4th SG13 Regional Workshop for Africa on "Future Networks for a better Africa: IMT-2020, Trust, Cloud Computing and Big Data"

(Accra, Ghana, 14-15 March 2016)

Big Data: considerations on perspectives, issues and standardization

Marco Carugi,
Senior Consultant, NEC Corporation
ITU-T Q2/13 and Q2/20 Rapporteur
marco.carugi@gmail.com

1956 2016
CCITT / ITU-T

Outline

- Perspectives: Big Data Analytics
- Some key technical issues for (Big) Data
- Semantic based technologies as key tools to manage data transformation and derive intelligence from Big Data (value, critical aspects, ITU-T studies)
- Big Data technologies (value, critical aspects, ITU-T studies and potential areas of standardization)

1956 2016
CCITT / ITU-T

Key perspectives of Big Data in the ICT market: process optimization and data monetization via analytics

Big Data analytics are a key revenue opportunity in the ICT market

Analytics of data can allow organizations to drive revenue by sharing, analyzing and interpreting data, for multiple purposes

- Extraction of tangible business and technology value
- Response and action in real time, improving productivity and business processes, lowering operational and other costs
- Long-range forecasts enabling strategic actions - to drive competitive advantage and business differentiation, to address society challenges (energy, environment, health, transport etc.)
- New/improved business models and service offer to customers and citizens, faster, more efficiently (e.g. targeted marketing, customer tracking) and agile (e.g. anticipation of societal issues)

Surveys indicate increasing Big Data analytics deployments and exponential growth of Big Data analytics revenues

1956 2016
CCITT / ITU-T

Solution for monitoring and predicting failures for plant facilities [example using NEC's analysis technologies]

