ISDE Seminar, 27 Mar 2024

Big Earth Data analytics in Digital Earth

Stefano NATIVI

National Research Council (CNR) and Permanent Representation of Italy to European Union

Consiglio Nazionale delle Ricerche

Big Data Analytics

Process of examining *large* and *varied* datasets to **uncover hidden patterns**, **correlations**, **trends**, and **insights** that can help organizations make informed decisions

VOLUME: traditional data processing applications are not equipped to handle large volumes of data (e.g. social media, sensors, mobile devices, and transactional systems)

VARIETY: data comes in different formats (including structured data, semi-structured data, and unstructured data) and specialized tools and techniques are needed to analyze them

VELOCITY: analytics is generated at a high velocity and needs to be processed quickly to derive timely insights (real-time or near-real-time) and make effective decisions (e.g. disaster mitigation, early warning, finance, healthcare, and cybersecurity)

VERACITY: data has of varying quality (including noisy, incomplete, and inconsistent data). Different techniques (e.g. data cleansing and data quality assessment) are essential to ensure the accuracy and reliability of analytical results

VALUE: analytics primary goal is to extract actionable insights from data (e.g. improve decision-making, enhance operational efficiency, identify new business opportunities, and drive innovation)

Big Data Analytics

Intelligence from latin “intellegentia” (*intus-legere*) = “intus-read” and means "to see into“ (**to see deep**)

FUTURE TRENDS: analytics evolves rapidly with advancements in technologies such as artificial intelligence (AI), machine learning (ML), and edge computing. Future trends include the integration of Big Data Analytics with IoT, blockchain, and augmented analytics

SUCCESS FACTORS: data security and privacy concerns, data governance issues, scalability issues, and the need for specialized skills and expertise to analyze large and complex datasets

ETHICAL CONSIDERATIONS: data privacy, consent, transparency, fairness, and accountability

APPLICATIONS: enabling different functionalities user segmentation and analytics, demand forecasting, personalized services, risk management, crisis detection, management optimization, predictive maintenance, supply chain optimization

TECHNOLOGIES: analytics must rely on a variety of powerful technologies and tools (e.g. distributed computing frameworks, NoSQL databases, decentralized data systems, ML/DL algorithms, natural language processing (NLP), and visualization tools)

Earth Observation AI and GeoAI

GEOSPATIAL FOUNDATION MODELS

- A model pre-trained on an extensive portfolio of remote-sensing (unlabeled) data, allowing to direct expert's efforts towards fine-tuning and seamless inference
- Effortlessly save time and resource allocation without compromising precision
    - Reusable models that can be applied to just about any domain or industry task
    - FM need a fine-tuning step to enable specific applications
    - Example: NASA and IBM

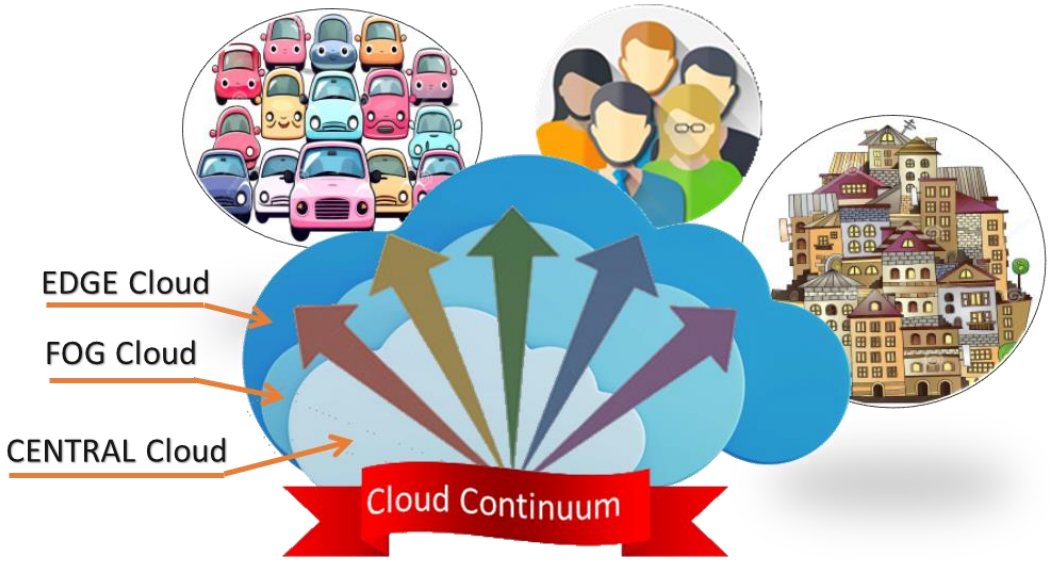
GEOSPATIAL ML/DL

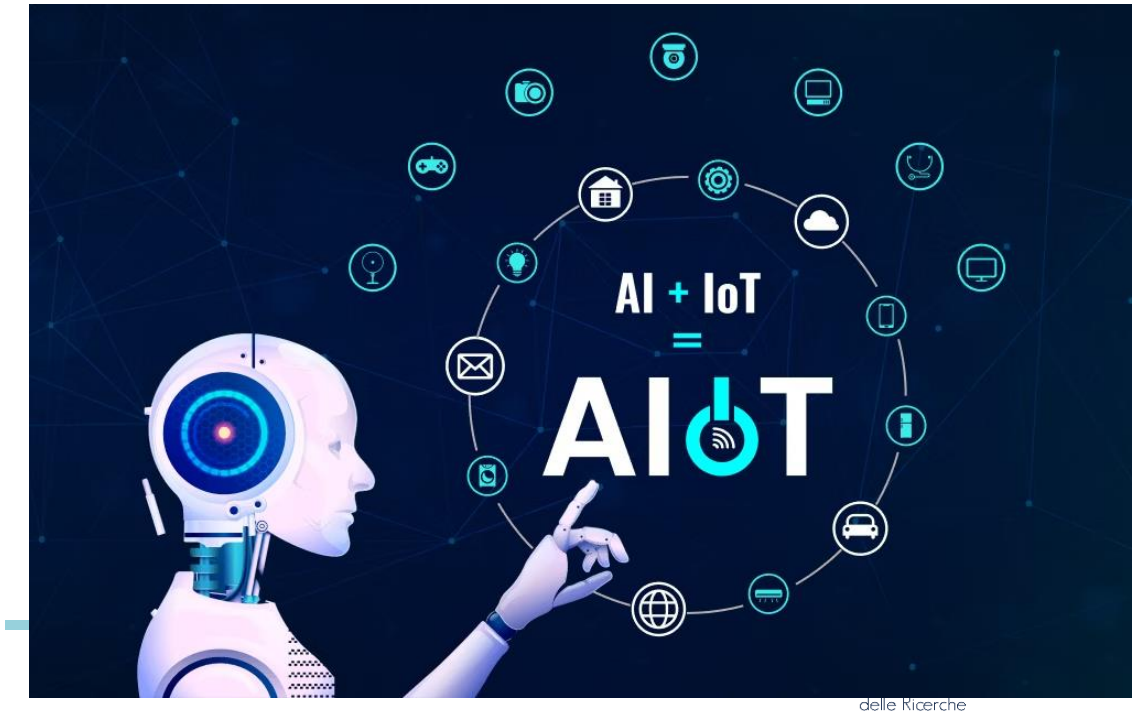
- Describe input data as feature (high-dimensional) vectors labelling phenomenon
- Find adequate feature spaces for geospatial entities

