detailed insights of data by reasoning , using knowledge (ontologies, rules) of relevant domains (second step of data enrichment)
- Actionable intelligence allows decision making

# Key goal : to have ready for use the Right Data, at the Right Time, at the Right Location

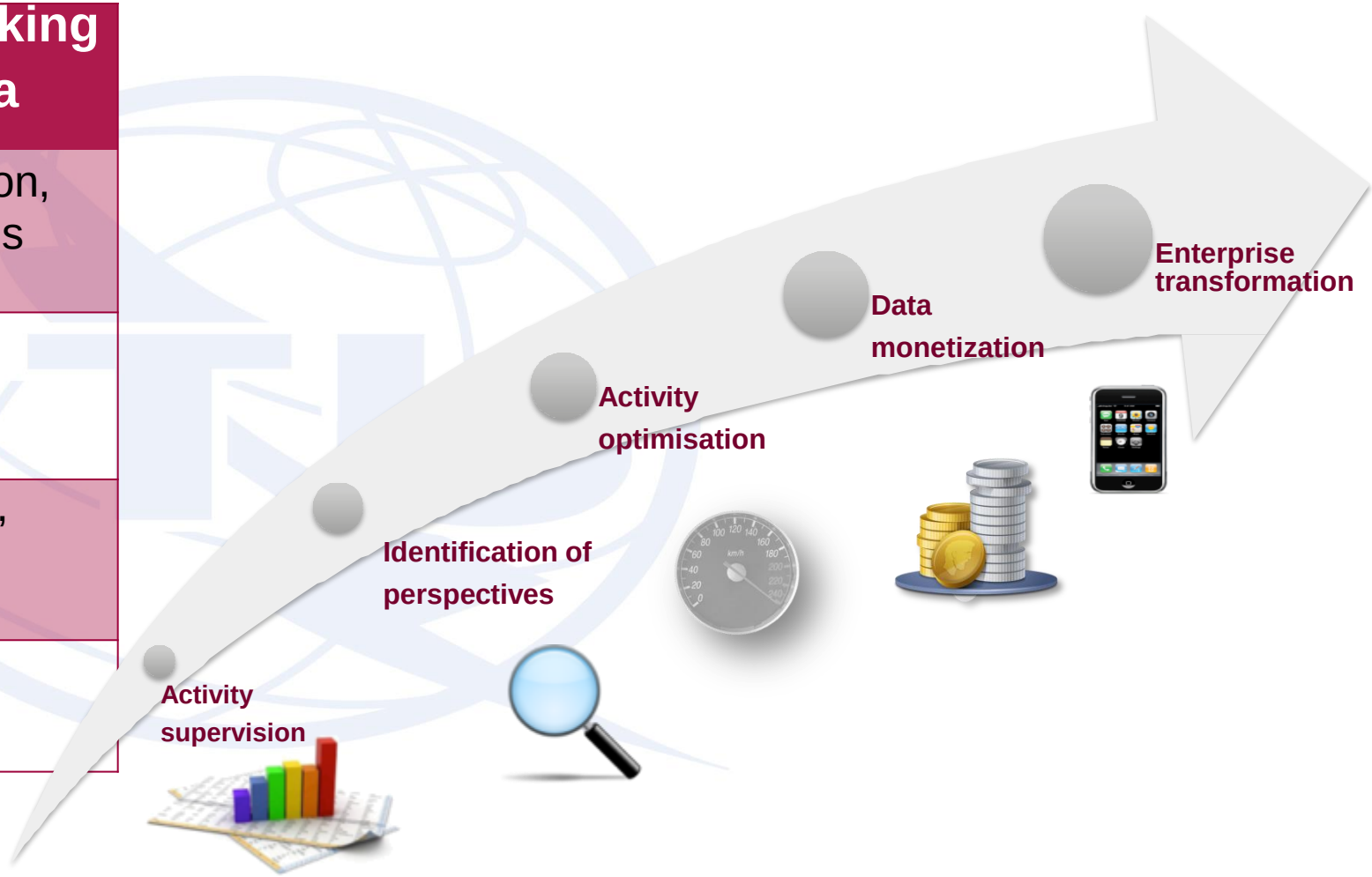
**The Industrial Internet Data loop** *[source: GE whitepaper]*



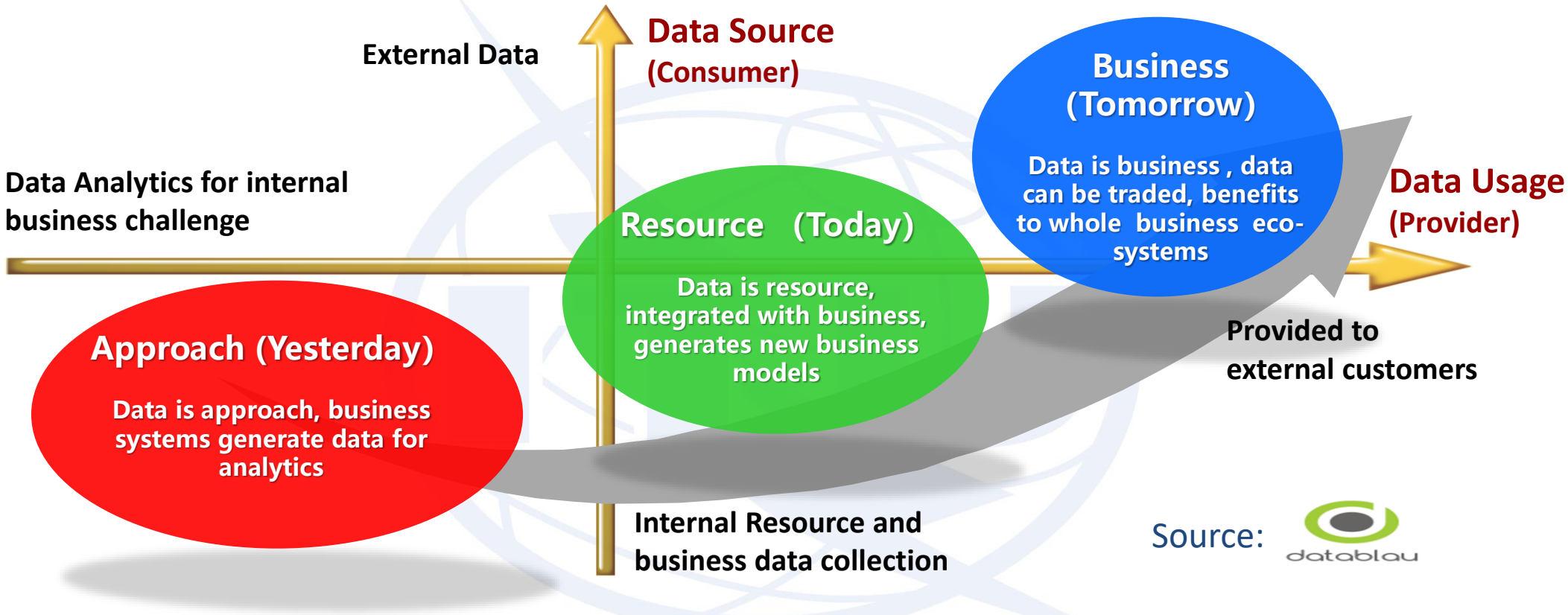
# Impact of Big Data on Enterprises

| Decision making without Big Data | Decision making with Big Data |
|---|---|
| Retroprospective vision | Prospective vision, recommendations |
| Less than 10% data are available | All sources of data can be exploited |
| Data in lots, disjoint, incomplete | Data in real time, correlated |
| **Supervision** | **Optimisation** |

*Source :Bill Schmarzo (CTO Dell EMC Services) Big Data*



# Evolution of Data Usage and Source




Three stages of big data: internal data analytics, data driven business innovation, data benefits & market

# Smart Cities: an incremental and participatory journey towards full support to Data Economy

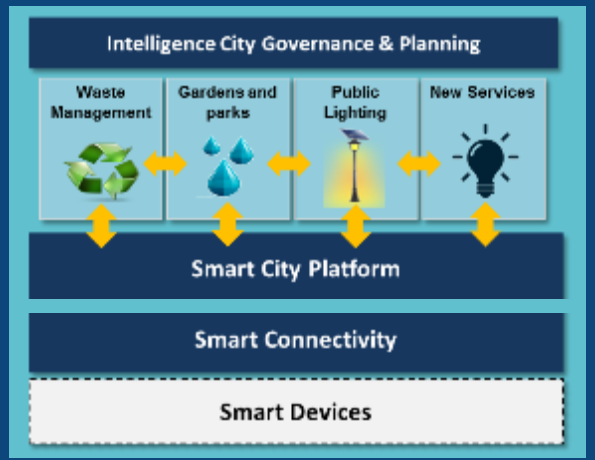
1 2 3 4

## Efficient and Open

- Vertical solutions bringing efficiency in silos
- Historic data as open data
- Information still in vertical silos, no global picture


## Truly Smart

- Horizontal platform integrating “right-time” context info from different vertical services
- Predictive and prescriptive models




## Unleashing Right-time Open Data

- Right-time context info published to third parties
- Exchange of context info with systems from other domains

## Support to Data Economy

- City as a platform including also 3rd party data enabling innovative business models
- Open and commercial data enabling multi-side markets