Customer's challenge

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

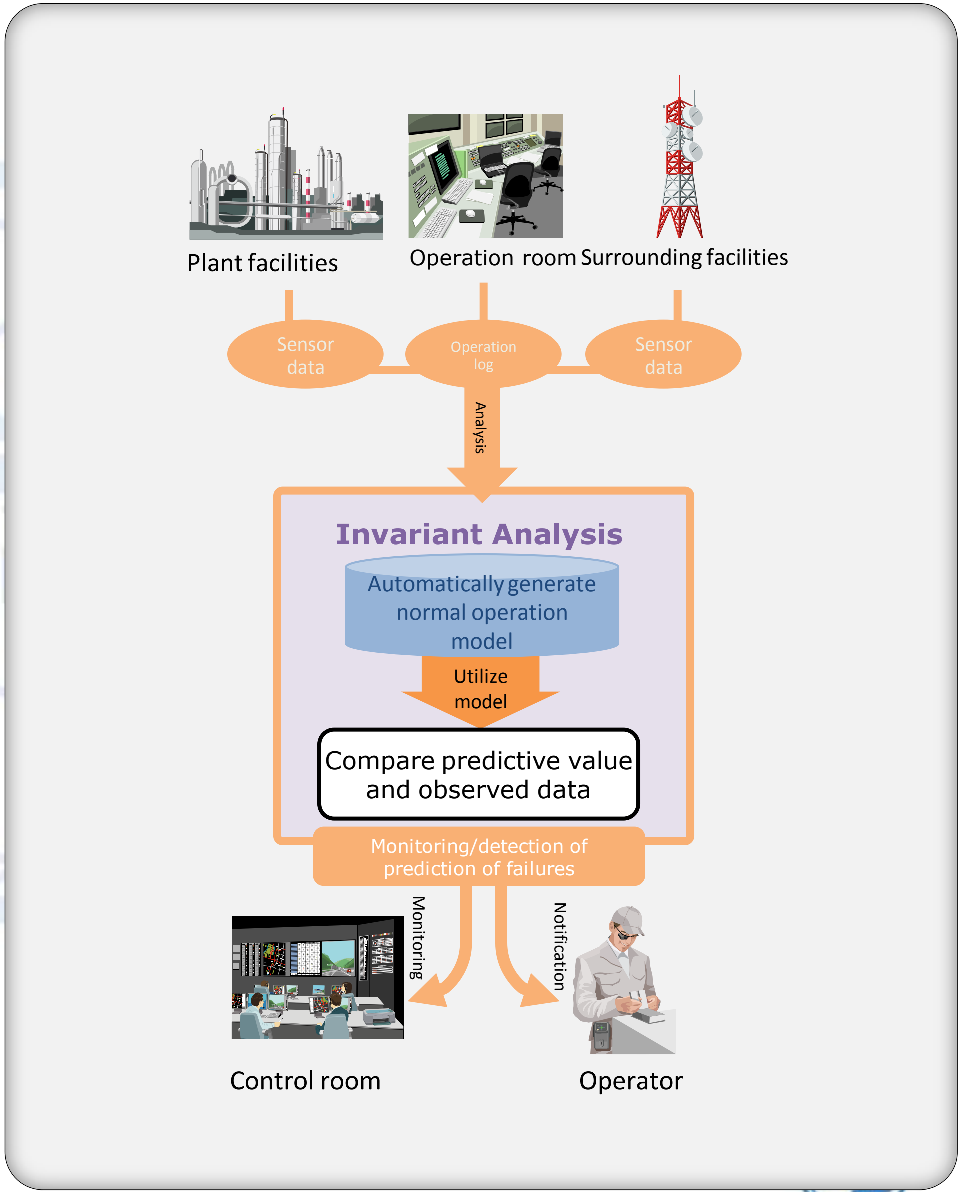
Effect of solution

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

Point of introduction

- Visualize operational status from the existing data by "Invariant Analysis Technology"
- Utilize massive data in real time and realize high accuracy monit. /detection of failure prediction

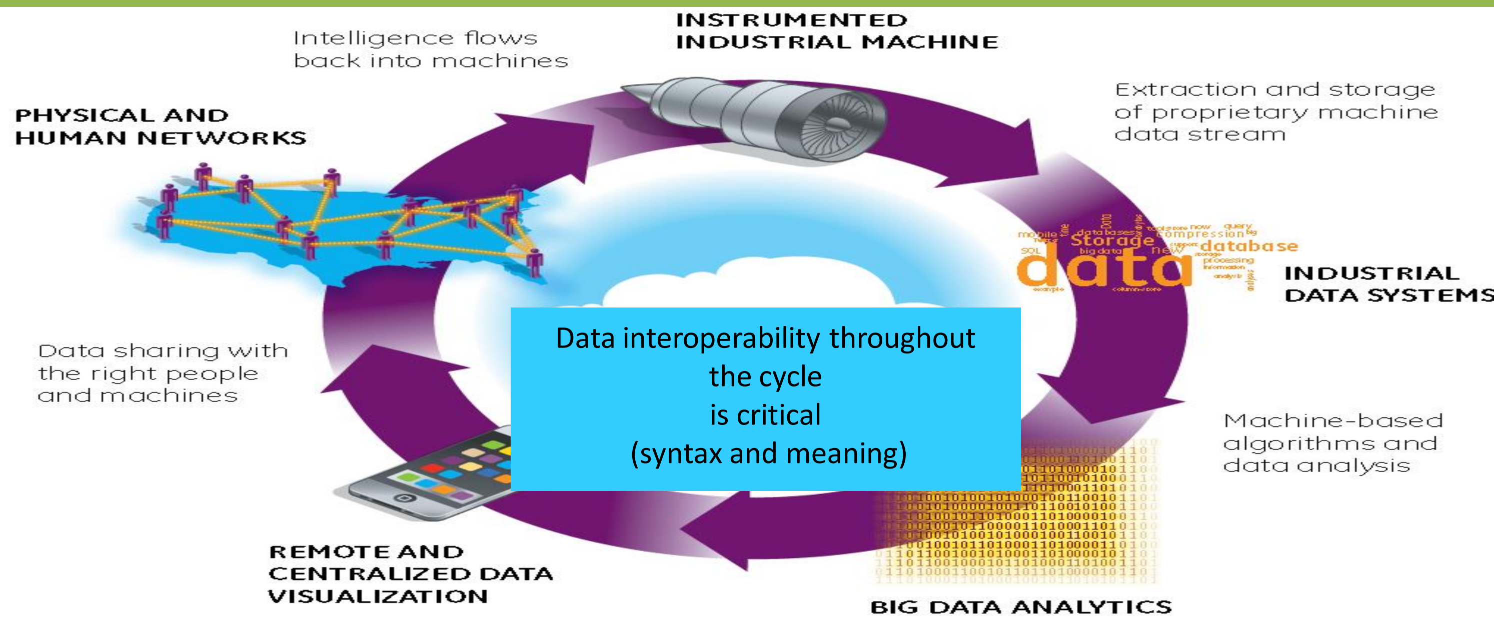
"Invariant Analysis" (NEC's leading analysis technology): analysis of large amounts of metric data collected from multiple sensors to automatically identify relationships and detect anomalies.