When AI meets Internet of Things (Smartification process)

AIoT is a subset of AI that **refers to AI performed on IoT-type data sources**

- **Edge AI** refers to processing data where it is generated (rather than sending it to a central cloud via the internet). This involves deploying ML models directly to the network edge
- **AIoT and 5G/6G** is going to realize real-time processing
- AIoT improves **human-machine interactions** and **society smartification** (e.g. smart cities, smart home, smart office buildings, autonomous vehicles, city brains, …)





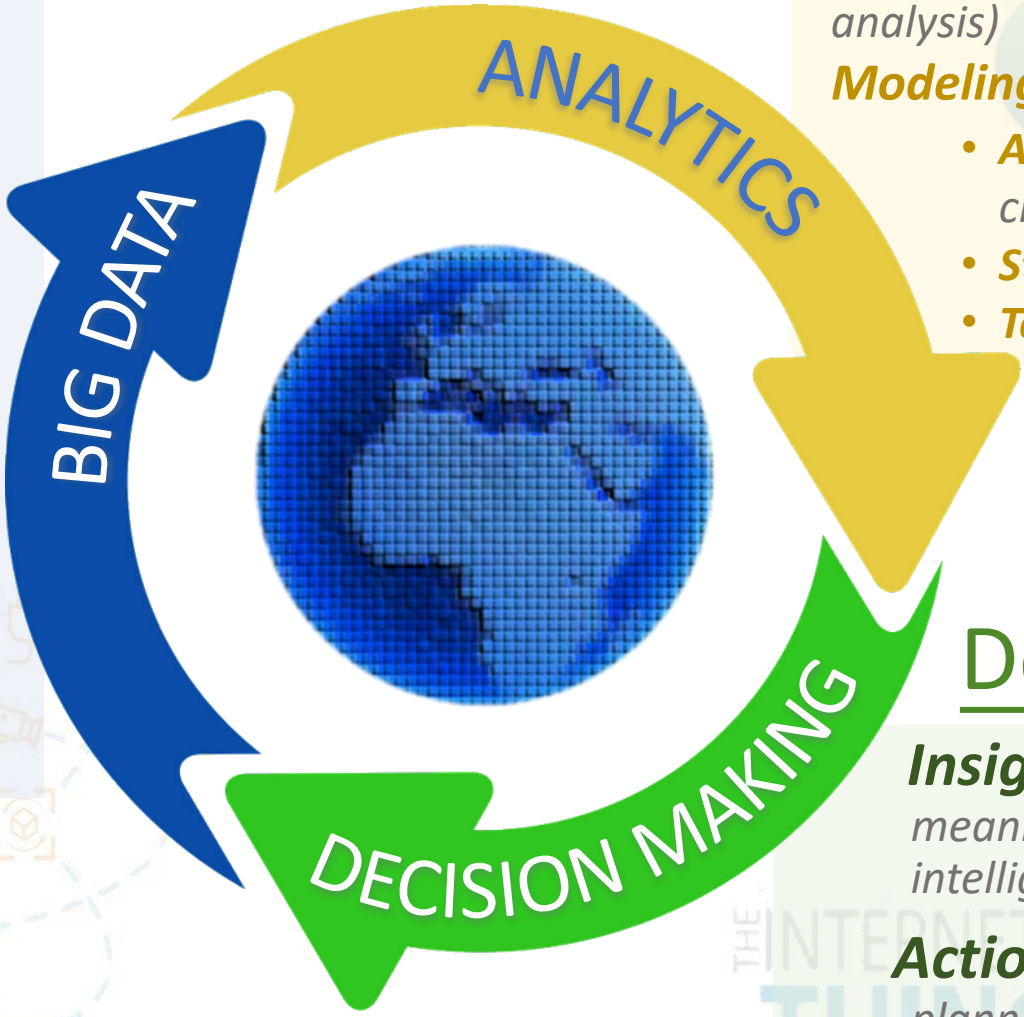
BIG DATA

Data Collection

- *SCIENTIFIC O&M*
- *SOCIAL SENSING & TRANSACTIONS*

Data Preprocessing

- *DATA CLEANING (removing duplicates, handling missing values)*
- *DATA INTEGRATION (combining data from multiple sources)*
- *DATA TRANSFORMATION (standardizing formats, converting data types)*
- *DATA REDUCTION (aggregation, dimensionality reduction)*



ANALYTICS

Exploratory Data Analysis *(descriptive statistics, data visualization, and correlation analysis)*

Modeling and Data Mining

- ***AI-ML*** *(regression, classification, clustering)*
- ***Statistical analysis***
- ***Text mining***

Decision Making

Insights generation *(extracting meaningful insights, insight interpretation and intelligence provision)*