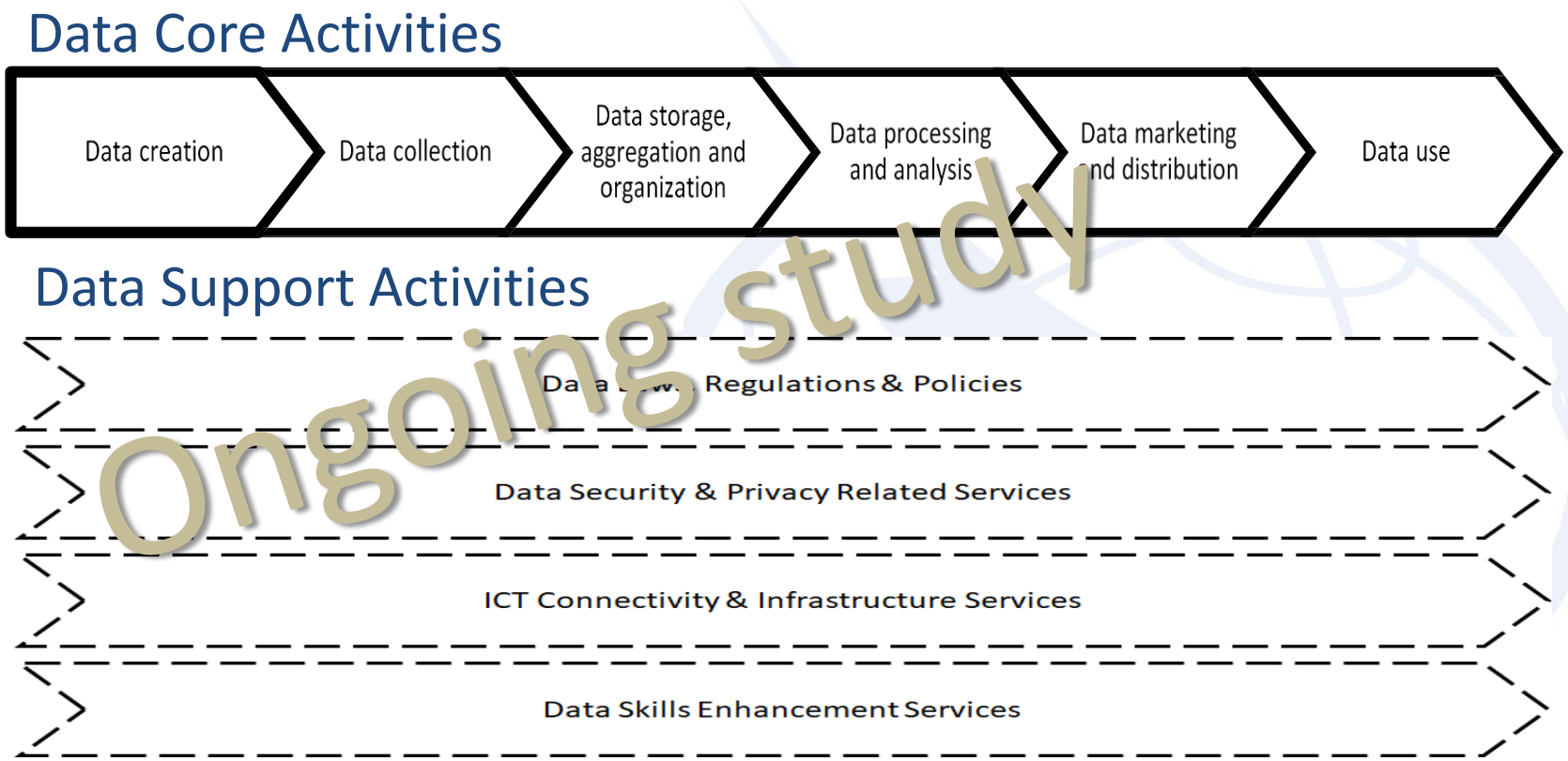

*Source:* FIWARE

# Business and social benefits from Big Data

## The Big Data Value Association vision for Europe in 2020

www.bdva.eu

# Data economy, commercialization and monetization




## Data Value Chain

**(business perspective)**

*Source : ITU-T FG-DPM*

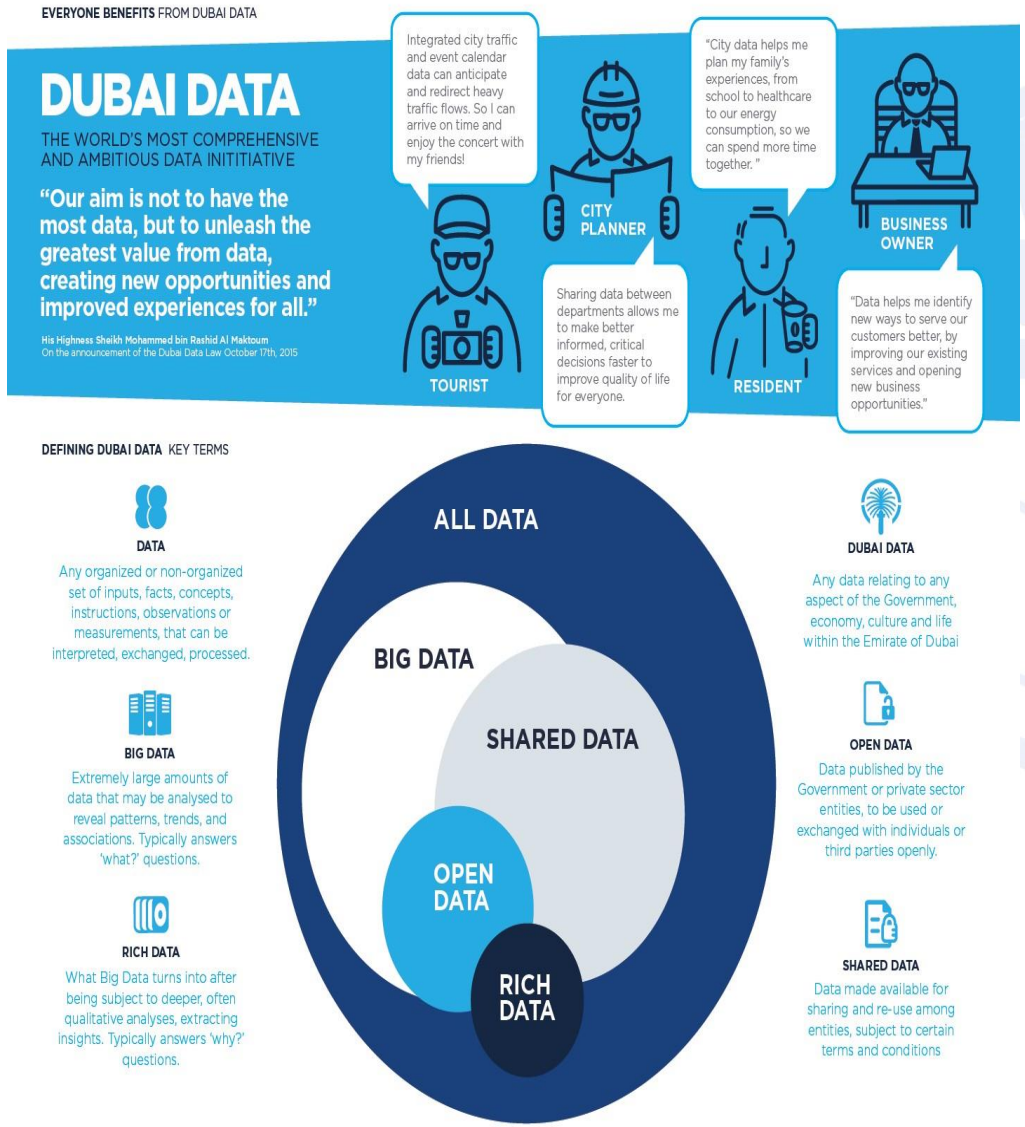
**Data laws, regulations and policies**: formulation and enforcement of data related laws, regulations and policies

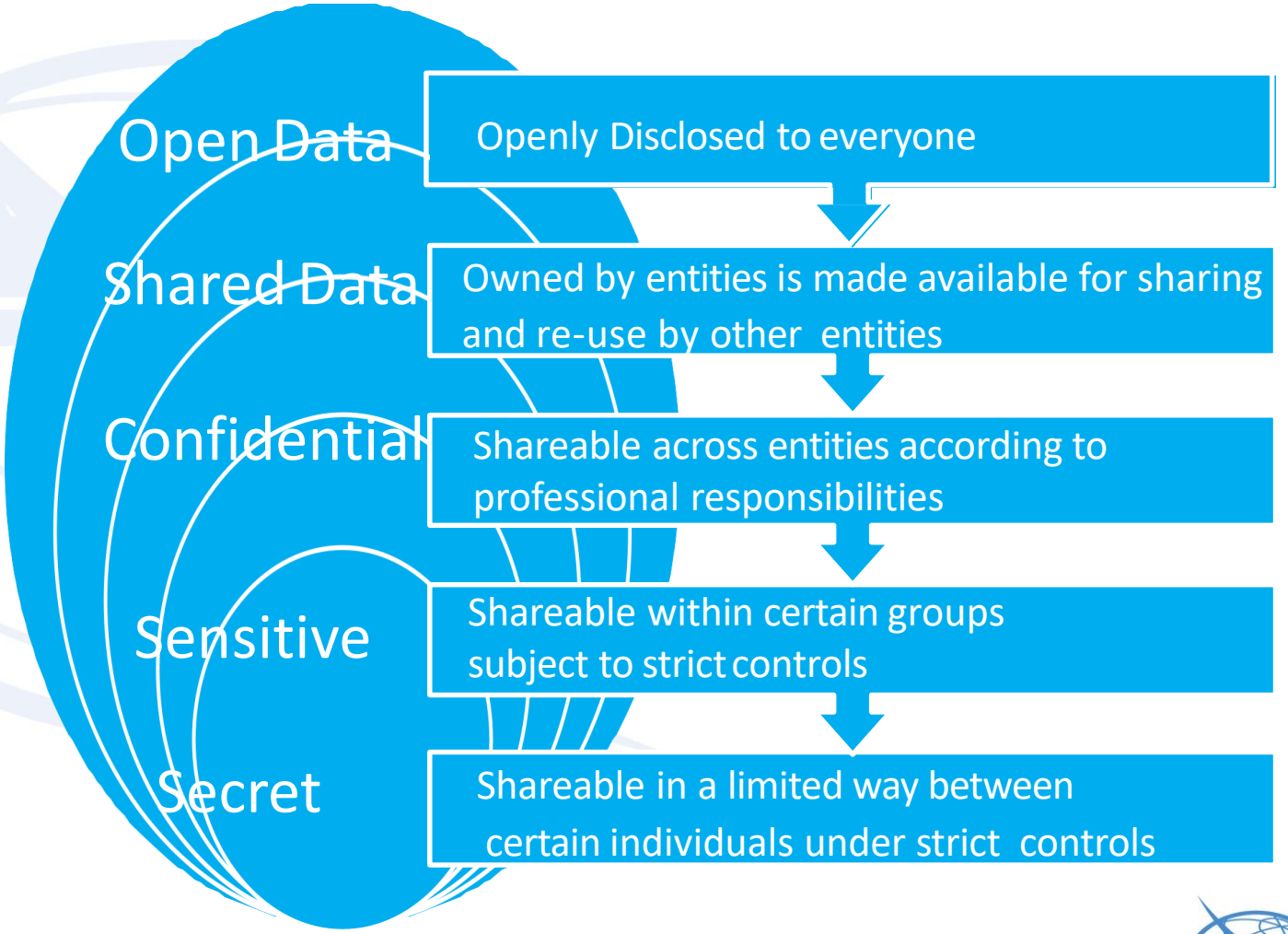
**Data security and privacy services**: provisioning of data related security and privacy services for implementing and enforcing data laws, regulations and policies.