An important application domain of Big Data: Internet of Things

- Some analysts indicate that by 2020 40% of data will come from sensors
- Multiple data sources: (real-time) data from things and context, historical and social data (cross-domain data exchange and correlation etc.)
- Data are mainly semi-structured and unstructured
- Data may have different precision and confidence levels
- Various operations to be made on data for the extraction of actionable intelligence (collection, de-noising, aggregation, adaptation, analysis)
- Raw data -> Information -> Knowledge (-> Wisdom)
- Target: the right data, at the right time, **at the right location (e.g. cloud versus edge computing)**

The Industrial Internet Data loop [source: GE whitepaper]

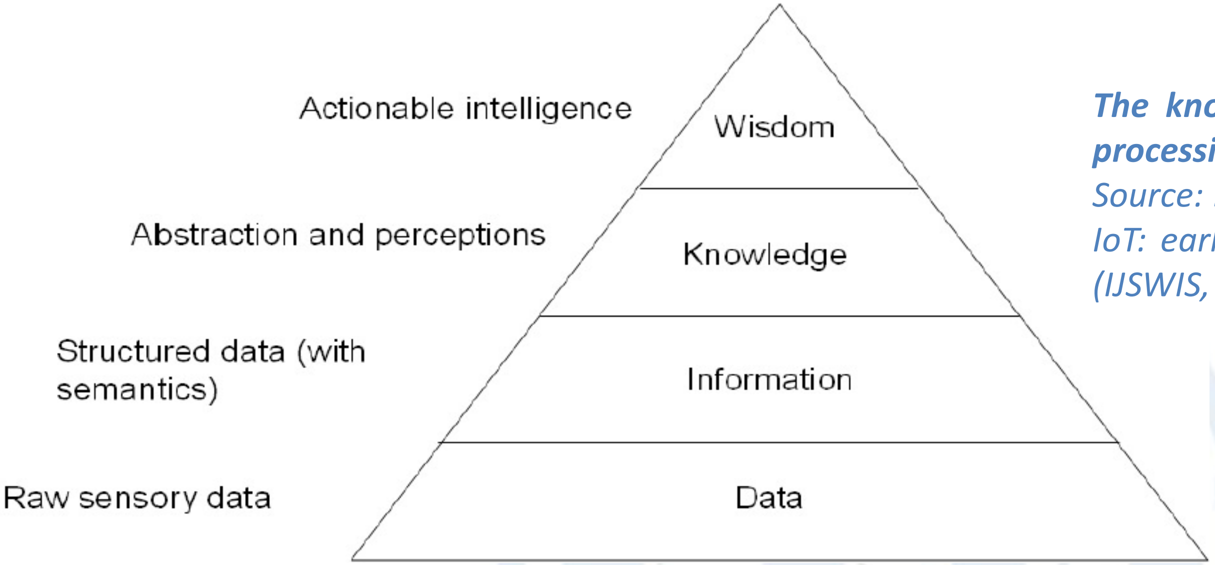


Some key technical issues for (Big) Data

- Dealing with the (Big) Data "V"s (volume, variety, velocity etc.)
- Integration of heterogeneous devices, networks and data
- Scalability to cope with large numbers of devices, diverse and huge data, computational complexity of data interpretation
- Discovery of data sources, data querying, (open) access to data resources/data
- Massive data mining, adaptable learning, efficient computing and processing
- Security and privacy of data (incl. doing mining, analytics)
- Interpretation: extraction of actionable intelligence from data
- Non-technical challenges are essential though -e.g. data ownership, governance

"Big Data" as defined in ITU-T Y.3600: A paradigm for enabling the allocation of storage, management, analysis and visualization, potentially under real-time constraints, of extensive datasets with heterogeneous characteristics. NOTE – Examples of datasets characteristics include high-volume, high-velocity, high-variety etc.