Actionable intelligence *(strategic planning to operational adjustments)*

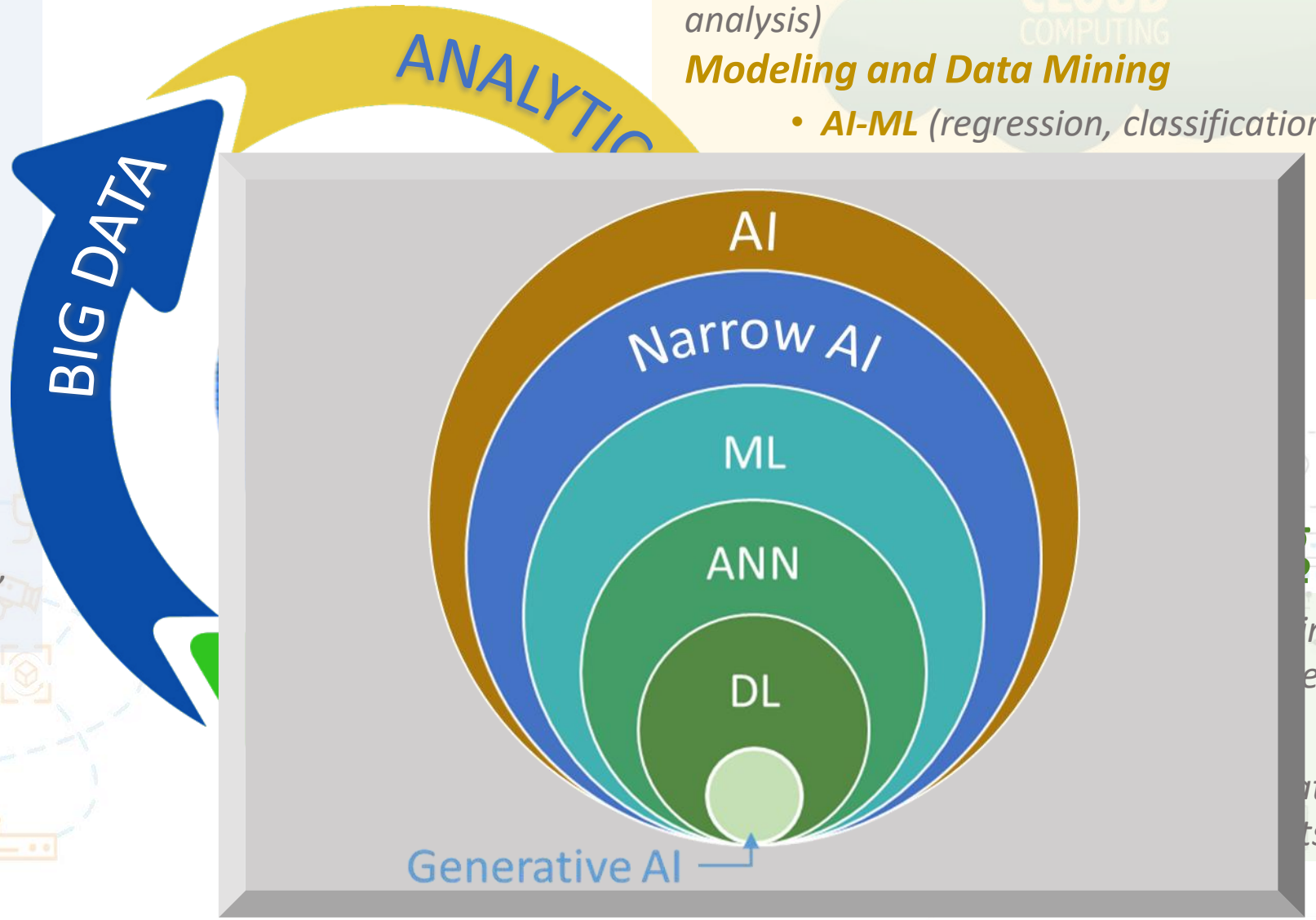
BIG DATA

Data Collection

- *SCIENTIFIC O&M*
- *SOCIAL SENSING & TRANSACTIONS*

Data Preprocessing

- *DATA CLEANING (removing duplicates, handling missing values)*
- *DATA INTEGRATION (combining data from multiple sources)*
- *DATA TRANSFORMATION (standardizing formats, converting data types)*
- *DATA REDUCTION (aggregation, dimensionality reduction)*



ANALYTICS

Exploratory Data Analysis *(descriptive statistics, data visualization, and correlation analysis)*

Modeling and Data Mining

- ***AI-ML*** *(regression, classification,*

ing
etation and
ategic
ts)

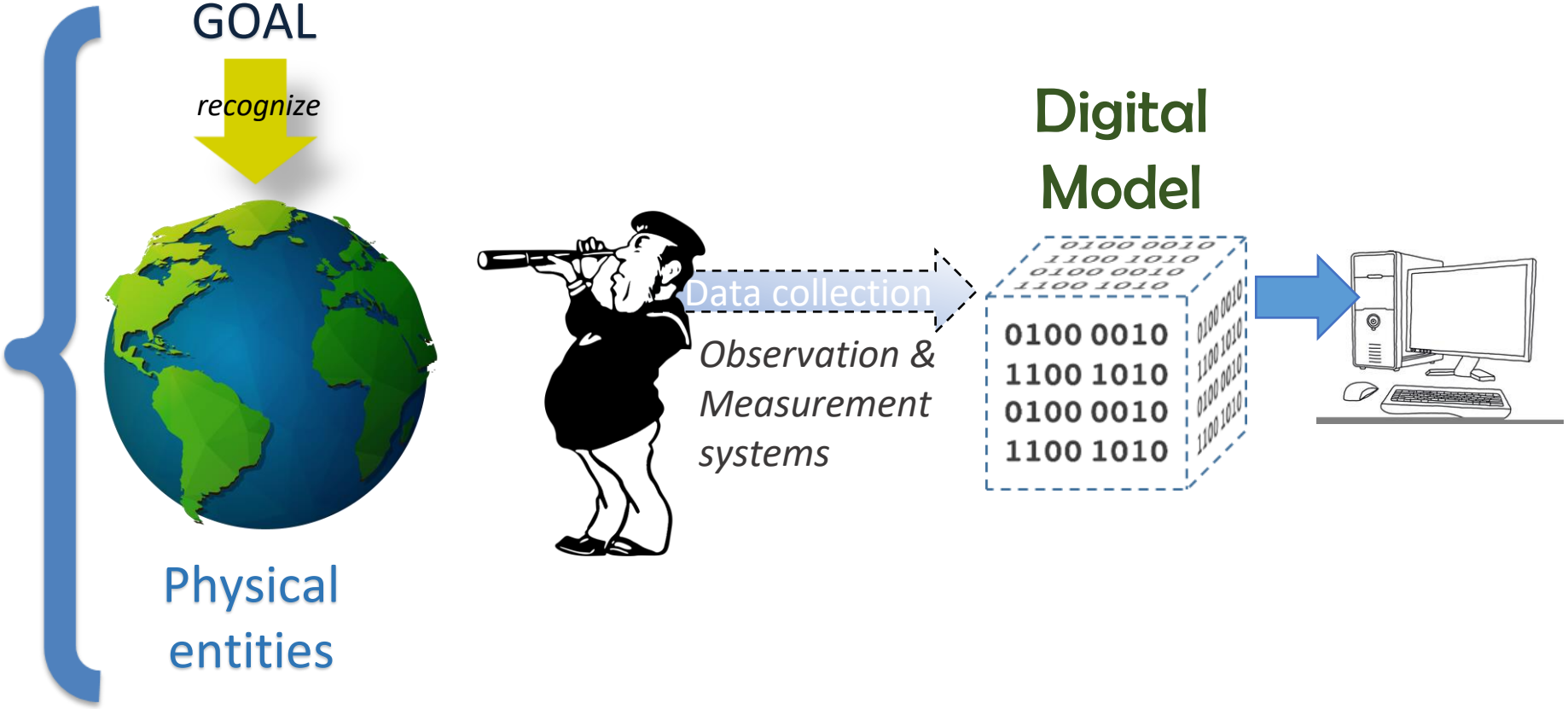
[cyber-physical model]