**ICT connectivity and infrastructure services**: provisioning of ICT connectivity and infrastructure services for implementing data value chain activities

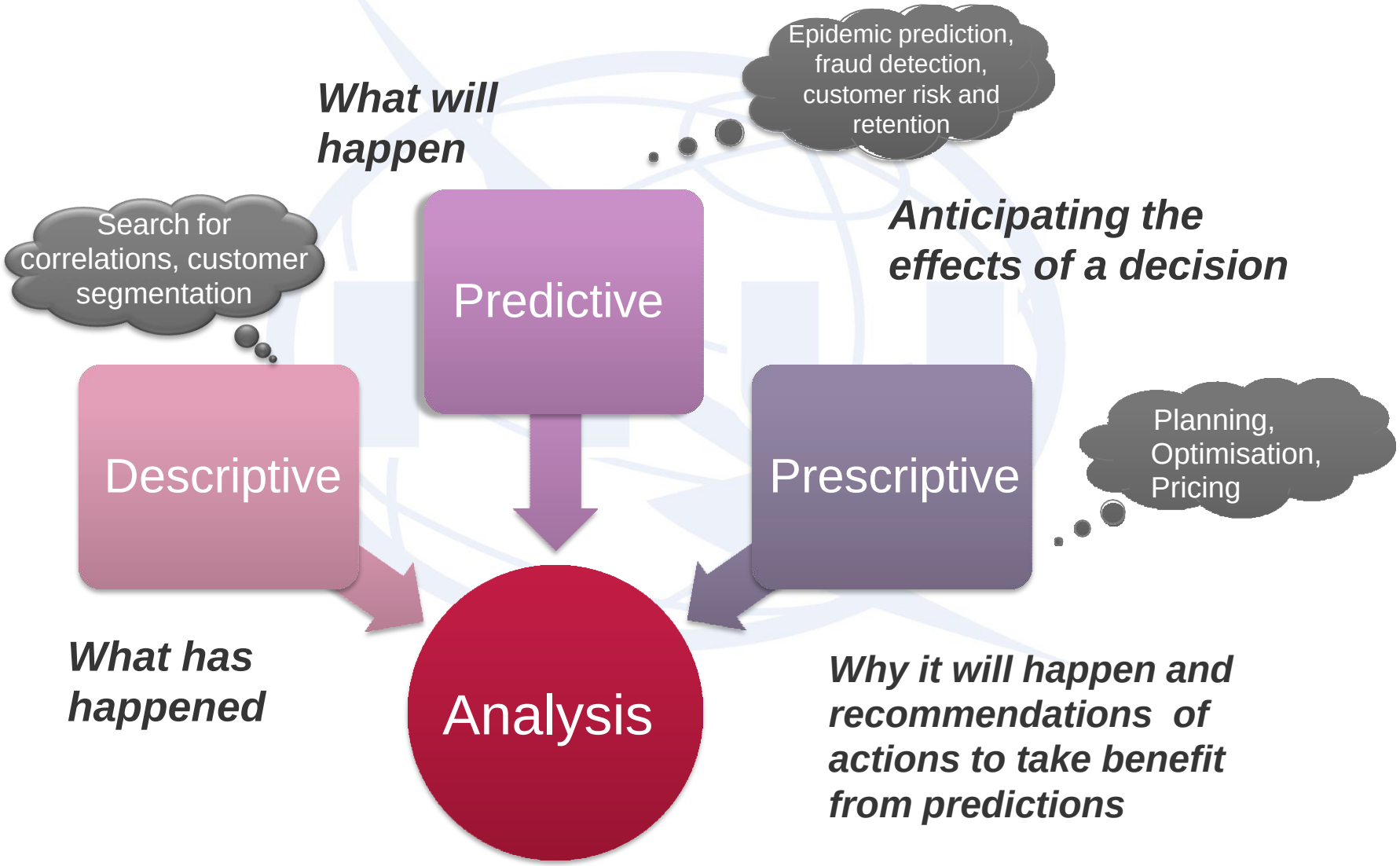
**Data skills enhancement services:** provisioning of data skills enhancement services, e.g. data related programs, degrees, certificates, courses

# Data classification: Big Data, Open Data and other (view from Dubai Data Initiative)



# Data Analytics techniques



# Application Domains of Big Data - few examples

# Telecommunication networks and services

**Process optimization and data monetization via analytics - driving revenue by sharing, analysing and interpreting data, for multiple purposes**

- Extraction of tangible business and technology value
- Real time response and action, improving productivity/business processes, lowering costs
- Long-range forecasts enabling strategic actions - business differentiation
- New/improved business models and service offer, faster, more efficiently and agile

# Vertical industries – example of monitoring and predicting plant failures

## Customer's challenge

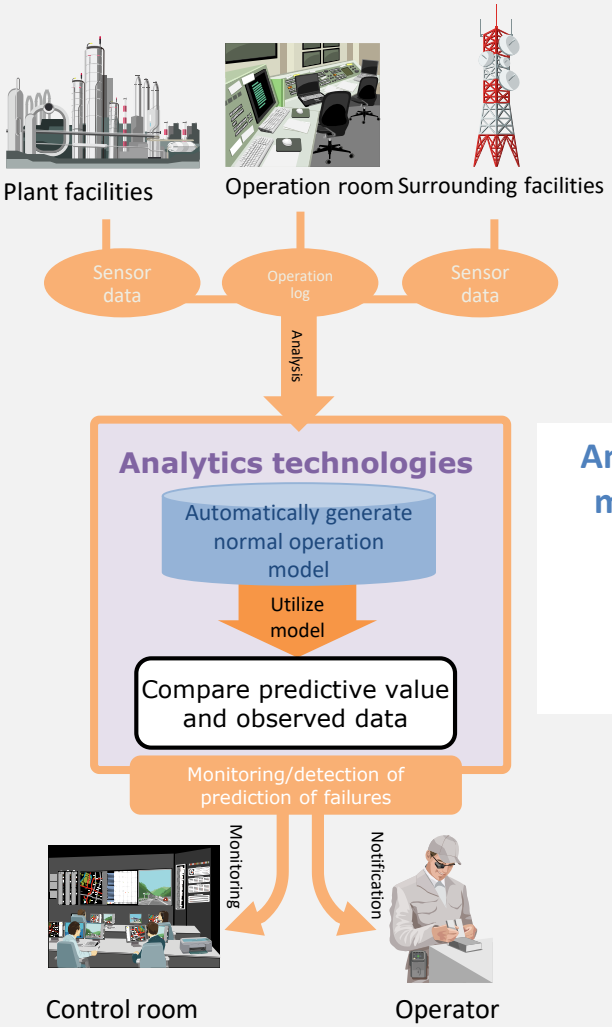
- Avoid damages by predicting failures, shorten the lead time to identify the cause of failure

## Effect of solution

- Monitor/detect prediction of failures of plant facilities
- Detect abnormalities from large volume of sensor data at an early stage, avoid large-scale damage before it happens

## Point of introduction

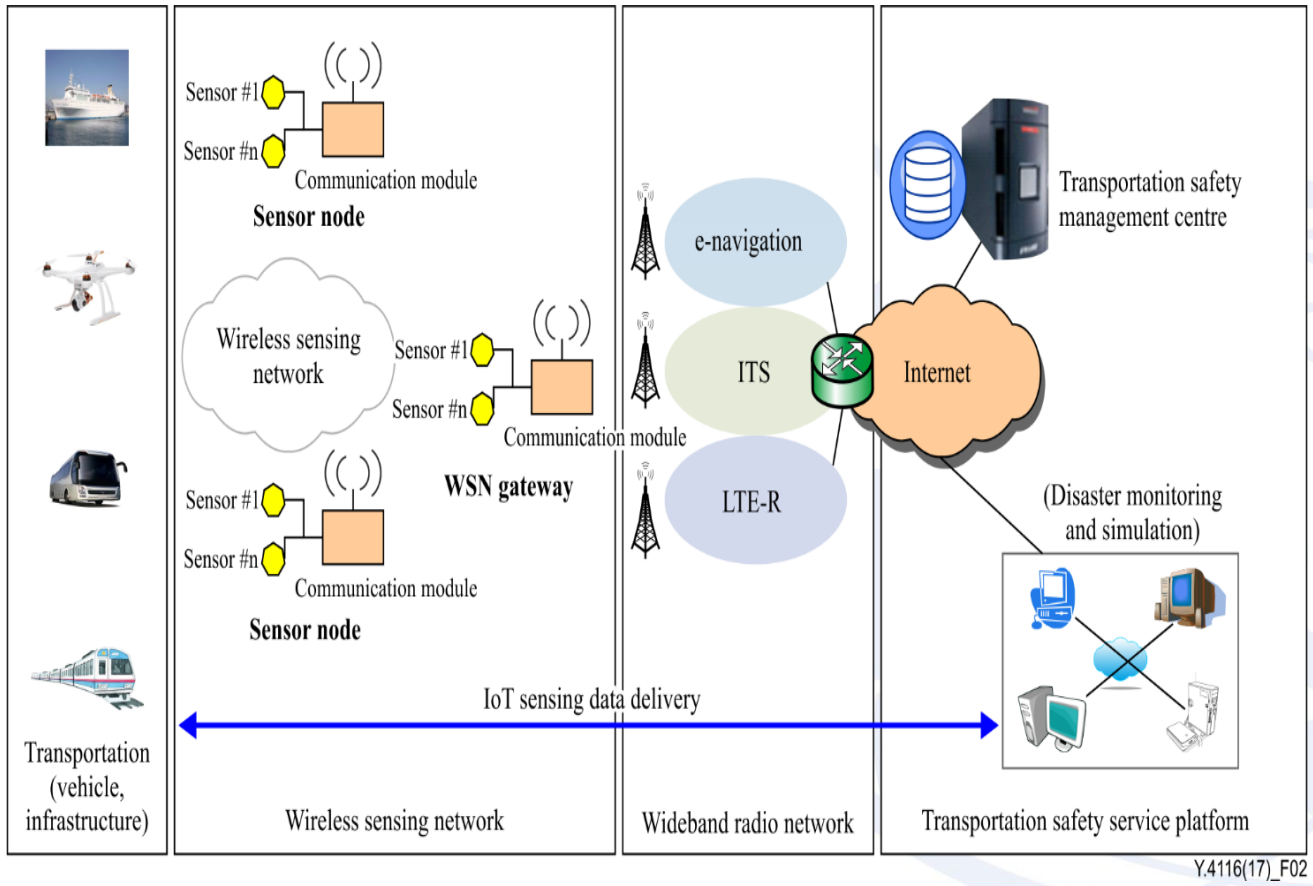
- **Visualize operational status from existing data by Analytics technologies**
- **Utilize massive data in real time and realize high accuracy monitoring/detection of failure prediction**

