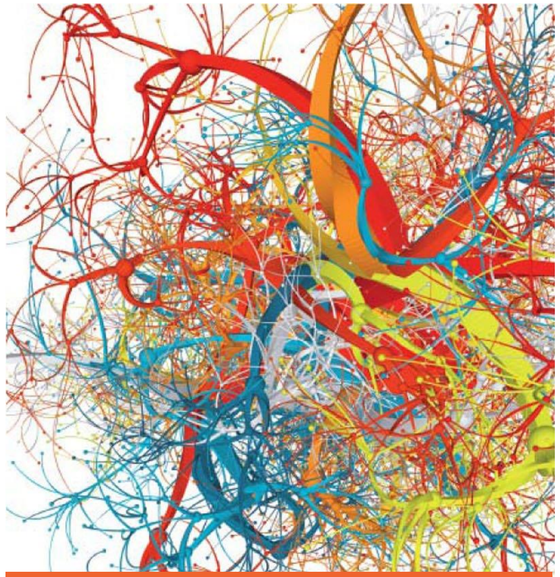
ROLE OF DATA SCIENCE IN FOOD CHAIN SAFETY DECISION MAKING: CURRENT STATUS AND FUTURE TRENDS

ÁKOS JÓŹWIAK
NATIONAL FOOD CHAIN SAFETY OFFICE (NÉBIH), HUNGARY

40 years of RASFF: “All you need is RASFF?” – theme: smarter data = better analysis