The (Geo) Digital Twin pattern

Digital twins provide a range of advantages in all areas of society and business

This is NOT a Digital Twin
(towards the Metaverse)

A Digital Model

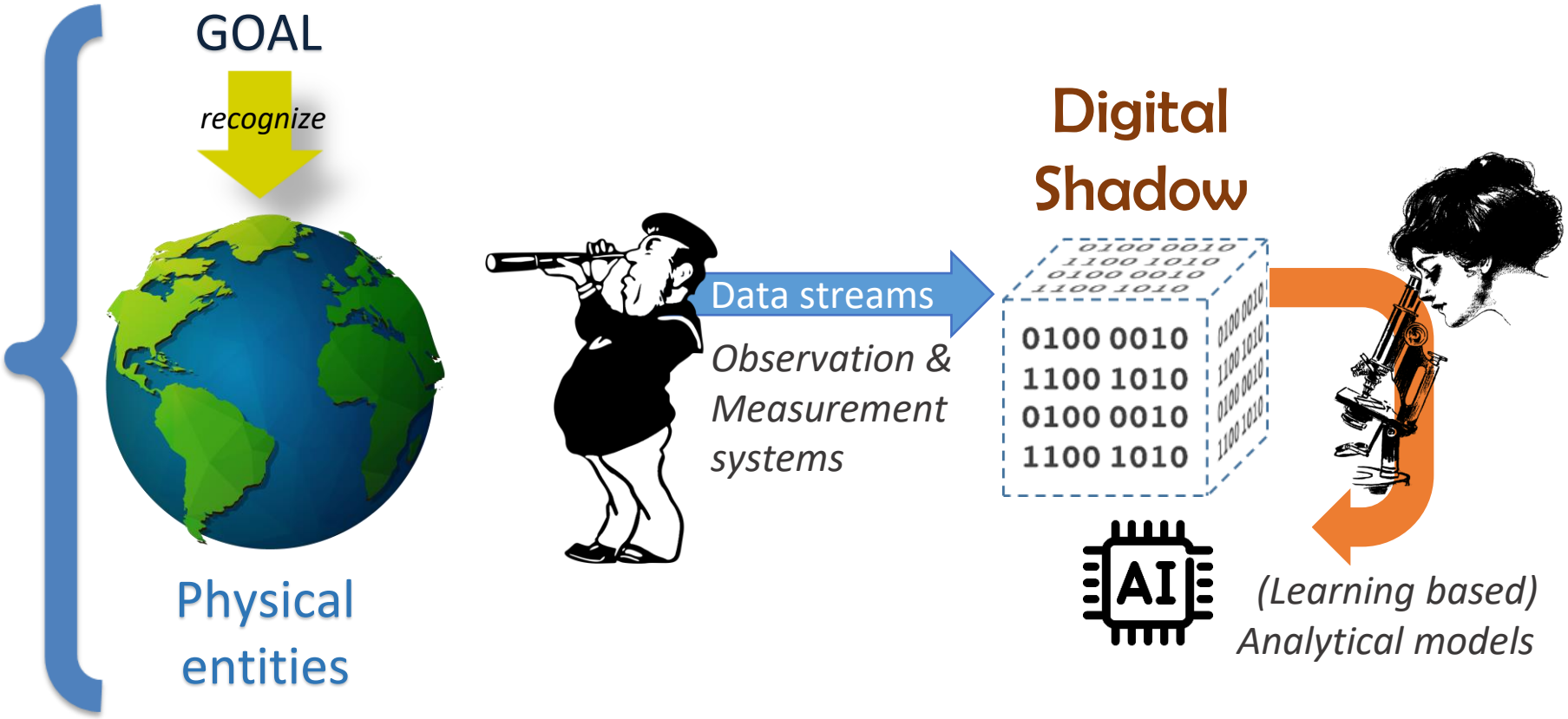


Digital Model → Static representation (NO behavior)

Consiglio Nazionale delle Ricerche

This is NOT a Digital Twin
(towards the Metaverse)