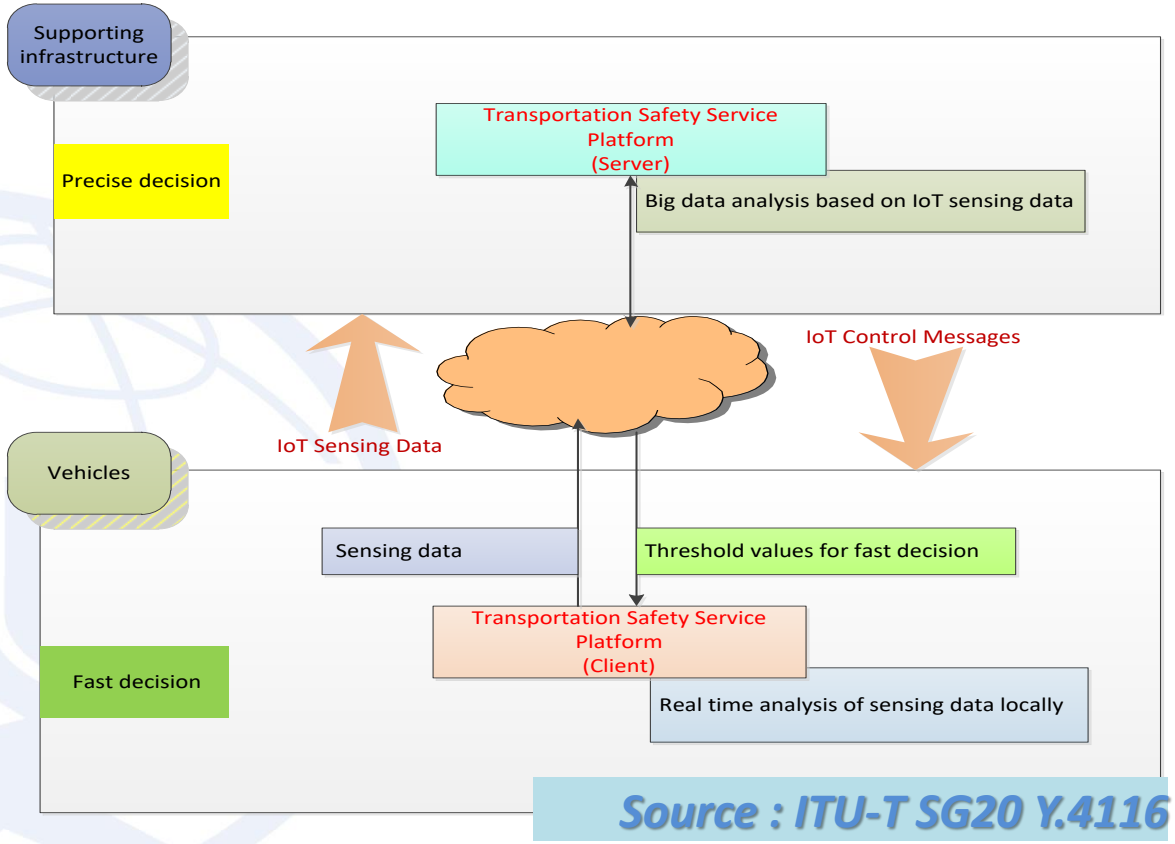



**Analysis of large amounts of metric data collected from multiple sensors to automatically identify relationships and detect anomalies**

*Analytics Technologies example from NEC*

# Big Data distributed processing for Transportation Safety




**Transportation service platform** monitors transportation safety relevant conditions and parameters, performs disaster simulations and decides the threshold values for disaster prediction and detection

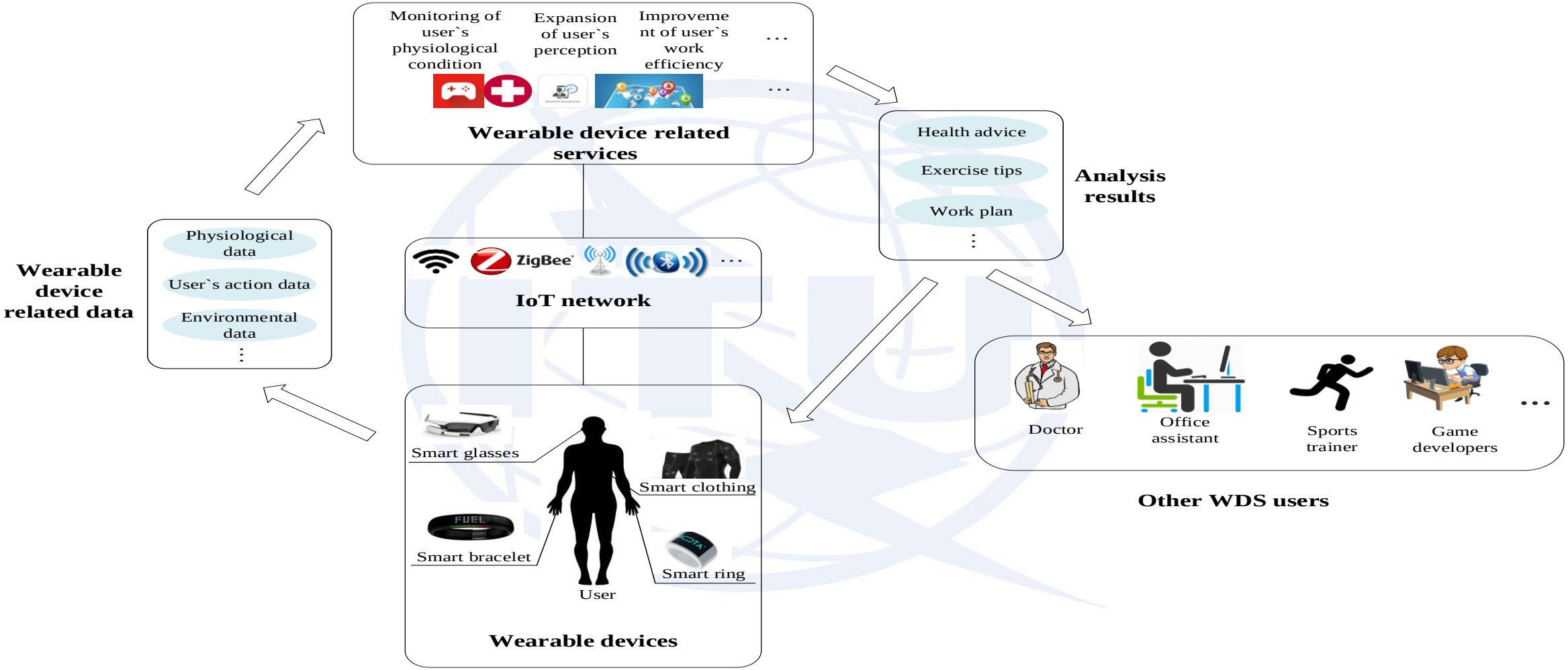
**Transportation safety management centre** monitors safety status of vehicles and transportation infrastructure, and influences operations of vehicles and infrastructure, by collaborating with transportation safety service platform, including generation of alarms

**Vehicles locally process and compare sensing data to threshold values for fast decision**. Sensing data from vehicles and transportation infrastructure are delivered to the transportation safety service platform (server side).

**The platform generates threshold values (e.g. safety indexes) for more accurate decision based on big data analysis.**

The generated threshold values are delivered to vehicles for appropriate adjustment of the local decision making process.

# Consumer applications: wearables




*Source: ITU-T Y.4117 «Requirements and capabilities of IoT for support of wearable devices and related services»*

# Big Data Technologies - highlights

# Some Big Data challenges

- Dealing with the “V”s of data : Volume, Variety, Velocity, Veracity
- Discovery (devices and data sources), integration (heterogeneous devices, networks and data)
- Scalability (number of devices, diverse and huge data, computational complexity of data interpretation)
- Availability and (open) access to data, data query
- Interpretation (extraction of actionable intelligence from data)
- Massive data mining, efficient processing
- **Trust, security and privacy of data (technical and non-technical)**
- **Other non-technical challenges are also essential, including data governance and ownership**

# Some critical deployment aspects

- **Data security, privacy and trust, extended and lawful data access, data liability (policies) and regulatory compliance**
  - **Definitely, these issues are not only technical**
- Dealing with the multi-dimensional challenge of Analytics (data at rest versus data in motion, and the related data cycle operations)
- Large spectrum of evolving technologies and products (e.g. which tools (identification of needs), which adequate evolution, deployment costs), organizational impact, skilled personnel
- System performances and reliability
- Best practices on integration and interoperability with legacy environments and applications

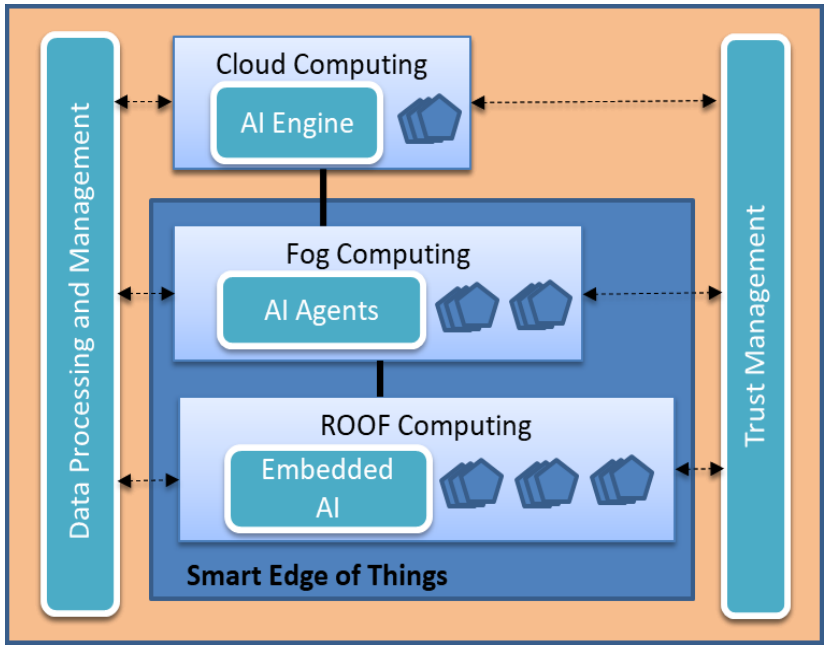
# Scalable Big Data processing and management solutions for stream processing, including video, are not yet there