Data transformation



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 the IoT)
- Additional information enables creation of structured metadata (first step of data enrichment - *semantics is applicable*)
- Abstractions and perceptions give detailed insights of data by reasoning , using knowledge (ontologies, rules) of relevant domains (second step of data enrichment - *semantics is applicable*)
- Actionable intelligence allows decision making

Semantics based technologies: key tools for (Big) Data

Semantics is the study of meaning

- A definition of semantics: “The rules and conventions governing the interpretation and assignment of meaning to constructions in a language” [ITU-T Z.341]
- Shared vocabularies for the network entities and their relationships

Requirements for interoperability, scalability, consistency, discovery, reusability, composability, automatic operations, analysis and processing of data and services are more and more essential in networks

- Examples of driving factors from IoT: growing number of interconnected things, variety of devices and connectivity, volume and types of generated data, number and type of services

Semantics based approaches reveal outstanding features to support these requirements in networks

1956 2016 CCITT / ITU-T

But still issues concerning the full applicability of semantics based technologies in networks

The benefits of semantics based technologies will be realized in incremental way: current issues include

- Lack of elaborated use cases as drivers to validate the value proposition (value for the different stakeholders of the value chain)
- Insufficient link with network architecture
- Immaturity of tools, essential to establish semantics based bridging (domain specific ontologies formalizing the meaning of domain data and information models; semantic merging, matching and alignment strategies; mediators to enable integration of disparate data resources; bridges etc.)
- Semantic discovery of services, devices, things and their capabilities
- Semantic metadata framework and base/core network ontology (horizontal cross-domain integration)
- Participation of domain specific communities and creation of social and business incentives for sharing
- Education (entrepreneurs/domain experts – developers interaction)

The initial results are promising, but further research, development, validation and standardization are required.

ITU-T studies on semantics

The initial study has focused on IoT: ITU-T Y.2076 "*Semantics based requirements and framework for the IoT " (SG13-led, approved Feb 2016)*

- Value proposition of the usage of semantics in IoT systems, semantics based requirements at different layers and cross-layer capabilities of the IoT Reference Model [Y.2060], semantics based capability framework of the IoT
- SDOs' collaboration promoted: oneM2M will integrate these results in future studies
- European Commission-launched Alliance for IoT Innovation (AIOTI) has referred Y.2076 in AIOTI WG3's Oct 2015 Technical Report on "Semantic Interoperability"

Further studies are expected (in SG20 (on IoT, initiated in Jan 2016), in other SGs including SG13 (TBD))

Two other SDOs with relevant studies on semantics

- W3C – has pioneered with the Semantic Web, other studies ongoing
- oneM2M – ongoing study on a base ontology for the IoT (ETSI TS 103 264 "Smart Appliances Reference Ontology (SAREF) " will be fully mapped to it)

1956 2016
CCITT / ITU-T

More in general about Big Data technologies: value proposition and some current critical aspects for deployment

Big Data technologies address several data challenges in networks

- Scalability, data integration, massive data mining, data accessibility

and create value (Big Data Analytics)

But some aspects are still critical from a deployment perspective

- Framework for best practices on integration and interoperability with legacy environments and applications
- System performances and reliability
- Wide spectrum of continuously evolving technologies and products (e.g. which tools (identification of needs), which adequate evolution, deployment costs), organizational impact, skilled personnel
- And analytics is becoming a multi-dimensional challenge (data at rest versus data in motion, and the related data cycle operations)
- Data privacy, extended data access, data security, data liability (policies)

1956 2016
CCITT / ITU-T

Big Data technical standardization

Some potential areas of standardization [this list includes – but is not limited to – the areas identified in SG13's Y.Suppl.BigData-RoadMap and AFNOR's 2015 Big Data white paper]

- Requirements and use cases (IoT use cases being one key driving set in industry and society)
- **Architecture, data model and APIs (APIs with network infrastructure, users, auditors)**
- **Flexible analytics (real-time, batch; remote, distributed and federated analytics; network-driven data analytics)**
- Security and data protection, anonymization and de-identification of personal data (and reversibility)
- Framework for data quality and veracity
- Standards and guidelines to address issues related to legal implications of Big Data in telecommunications (e.g. data ownership)
- **Framework and standards for (Telecom) Big Data Exchange** - data sharing, transaction, interconn.
- Benchmarks for system performance evaluation (e.g. Hadoop)
- Standardized visualization methods
- NoSQL query languages, NoSQL-SQL interfaces
- Domain specific languages
- **Open data frameworks**
- Integration of Big Data requirements in cloud/distributed computing solutions for both infrastructure and services (interoperability, data/process security, traceability, personal data protection)