A Digital Shadow



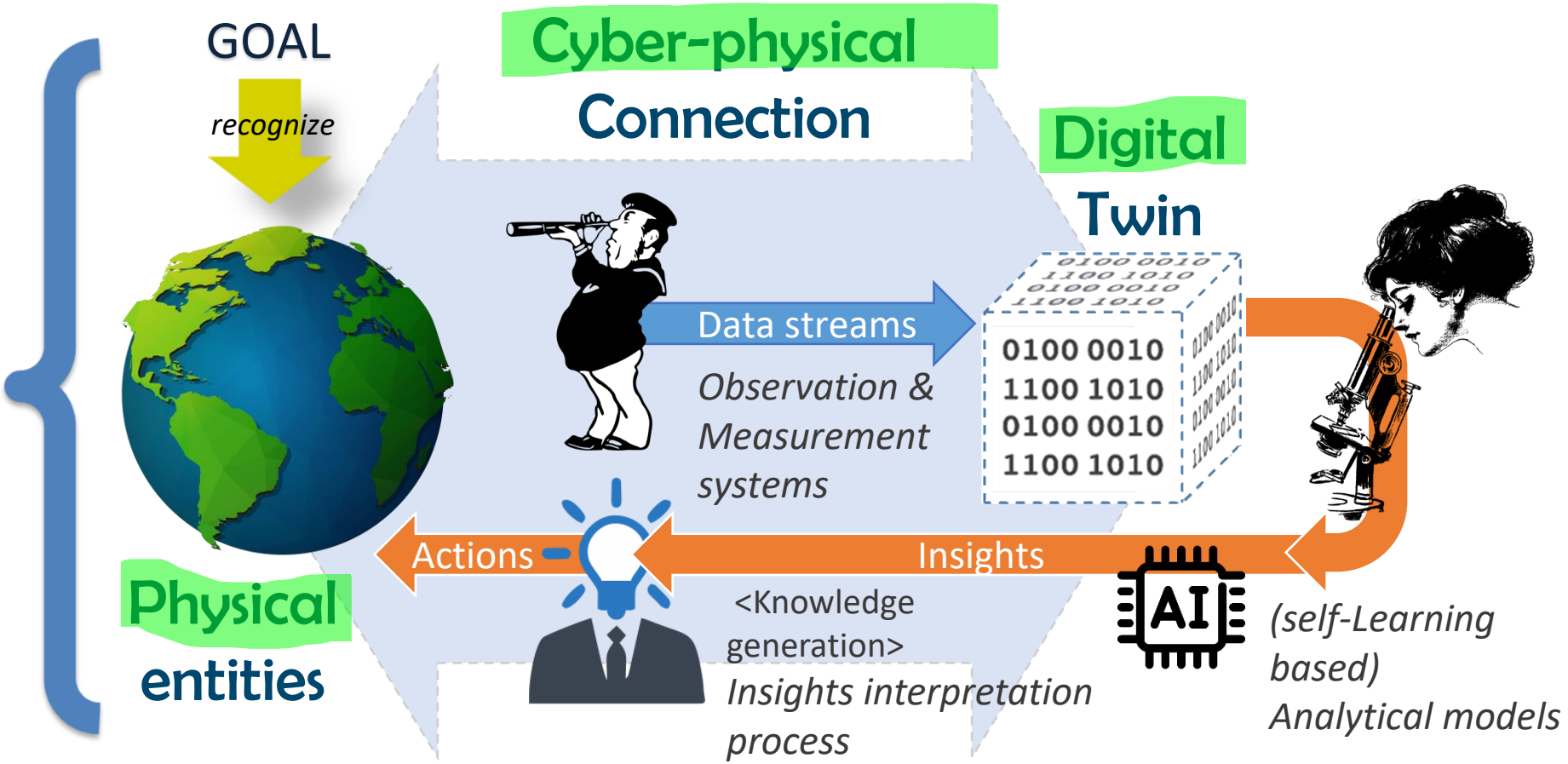
Digital Shadow → *From static to dynamic representation (**Behavior**)*
*Learning analytics (**Simulation and Prediction**)*

Consiglio Nazionale delle Ricerche

(Geospatial) Digital Twin pattern
(towards the Metaverse)

Cyber-physical interconnections must be

- Continuity
- Autonomy
- Bi-directionality



Digital Twin → Behavior + Simulation & Prediction + Action (Smartification)

(Geospatial) Digital Twin pattern
(towards the Metaverse)

Geospatial Digital Twin

(*Encyclopedia of Mathematical Geosciences, Springer Nature*)

Digital replica of an Earth system component, structure, process, or phenomenon obtained by merging digital modelling and real-world observational continuity

A Geospatial DT continuously learns and updates itself and must be seen as a living digital simulation model that modifies and changes itself as its physical counterpart changes

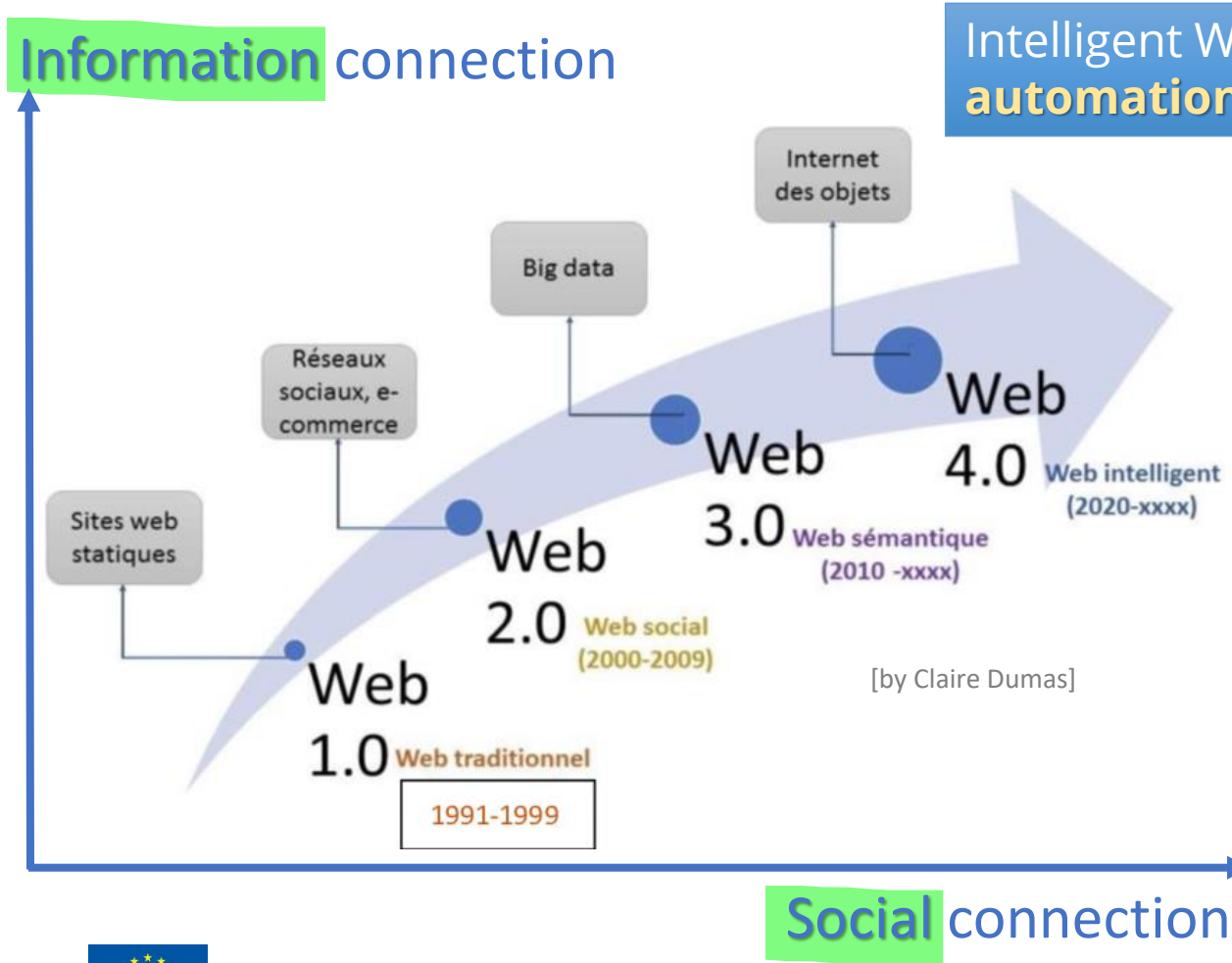
Digital Twin → Behavior + Simulation & Prediction + Action (Smartification)