Pure centralized cloud solutions will not be able to scale for continuous and timely processing of growing amounts of real-time streams

Solutions **mixing edge and central cloud processing** with high performance computing capabilities are required

- Faster response to emergencies
- Improved decision-making
- Increased operational efficiency

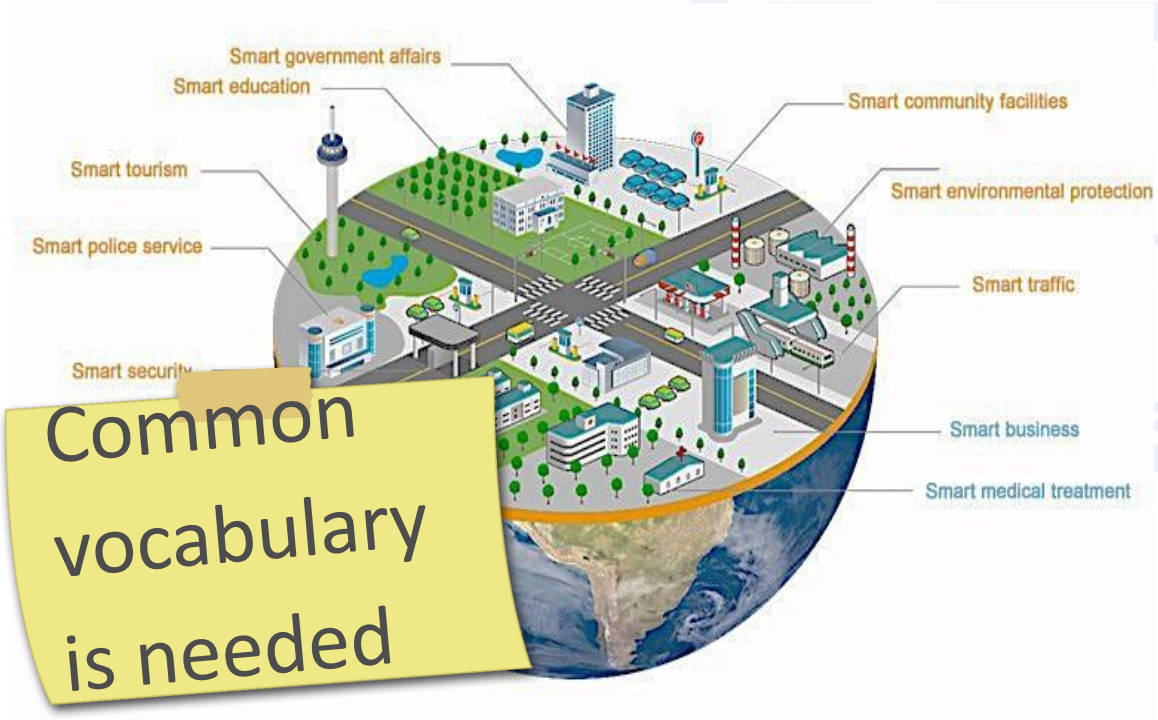
***Distributed computing: Edge Computing and more (Fog/ROOF/Device Computing)***

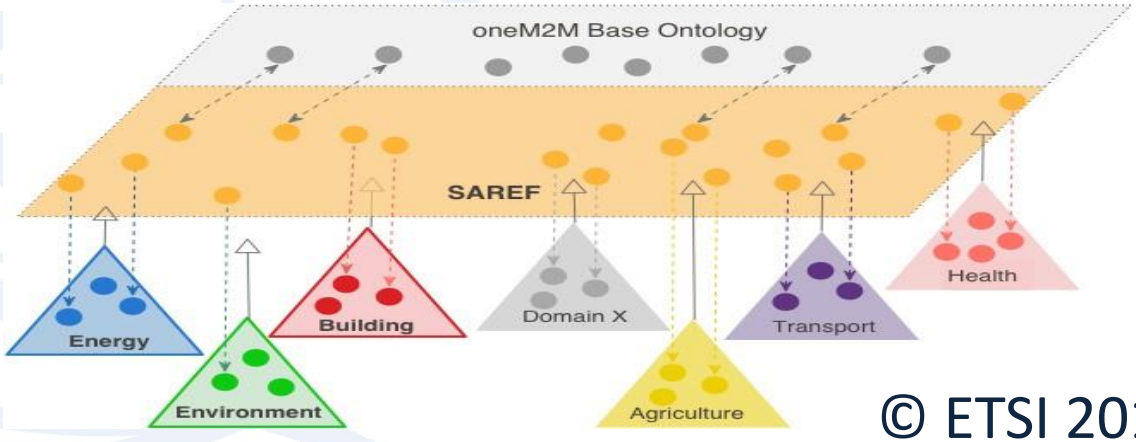


# Semantics based technologies and ontologies for semantic data integration: tools for intelligence enablement from (Big) Data

**Requirements for interoperability, scalability, consistency, discovery, reusability, composability, automatic operations, analysis and processing of data in telecom networks**

- **Semantics based approaches have outstanding features towards these requirements**



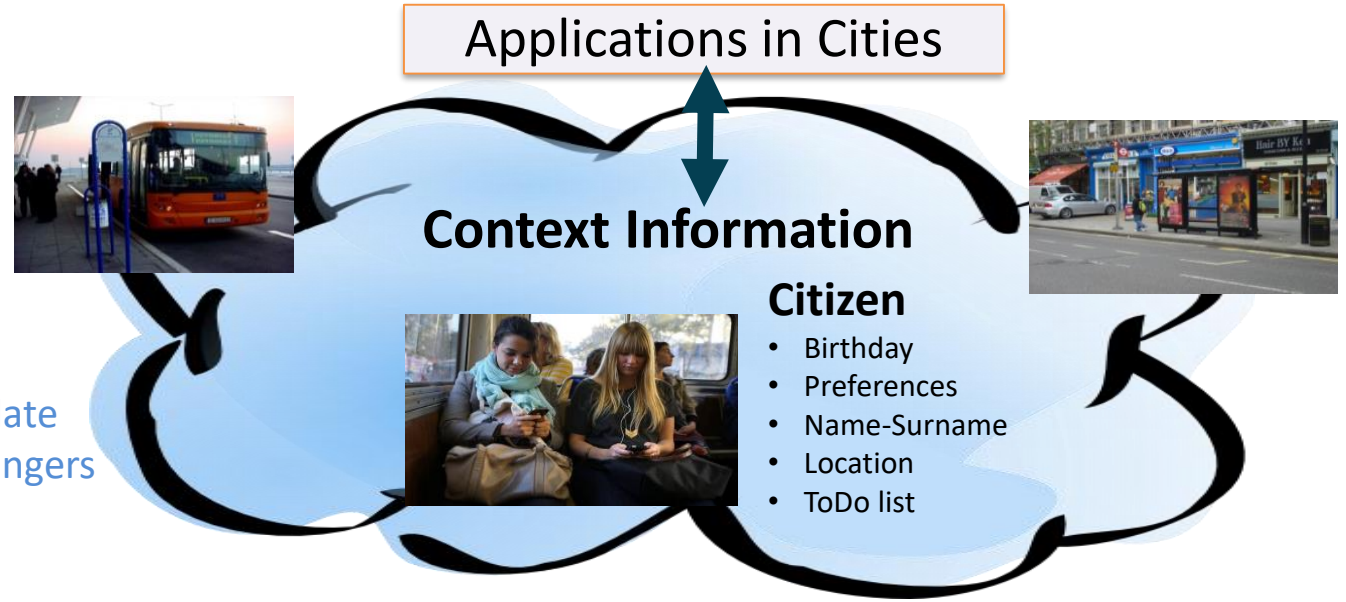






SAREF (Smart Appliances Reference Ontology) and ongoing SAREF extensions to better integrate semantic data from different vertical domains

# Technologies for Cross-Domain Data Sharing - context info management

**Cross-domain applications require access to information from different domains that is normally held in separate silos - e.g. they need to share context information**

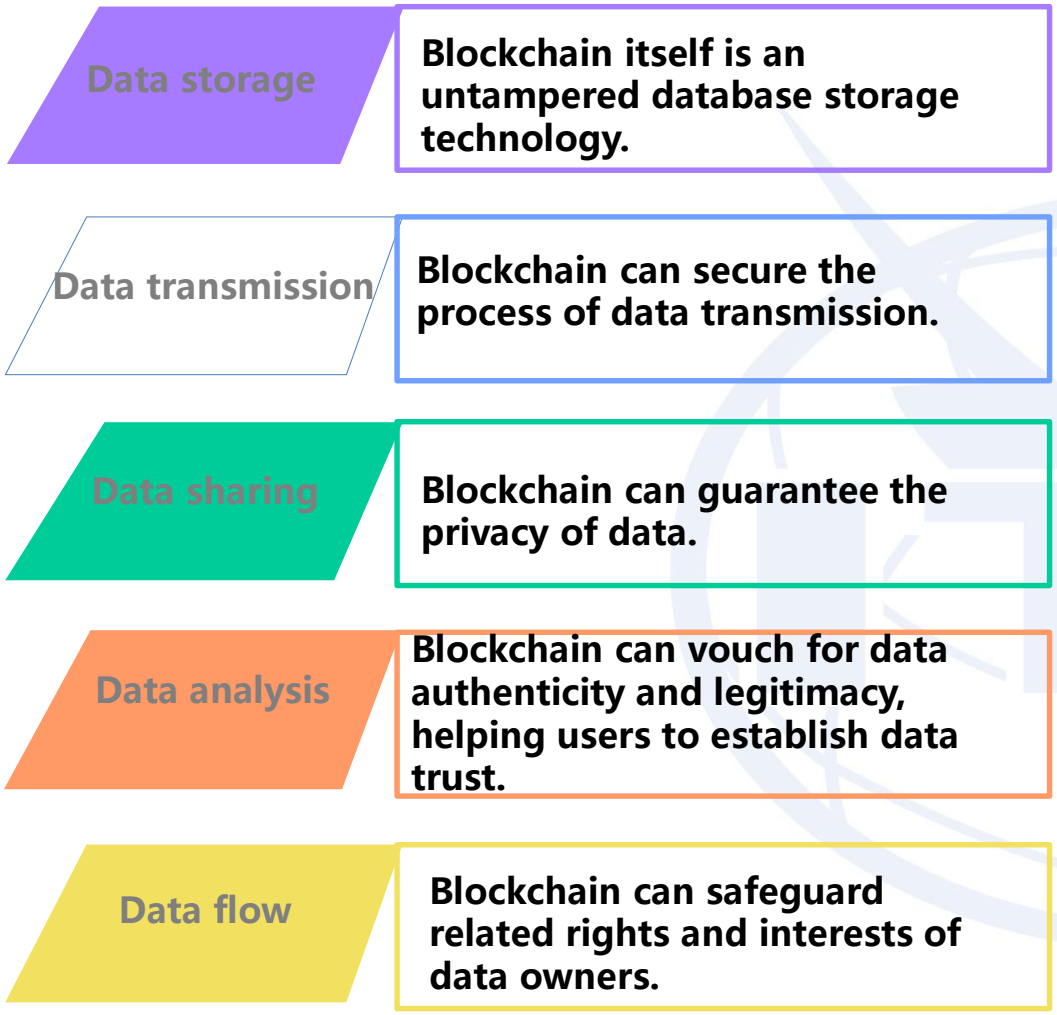



**Standardized solutions for management of context information are under work (ITU-T, ETSI, …)**

- To ensure vendor neutrality for users such as Cities
- To reduce technological barriers to development/deployment, to enable innovative services