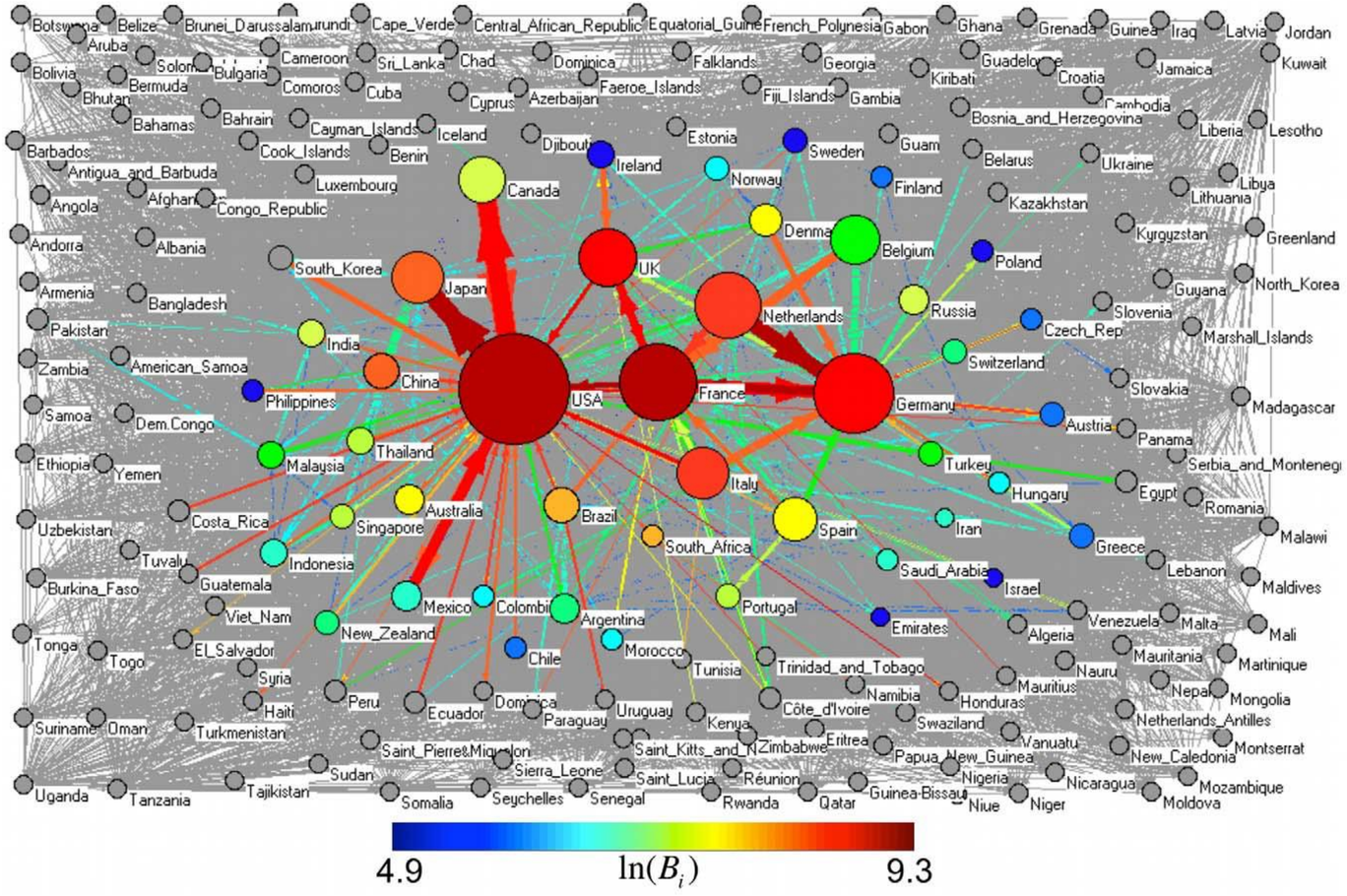
13.12.2019. Brussels



Increasing **volume & complexity** of the food chain

7 COUNTRIES FORM THE CORE OF THE AGRI-FOOD TRADE NETWORK

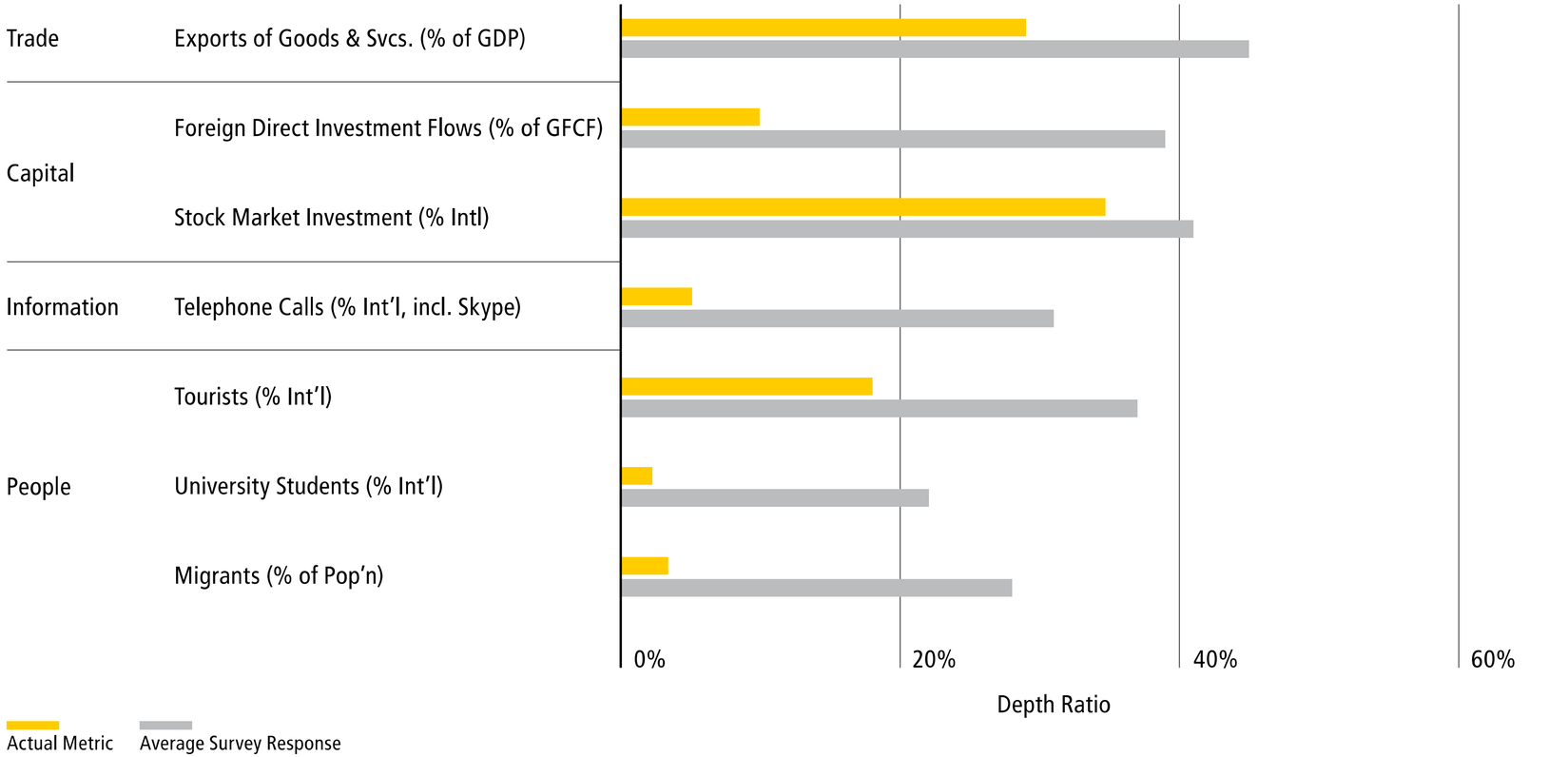
Each trading with over 77% of all the countries in the world