Consiglio Nazionale delle Ricerche

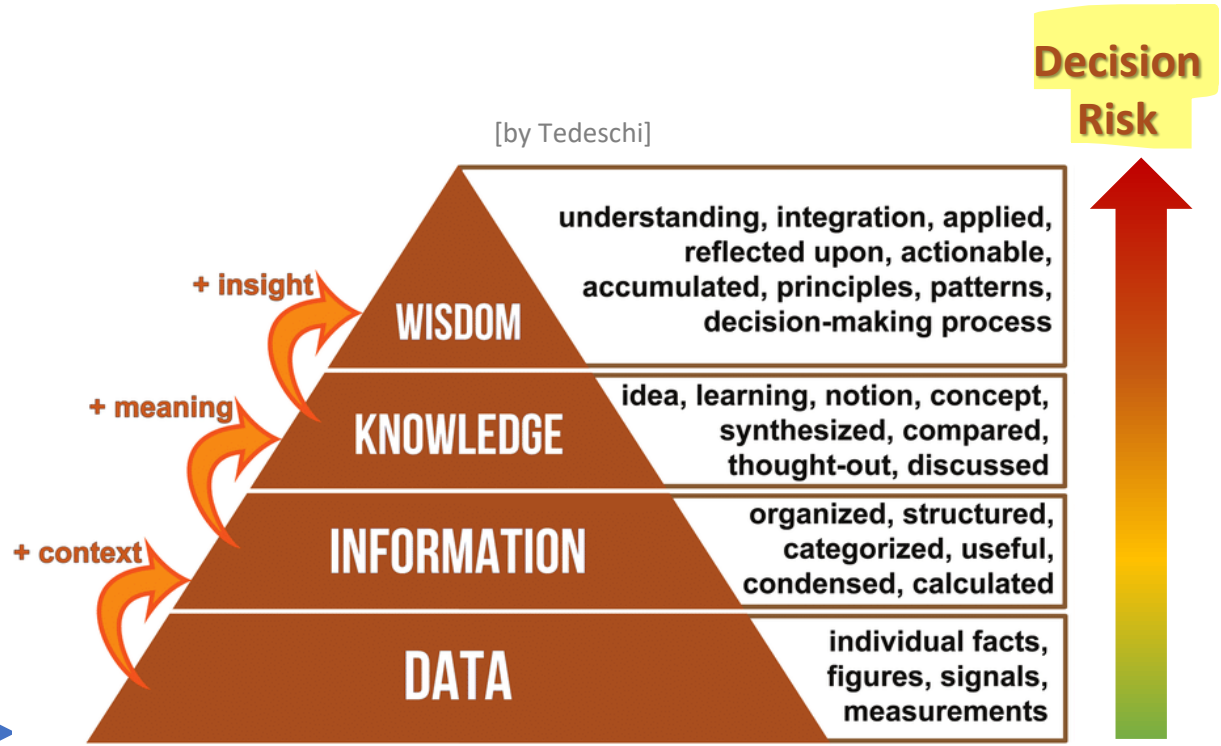
Web 4.0: the Smart Web & Virtual Worlds

Web 4.0 creates a seamless interaction between humans and machines

Intelligent Web enables **personalized experiences**, **efficient automation**, and **enhanced decision-making**



[by Claire Dumas]