**Promoting the "Network Effect" of Data**

# Blockchain for Big Data



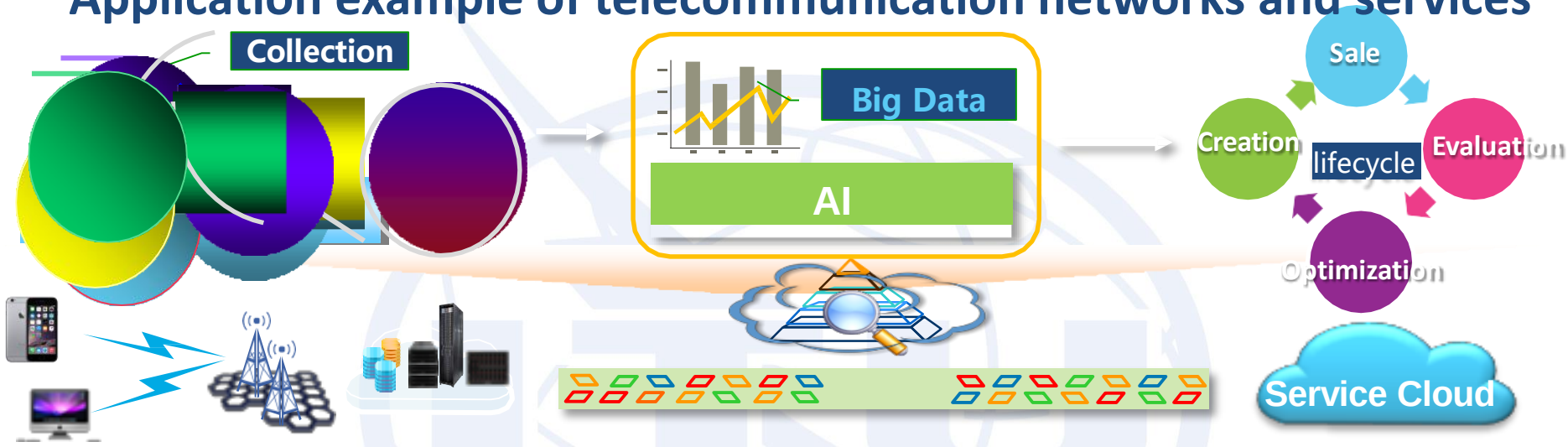

**Source: Tai Cloud Corporation**

- Blockchain and other Distributed Ledger Technologies are booming these days, with new technology approaches, uses cases and business models coming up regularly
- Still not completely prepared to support many real processes in today's economy
- Standardization is expected to play a relevant role in their application (but should not slow down innovation)
- ITU-T strongly involved in related standardization (incl. SG13, SG17, SG20, FG-DPM, FG DLT)

*Blockchain technology fosters a new generation of transactional applications that **establish trust, accountability, transparency and efficiency.** It shows great promises across a wide range of business applications in many fields. **However, its applicability needs to be evaluated according to the specific scenarios.***

# Augmented capabilities via Big Data + AI/ML integration

## Application example of telecommunication networks and services




**Enhanced network design, operation and optimization**

- coping with highly increased complexity
- enhancing efficiency and robustness of network operations (e.g. by reducing number of measurements and facilitating robust decisions)
- increasing network self-organization feasibility (cognitive network management)
- providing reliable predictions [pro-active strategies]

**Enablement of new advanced applications**

**Intelligent support of the management of the whole life cycle**

***But also challenges (including the general TRUST challenges of AI/ML) !***

# For an **Accountable and Ethical** Data Management and Analytics

- Opening the black box of Deep Learning
- Data provenance and usage monitoring
- Progressive user-centric analytics
- New paradigms for information flow monitoring
- Fact-cheking requiring explicit, verifiable argumentation integrating heterogeneous data sources and explainable reasoning

Source: N. Boujemaa, INRIA, BDVA Board member

***Marco's opinion: International coordination on Big Data and AI/ML standardization (for networks and services) is definitely necessary between ITU-T and relevant SDOs, Alliances, Consortia***

# “Personal Data” processing and management

***This is a very relevant topic at both technical and policy & regulation levels***
***Big Data technologies have to address this matter - research, solutions, …***

## *Personal data in GDPR*

**‘personal data’ means any information relating to an identified or identifiable natural person (‘data subject’)**

An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. *Source: GDPR*

*Article 4, Definitions*

***GDPR is the European Union’s “General Data Protection Regulation ” on usage and protection of personal data in EU - enforceable since 25 May 2018***

## *GDPR principles for personal data*

- Processed lawfully, fairly and in a transparent manner (“**Lawfullness, fairness and transparency**”)
- Collected for specified, explicit and legitimate purposes (“**Purpose limitation**”)
- Adequate, relevant and limited to what is necessary (“**Data minimization**”)
- Accurate and, where necessary, kept up to date (“**Accuracy**”)
- Identification no longer as necessary for the purposes (“**Storage limitation**”)
- Processed in an appropriate manner to maintain security (“**Integrity and confidentiality** ”)
- **Accountability** (documentation)

*Source: Austrian Data Protection Authority*

# Thank you very much for your attention



30

# Backup information - few highlights on Big Data Standardization

# Big Data technical standardization

## "Some" areas for standardization [ITU-T Y.Sup40 on Big Data standards roadmap and more]

- **Big Data architecture and APIs** (APIs with network infrastructure, users, auditors)
- **Flexible Analytics** (real-time, batch; remote, distributed and federated analytics; network-driven analytics)
- **Data access**, including Open Data frameworks and Data Governance within companies
- **Framework for Data quality and trust** (context dependent)
- **Framework and standards for Data Exchange** - data sharing, transaction, interconnection
- **Security and Data protection**, anonymization and de-identification of Personal Data (and reversibility)
- **Integration of Big Data requirements in central/distributed cloud computing solutions** for both infrastructure and services (interoperability, data/process security, traceability, personal data protection, SLAs, data storage)
- **Standards and guidelines to address issues for legal implications of (Big) Data** in telecom (e.g. Data ownership)
- **Benchmarks for system performance evaluation**
- **Standardized visualization methods**
- **Domain-specific languages**

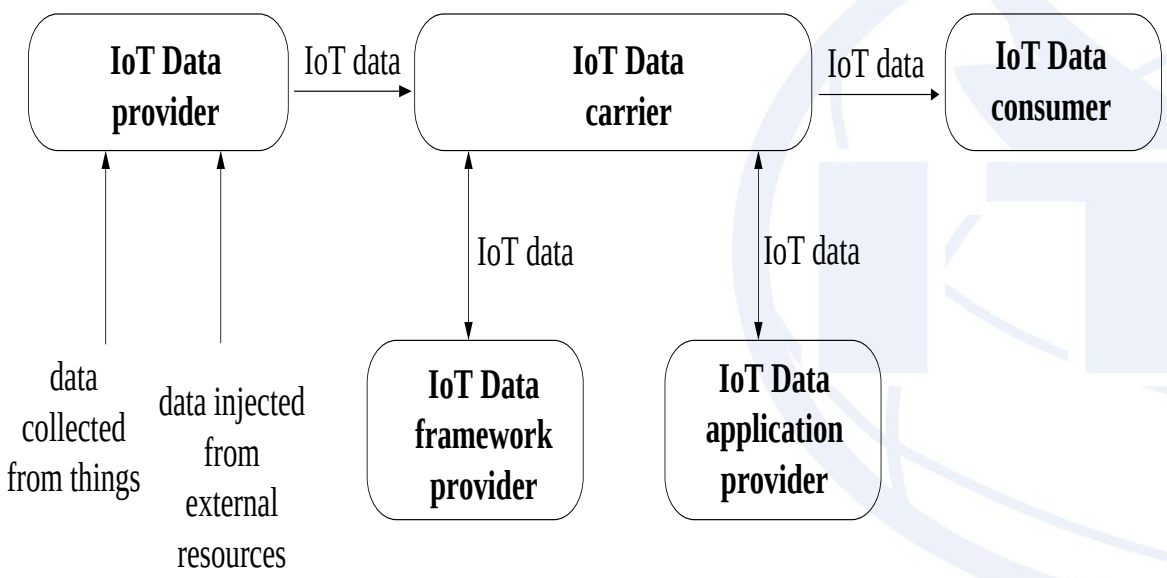
***Different standards initiatives are addressing the different technical Big Data areas***