1956 2016
CCITT / ITU-T

Main ITU-T developments related to Big Data

Completed studies (all completed in SG13)

- **Y.3600**: *Big data – **cloud computing based** requirements and capabilities*
- **Y.2066:** *Common requirements of the **IoT*** (includes requirements on IoT data aspects)
- **Y.2068:** *Functional framework and capabilities of the **IoT*** (includes IoT capabilities for data management and for integration of Big Data technologies and Cloud Computing technologies with the IoT)

Ongoing studies

- **In SG13: Y.BigDataEX-reqts:** *Big Data Exchange Framework and Requirements*
- **In SG13***:* **Y.BDaaS-arch:** *Functional architecture of Big Data as a Service*
- **In SG13: Y.Suppl.BigData-RoadMap:** *Big Data Standardization Roadmap*
  - standards landscape in ICT sector and al., feasibility studies on Big Data standardization, identification of technical areas (e.g. cloud comp. and IoT), roadmap
- **In SG20: Y.IoT-BigData-reqts**: *Specific requirements and capabilities of IoT for Big Data*

A number of other standards initiatives beyond ITU-T

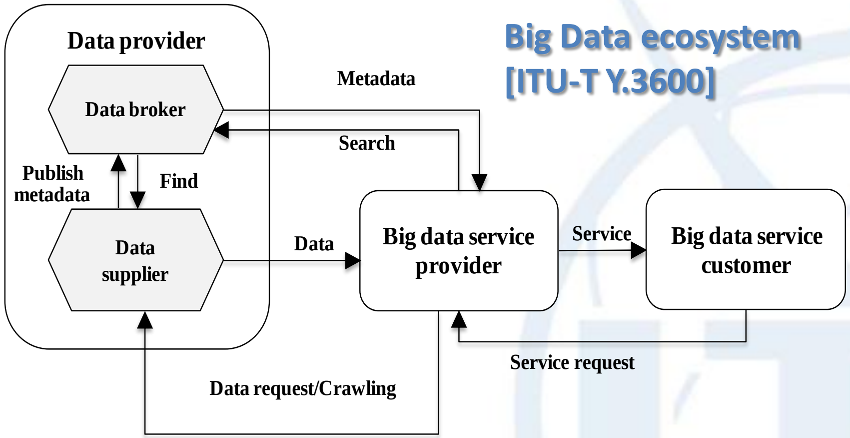
- ISO/IEC JTC1 WG9, W3C Data Activity group, others

1956 2016
CCITT / ITU-T

Potential standardization areas: architectures & (different level) APIs

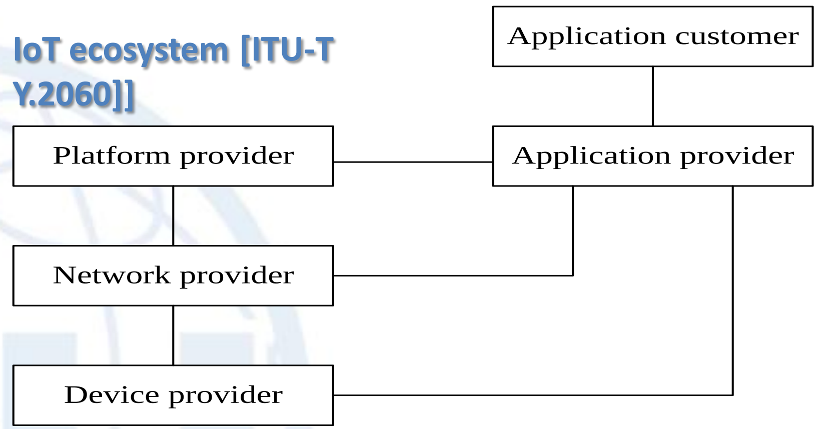
How to integrate Big Data platforms with Telecom Infrastructures : putting things together