Consiglio Nazionale delle Ricerche

Web 4.0: the Smart Web & Virtual Worlds

Web 4.0 creates a seamless interaction between humans and machines

Information connection

Intelligent Web enables **personalized experiences**, **efficient**

Sites web statiques

Political, Economic, and Societal FILTERS

Social costs

Ethical principles

Industry process

Governance lobbies

Decision Risk

Social connection

Consiglio Nazionale delle Ricerche

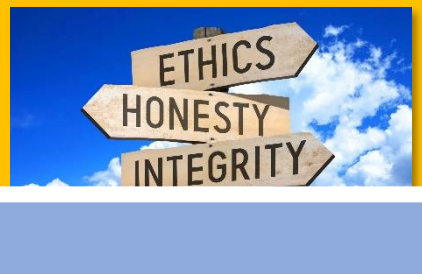
Web 4.0: the Smart Web & Virtual Worlds

Web 4.0 creates a seamless interaction between humans and machines

Information connection

Intelligent Web enables **personalized experiences**, **efficient**

Political, Economic, and Societal FILTERS



Sites web statiques

Decision Risk

Digital open (sovereignty) strategies

- EU Regulations
- Regional/multi-lateral Standards

Consiglio Nazionale delle Ricerche

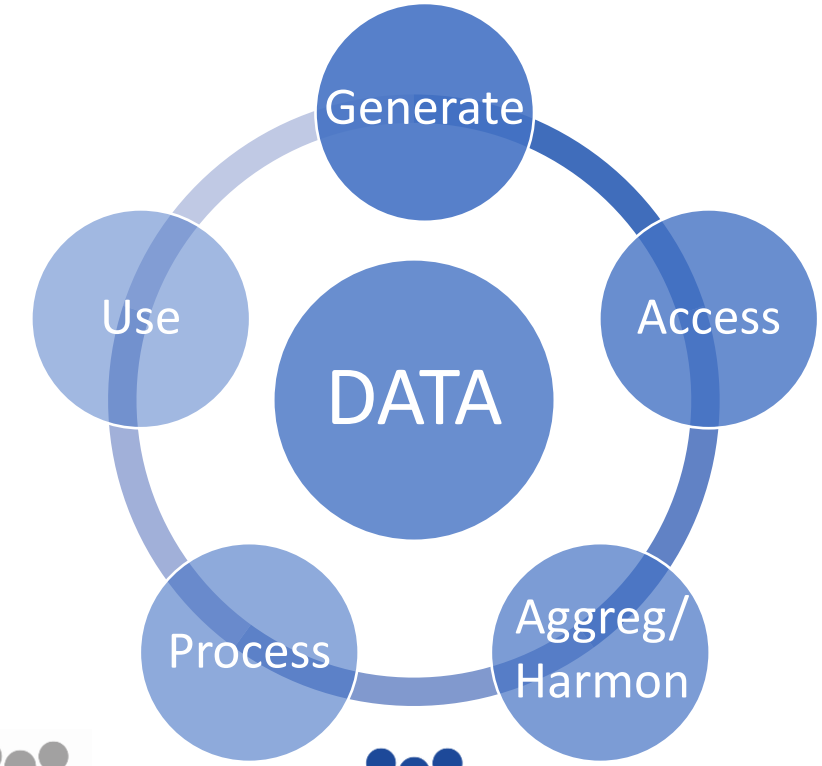
ESSENTIAL SHIFTS NEEDED

- Value-chain model
- SoS Engineering & Interoperability model
- Analytics model access paradigm

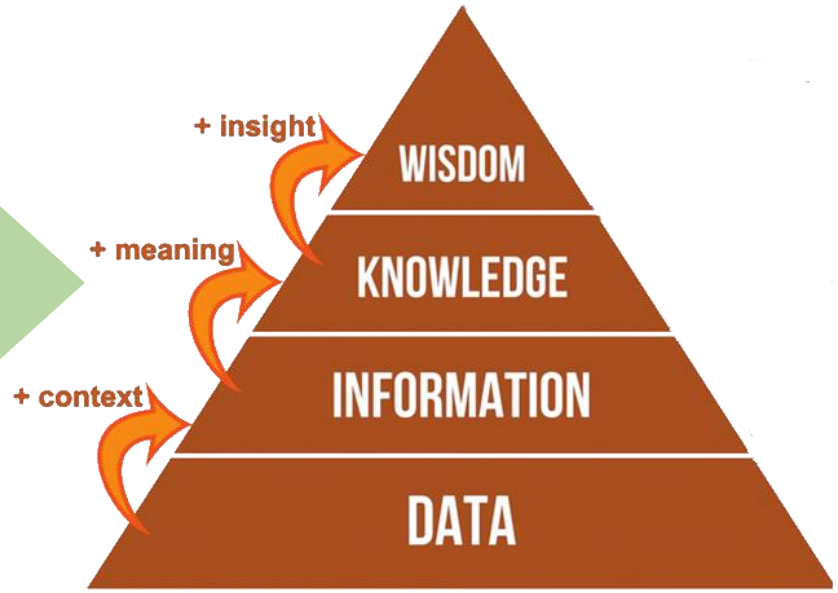
Value chain model shift

[*Data-exchange → Resources orchestration*]

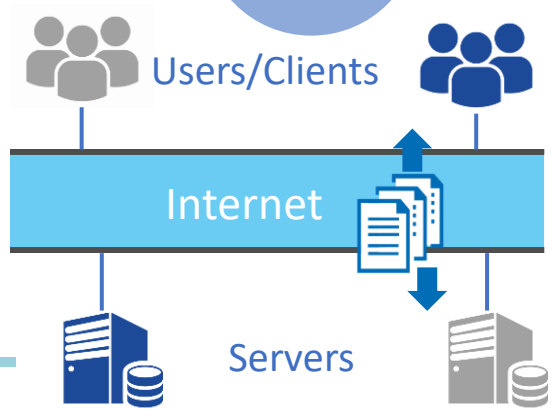
largely [OFFLINE]



fully [ONLINE]



Spatial Data Infrastructures



Digital (resources) Ecosystems

Consiglio Nazionale delle Ricerche

Value chain model shift

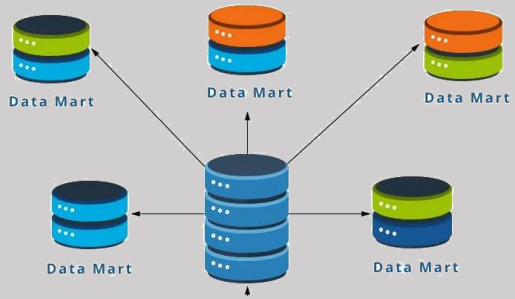
[*Data-exchange → Resources orchestration*]

1. **Move** (all) tasks and transactions from the *physical* (off-line) **to the *virtual* world** (on-line)
2. **Data is only one ingredient of the whole smartification recipe**
3. **Orchestrating resources via online services** (sensors, datasets, analytical models, tools, computing, storing, networking, security capacities)
4. **Move** from the "data discovery-access-download-use" paradigm to the **"digital resources orchestration"** one
5. **NO Big Data egress** principle: move algorithms where effective capacities (e.g. Big Data, HP computing, secure platform, …) are
6. **Move** from "Spatial Data infrastructure" to "**Geospatial Digital resources Ecosystem**"

SoS Engineering & interoperability model shift

[Data Systems Federation → Digital Ecosystem]