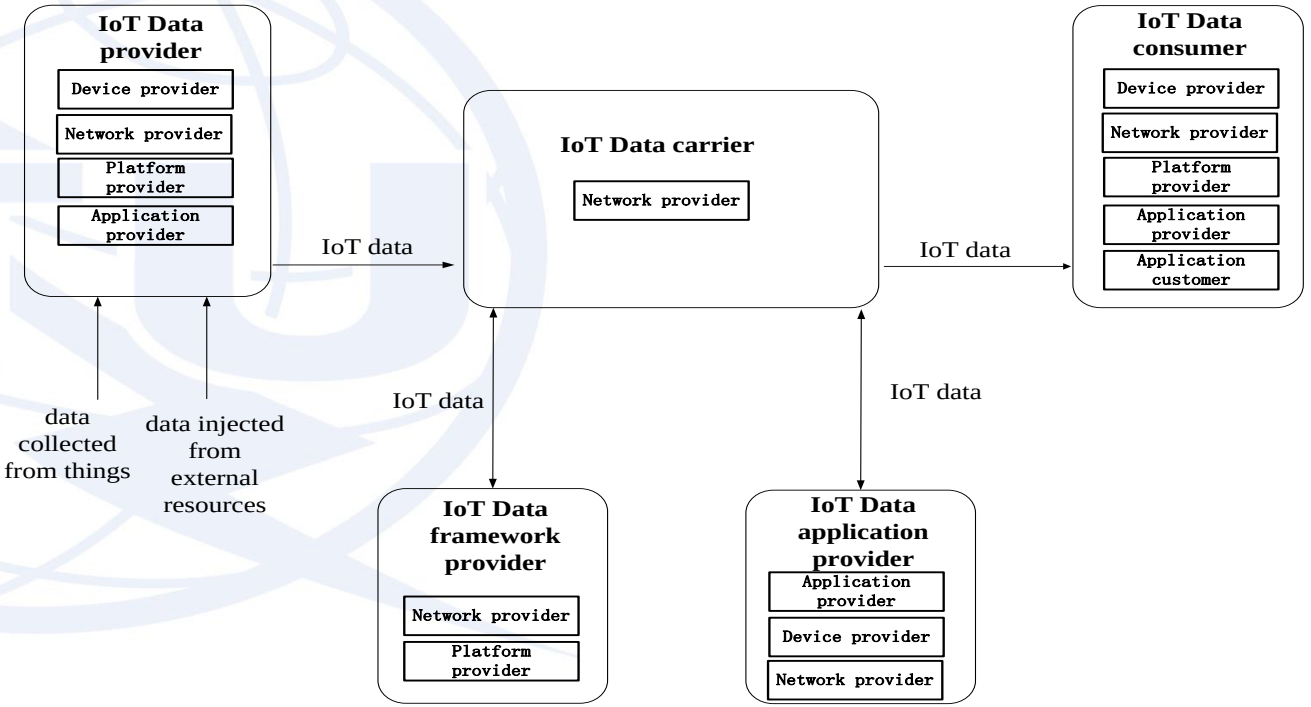
- *ITU-T [SG13 for non-IoT, SG20 and FG-DPM for IoT], ISO/IEC JTC1 SC42, BDVA, others*

# A foundational ITU-T Recommendation on Big Data in IoT:

## ITU-T SG20 Y.4114 "Specific requirements and capabilities of the IoT for Big Data"

**Specific requirements and capabilities the IoT is expected to support to address the challenges related to Big Data**



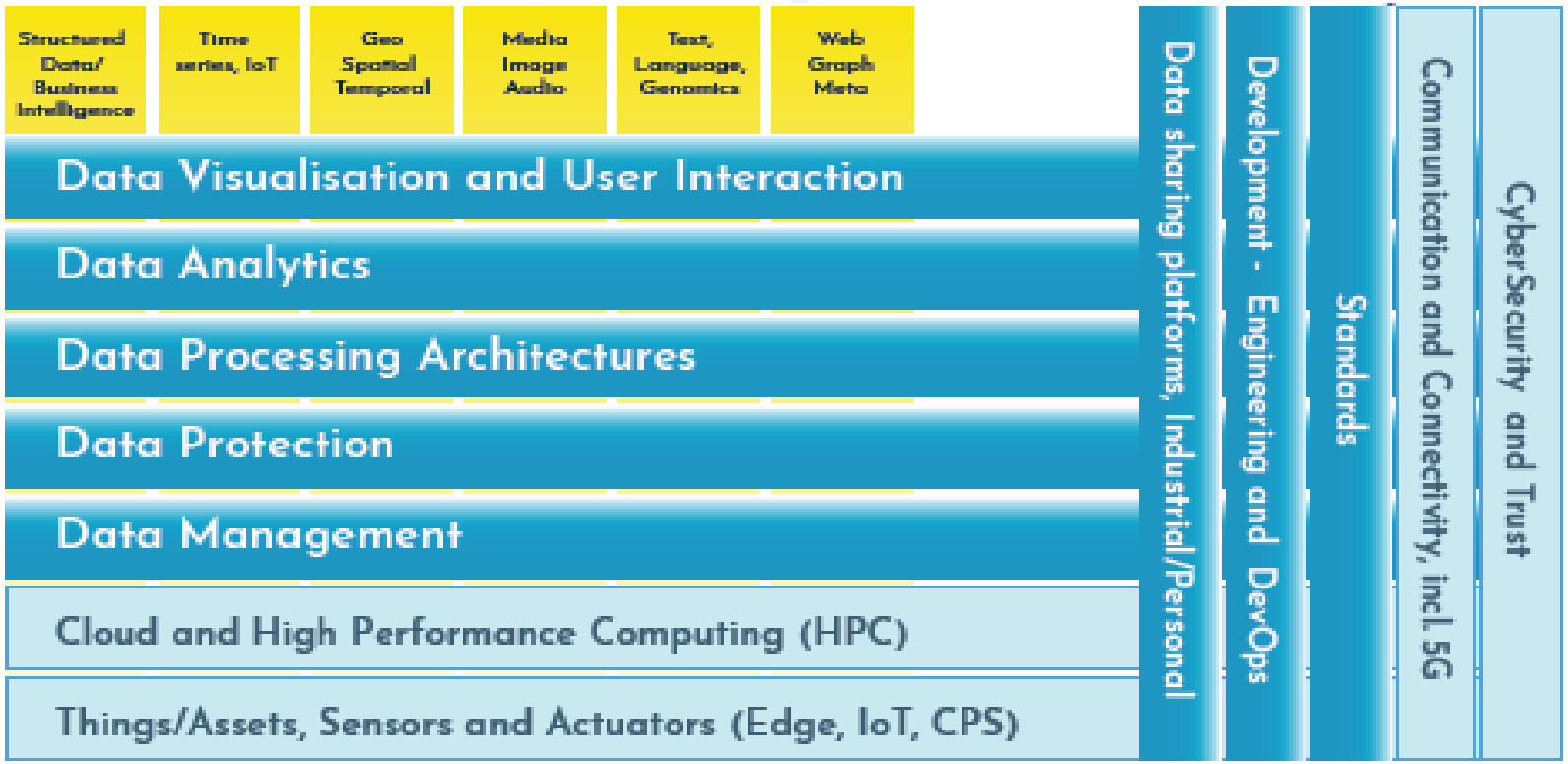




**The IoT data roles identified in Y.4114**

[the key roles relevant in an IoT deployment from a data operation perspective]

**Mappings of IoT business roles (ITU-T Y.2060) to IoT data roles**

# Big Data Value Association: BDV Reference Model




BDV RM is structured into technical areas (capabilities) – there is no layering connotation

***Source : BDVA***

**A key step in front of the IoT standardization work plan:**
**Big Data-IoT architectural integration**

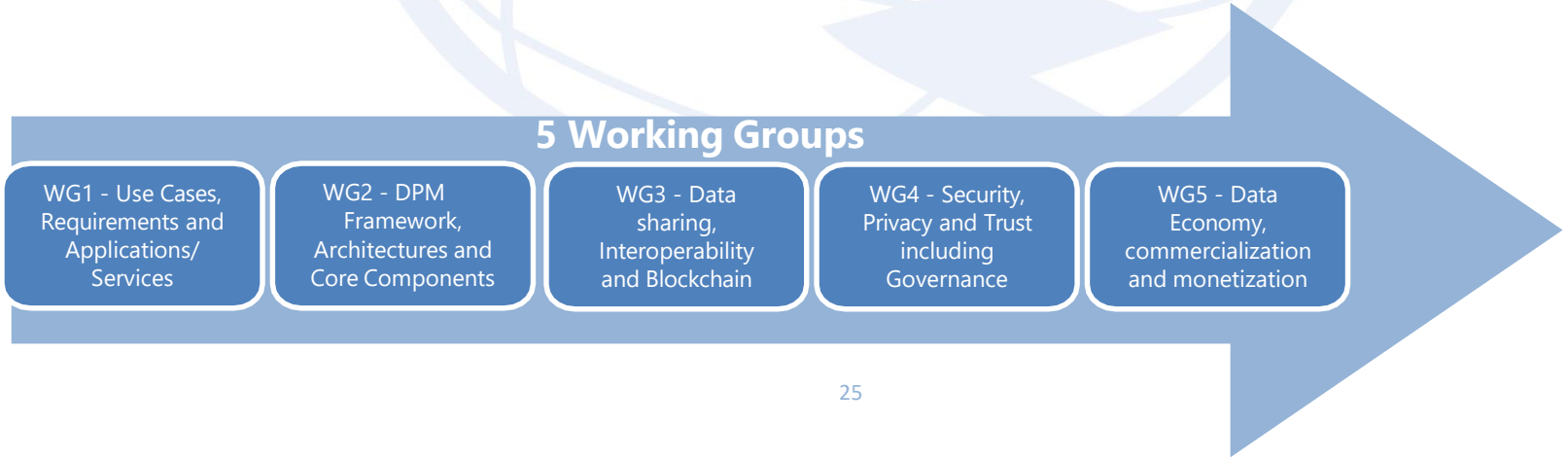
# ITU-T Focus Group on Data Processing and Management to support IoT and Smart Cities & Communities (ITU-T FG-DPM)

## *Essential tasks*

- Identify challenges in IoT and smart cities for DPM
- Identify key requirements ad capabilities for DPM
- Promote the establishment of trust-based data management frameworks for IoT and SC&C
- Investigate the role of emerging technologies to support data management incl. blockchain
- Identify and address standards gaps and challenges