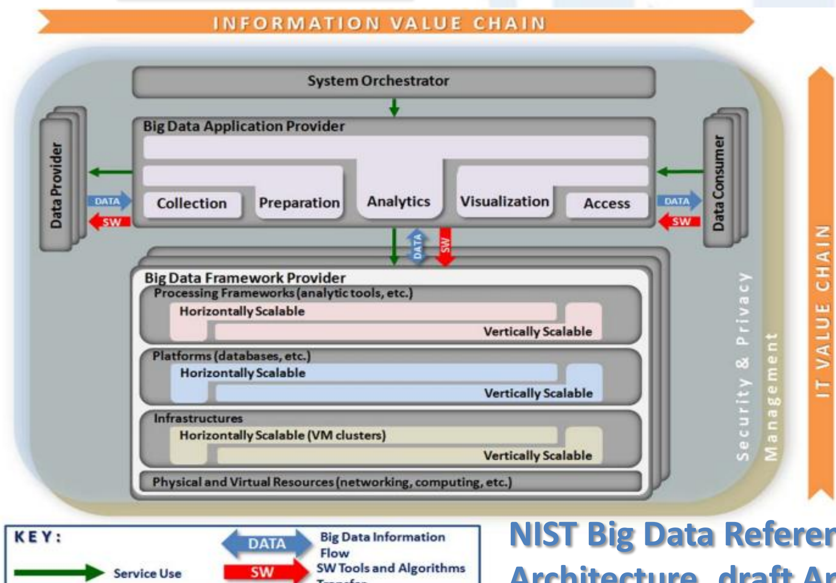
(this slide uses IoT as a relevant domain of application)



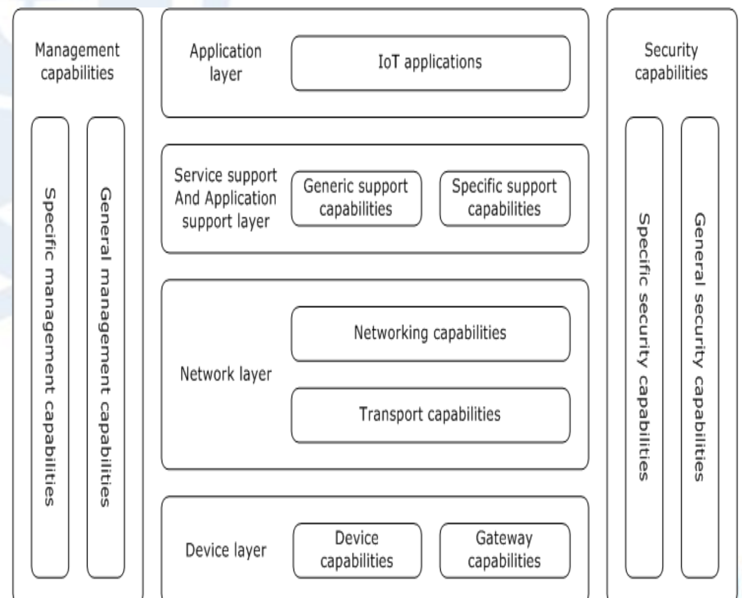
Big Data ecosystem [ITU-T Y.3600]

IoT ecosystem [ITU-T Y.2060]]





NIST Big Data Reference Architecture, draft April 2015 – an example of Big Data Ref. Arch. study