SoS INTEROPERABILITY



Data Warehouse

Resources/Services Orchestration

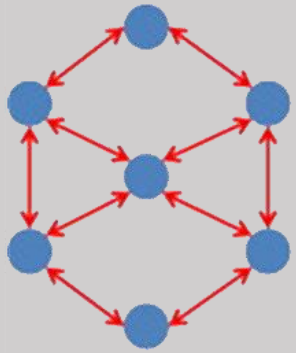
Online Analytics access

Virtual computing capacities

DT generation

AIOT & WaaP

SoS ARCHITECTURE



Directed Federation

Modularity

Evolvability

Flexibility

Viability

Enterprise systems autonomy

Content and digital Diversity

Digital Ecosystem

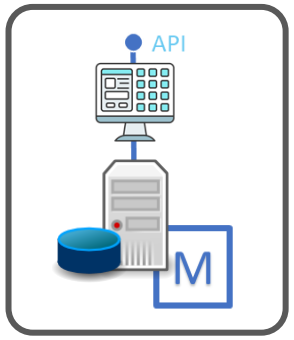
Consiglio Nazionale delle Ricerche

1. CAPACITIES
   - Data is **only one** of the resource types to be shared and used
2. INTEROPERABILITY
   - Balance between the SoS "belonging" and the **autonomy and diversity of an** enterprise system
3. ARCHITECTURE ENGINEERING
   - **Move** from stiff digital Federations to more resilient and evolvable geospatial **digital Ecosystem**
   - **Implement** SoS **viability and evolvability by design**
4. GOVERNANCE
   - **For the Society, the ecosystem common utility is as important as the enterprise utility**
   - **Move** from a fully directed-style to a more **collaborative** one
   - Even **to allow** systems **competition** (when useful for the ecosystem)

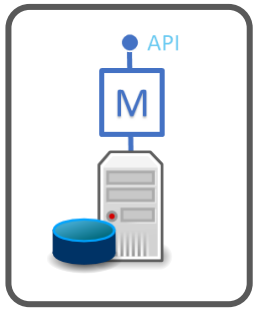
Analytical models access shift

[MaaS → MaaR]

1. **Model as a multi-functional system** (interaction with the system containing the model and not with the model itself)
   - **System API is published** to allow M2M interactions
   - **Client request** sent to the multifunctional system, via API



2. **Model-as-a-Service** (APIs exposed for model configuration and execution, on provider infrastructure)
   - **Traditional process-driven analytics** (Client must send request to the service interface; data should be moved to the model provider infrastructure)
   - **AI/ML analytics** (client might be asked to use lambda functions (e.g. cloud serveless functionalities)



3. **Model-as-a-Resource** (client get an instance of the model)
   - **Resource-as-a-procedure** (client get a model algorithm encoded in a standard language and available on a Git)
   - **Resource-as-a-trained NN** (client get a ML module encoded in a standard format and available on a Git)

Analytical models access shift

[MaaS → MaaR]

1. **Reusability of computational models** is required by the Open Science and Open Knowledge paradigms
   - Sharing models as runnable services is not sufficient for **transparency and reusability**
2. **No Big Data egress principle**: move algorithms (e.g. containerization)
3. Move from MaaS to MaaR frameworks, that allows sharing **models as full Web resources** (see also Web 4.0)
4. In a MaaR framework models can be reused beyond the intention of providers and where is more appropriate for **resources orchestration**
5. Recent manuscript (e.g. Mazzetti and Nativi) on **MaaR framework technical implementations**

Digital Ecosystem paradigm

- ***A dynamic and holistic view of a (virtual) space evolving over time***
  - *New digital resources arise*
  - *New Organization arise*
  - *New external pressures (e.g. from policy, economy, society)*

- ***Introducing the Ecosystem utility, in addition to the Organization one***
  - *To make the virtual space robust, resilience and scalable*
  - *To make the Organization systems viable*

References

- Digital earth: yesterday, today, and tomorrow
  A Annoni, S Nativi, A Çöltekin, C Desha, E Eremchenko… - International Journal of Digital Earth, 2023
- Big Earth Data science: an information framework for a sustainable planet
  H Guo, S Nativi, D Liang, M Craglia, L Wang, S Schade… - International Journal of Digital Earth, 2020
- The model-as-a-resource paradigm for geoscience digital ecosystems
  P Mazzetti, S Nativi - Environmental Modelling & Software, 2024
- Environmental model access and interoperability: The GEO Model Web initiative
  S Nativi, P Mazzetti, GN Geller - Environmental Modelling & Software, 2013
- Virtual earth cloud: a multi-cloud framework for enabling geosciences digital ecosystems
  M Santoro, P Mazzetti, S Nativi - International Journal of Digital Earth, 2023
- Digital ecosystems for developing digital twins of the earth: The destination earth case
  S Nativi, P Mazzetti, M Craglia - Remote Sensing, 2021
- A future for digital public goods for monitoring SDG indicators
  D Liang, H Guo, S Nativi, M Kulmala, Z Shirazi, F Chen… - Scientific data, 2023
- MyDigitalTwin: Exploratory Research Report
  S Nativi, M Craglia, L Sciullo - 2022
- Geosciences digital ecosystems
  Nativi, S., and P. Mazzetti. Encyclopedia of mathematical geosciences. Encyclopedia of earth sciences series (2021)
- Digital twins of the earth
  Nativi, S., and M. Craglia. *Encyclopedia of mathematical geosciences. Encyclopedia of earth sciences series* (2021)
- Destination Earth: Ecosystem Architecture Description
  Stefano NATIVI and Max CRAGLIA,, EUR 30646 EN, Publications Office of the European Union, Luxembourg, 2021
- AI Standardisation Landscape: state of play and link to the EC proposal for an AI regulatory framework
  Nativi, S. and De Nigris, S. EUR 30772 EN, Publications Office of the European Union, Luxembourg, 2021
- The VLab framework: An orchestrator component to support data to knowledge transition
  M Santoro, P Mazzetti, S Nativi - Remote Sensing, 2020
- Big data challenges in building the global earth observation system of systems
  S Nativi, P Mazzetti, M Santoro, F Papeschi, M Craglia… - Environmental Modelling & Software, 2015


Thank you
Dziękuję
Ευχαριστώ
Kiitos
Obrigado
谢谢
Hvala
有り難う
Merci
Danke
Tack
תודה
Thank you
Gracias
Grazie
감사합니다
شكرا
謝謝
ありがとう
Спасибі