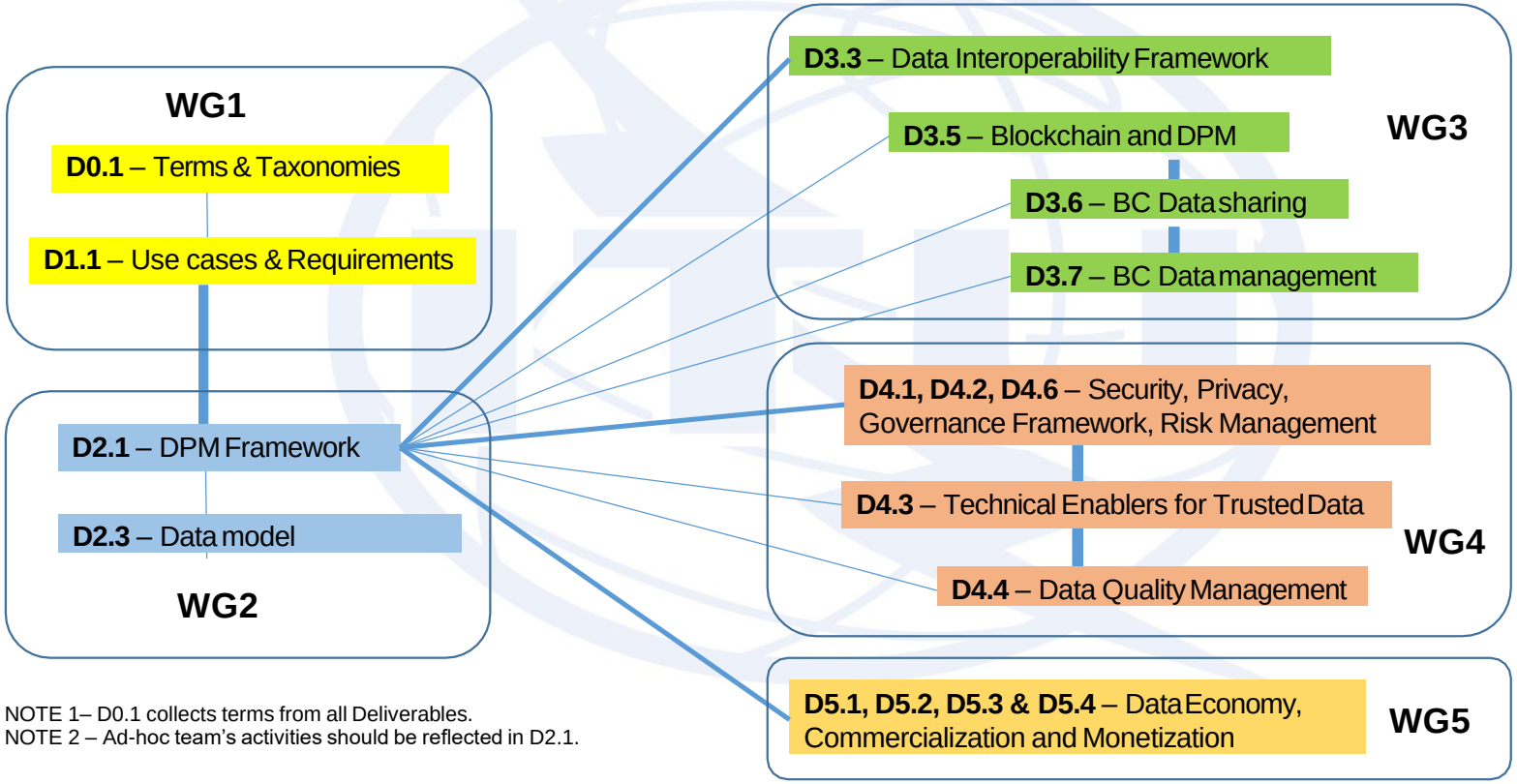
**1st meeting in July 2017 (SG20 as parent SG)**

**Last meeting is planned in July 2019**




25

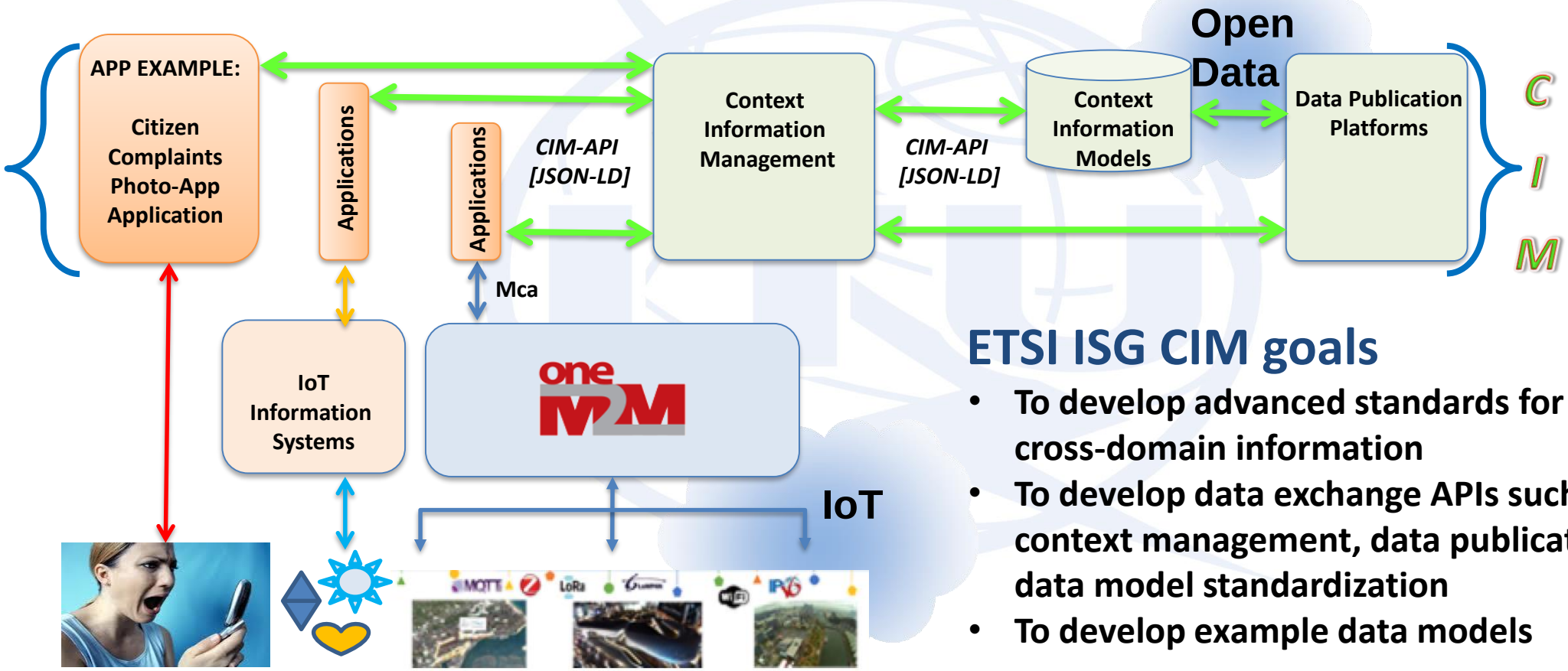
# ITU-T FG-DPM working deliverables




NOTE 1– D0.1 collects terms from all Deliverables.
NOTE 2 – Ad-hoc team's activities should be reflected in D2.1.

# Data sharing: cross-domain Context Information Layer in ETSI ISG CIM

An info-exchange layer on top of IoT platforms - especially targeting Smart City applications

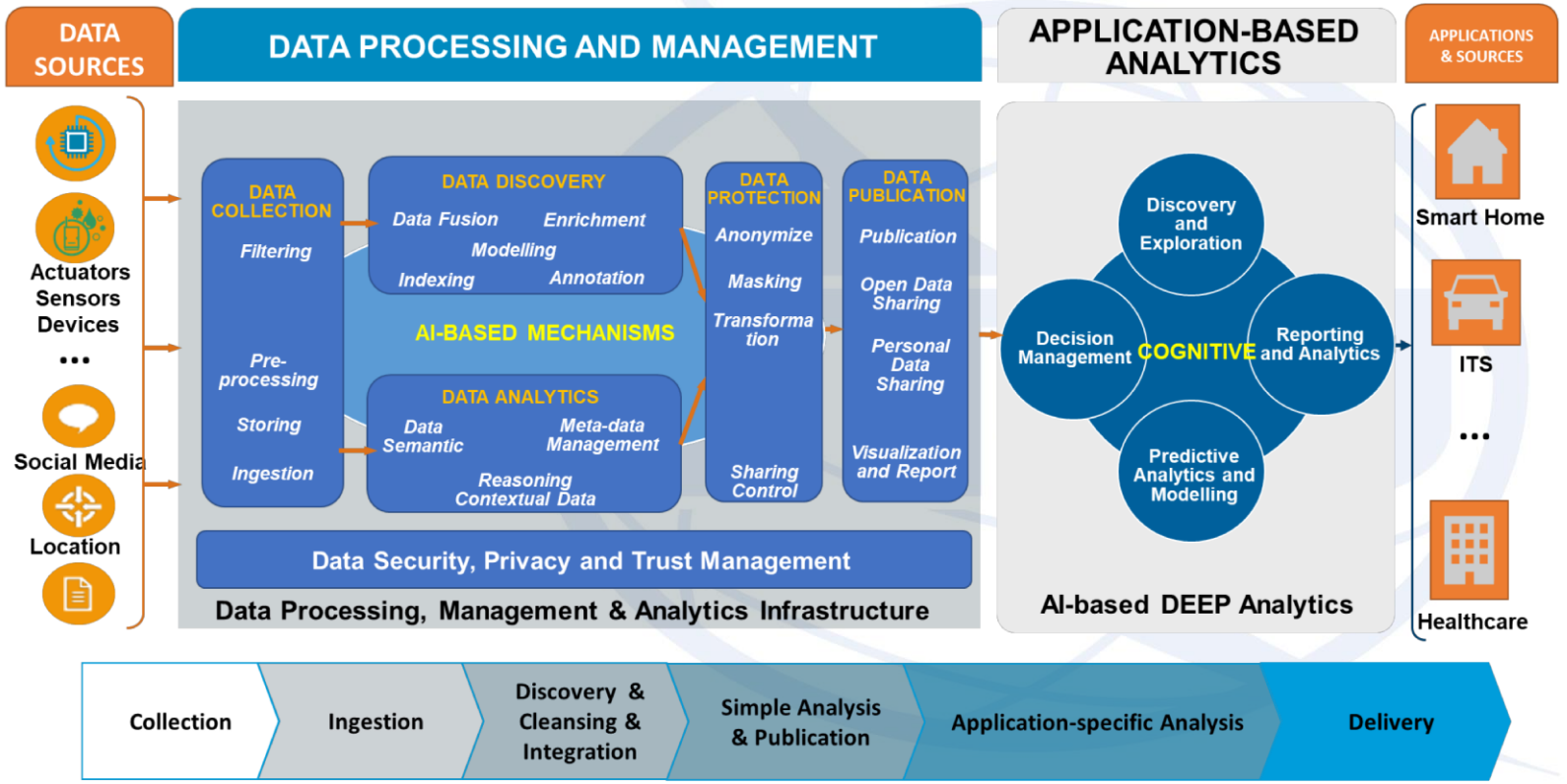



## ETSI ISG CIM goals

- **To develop advanced standards for exchanging cross-domain information**
- **To develop data exchange APIs such as APIs for context management, data publication platform, data model standardization**
- **To develop example data models**

Collaboration with ITU-T FG-DPM, oneM2M, W3C ..., and open-source implementations

# Towards Data-driven Artificial Intelligence of Things





New Data Economy

Data-driven services





**Data Security, Privay, Trust and Governance for trustworthiness in AIoT**

**AIoT**: Artificial Intelligence of Things

***Source : Dr. G.M. Lee***