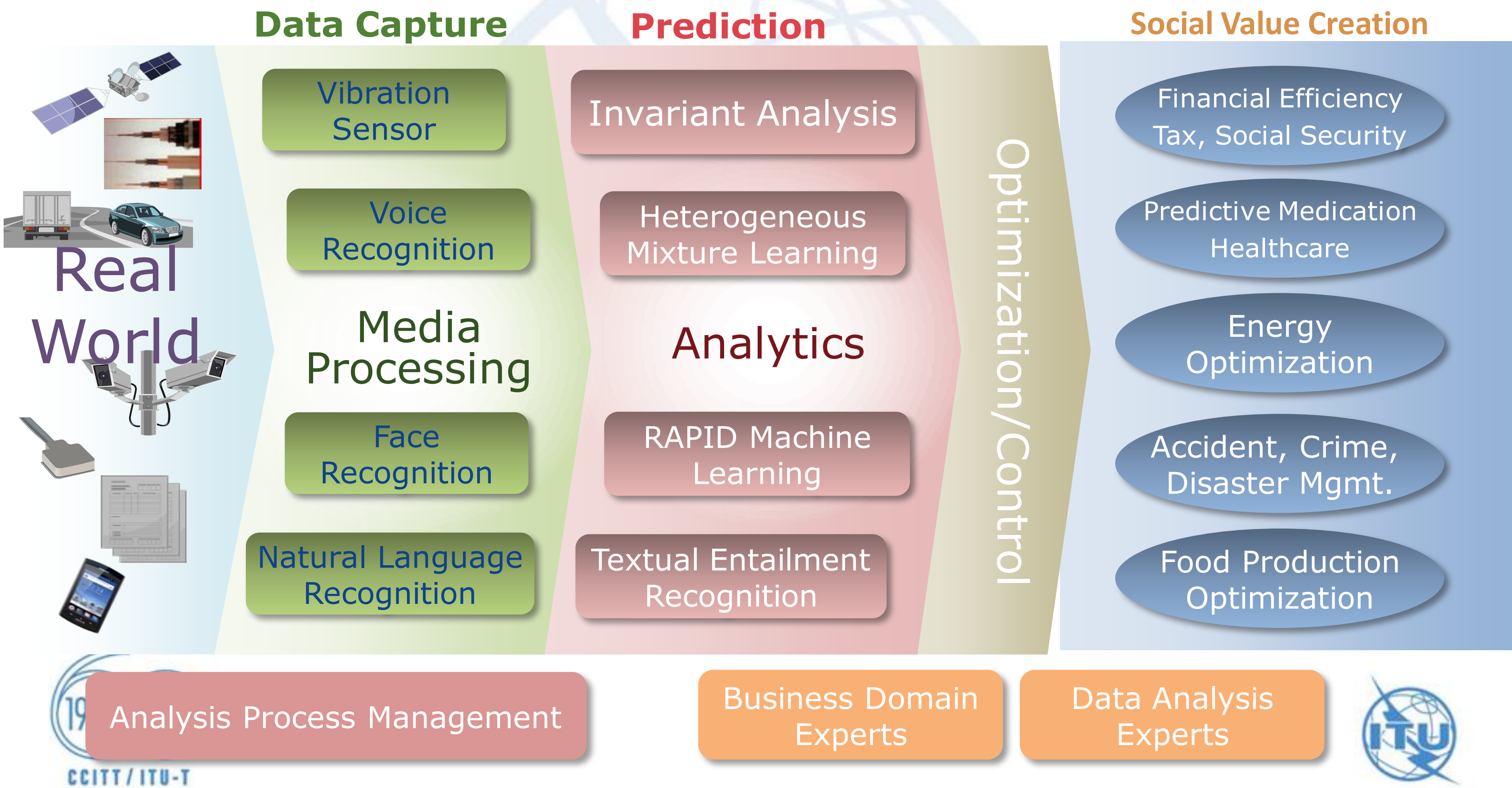


IoT reference model [ITU-T Y.2060]

CCITT / ITU-T

NEC's technologies to transform Big Data into Values

Social value creation based on 1) edge devices and data collection platform, 2) advanced analytics technologies, and 3) advanced analysis expertise and process management



World leading analysis technologies from NEC

World leading analysis technologies for knowledge acquisition from big data

World 1st

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

Invariant Analysis

World 1st

Automatically detects massive patterns hidden in big data. Achieves high-precision predictions and anomaly detections that are difficult for handcrafted data analytics.

Heterogeneous Mixture Learning(HML)

World #1※1

High speed and light processing deep learning based matching engine

RAPID Machine Learning

World #1※2

Delivers best-in-class performance in NIST testing. Recognizes semantic content between texts.

Recognizing Textual Entailment

World 1st

Collecting personal face features from surveillance camera, and checking with particular person in real time

Face Recognition

※1 PTP: Penn Treebank Project 2011
※2 NIST : National Institute of Standard and Technology

Thank you very much for your attention

Marco Carugi
Tel. +33 6 64047454
marco.carugi@gmail.com

CCITT / ITU-T

ITU

1956 2016

CCITT / ITU-T