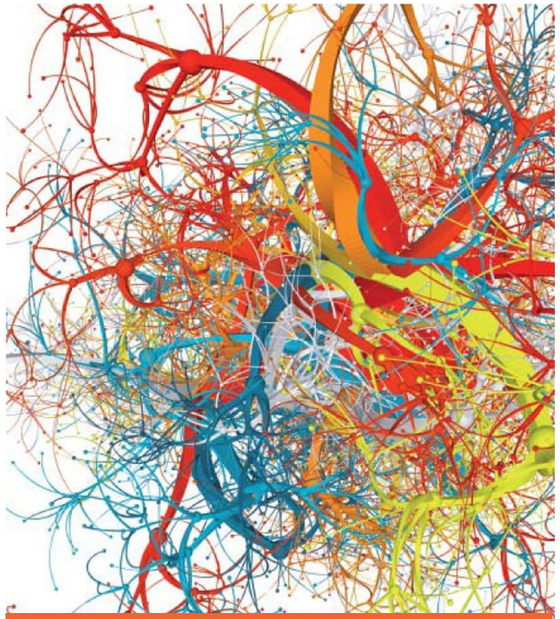
Source: Ercsey-Ravasz M, Toroczkai Z, Lakner Z, Baranyi J (2012) Complexity of the International Agro-Food Trade Network and Its Impact on Food Safety. PLoS ONE 7(5): e37810. doi:10.1371/journal.pone.0037810

THE WORLD WE KNOW IS **MORE LOCAL THAN GLOBAL**

GLOBAL DEPTH MEASURES VERSUS US SURVEY ESTIMATES



Source: Pankaj Ghemawat and Steven A. Altman: DHL GLOBAL CONNECTEDNESS INDEX 2016. The State of Globalization in an Age of Ambiguity



Increasing **volume & complexity** of the food chain

Growth of the amount of **data available** for analysis

INFORMATION **BOOM**

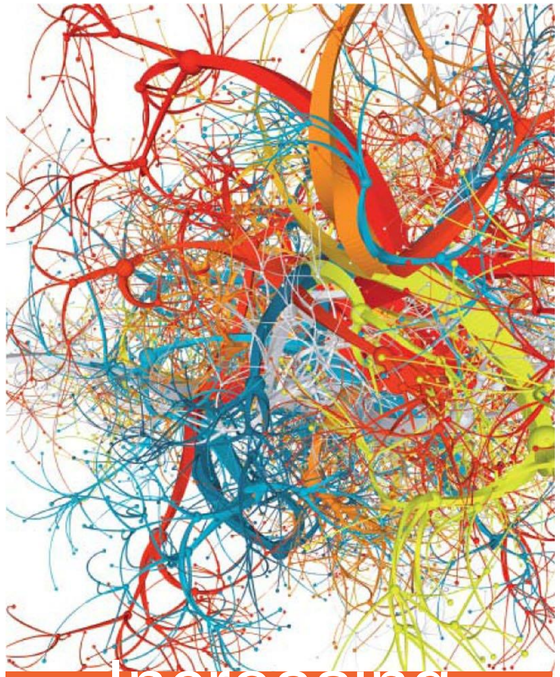
- **From the dawn of the civilization to 2003** humans produced **5 exabytes** of data in total
- IBM has estimated in 2016 that 2.5 exabytes (2.5 million terabytes) of data are produced every day
- Now it is around 5 exabytes **daily**
- We are in the middle of a **transition from a society of facts to a society of data**
- Numbers are being generated much faster than we have any specific use for

Source: https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

POST-NORMAL SCIENCE

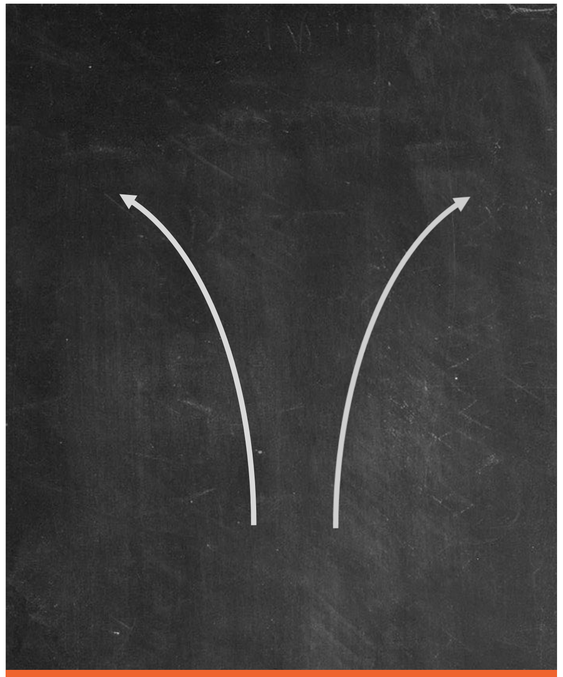
- Policy-related scientific problems:
  - uncertain facts
  - disputes over ethics and values
  - urgent decisions needed
  - that may have far-reaching consequences

- Policy makers are required to make difficult and firm decisions based on data characterized by high levels of uncertainty.



Increasing **volume & complexity** of the food chain

Growth of the amount of **data available** for analysis

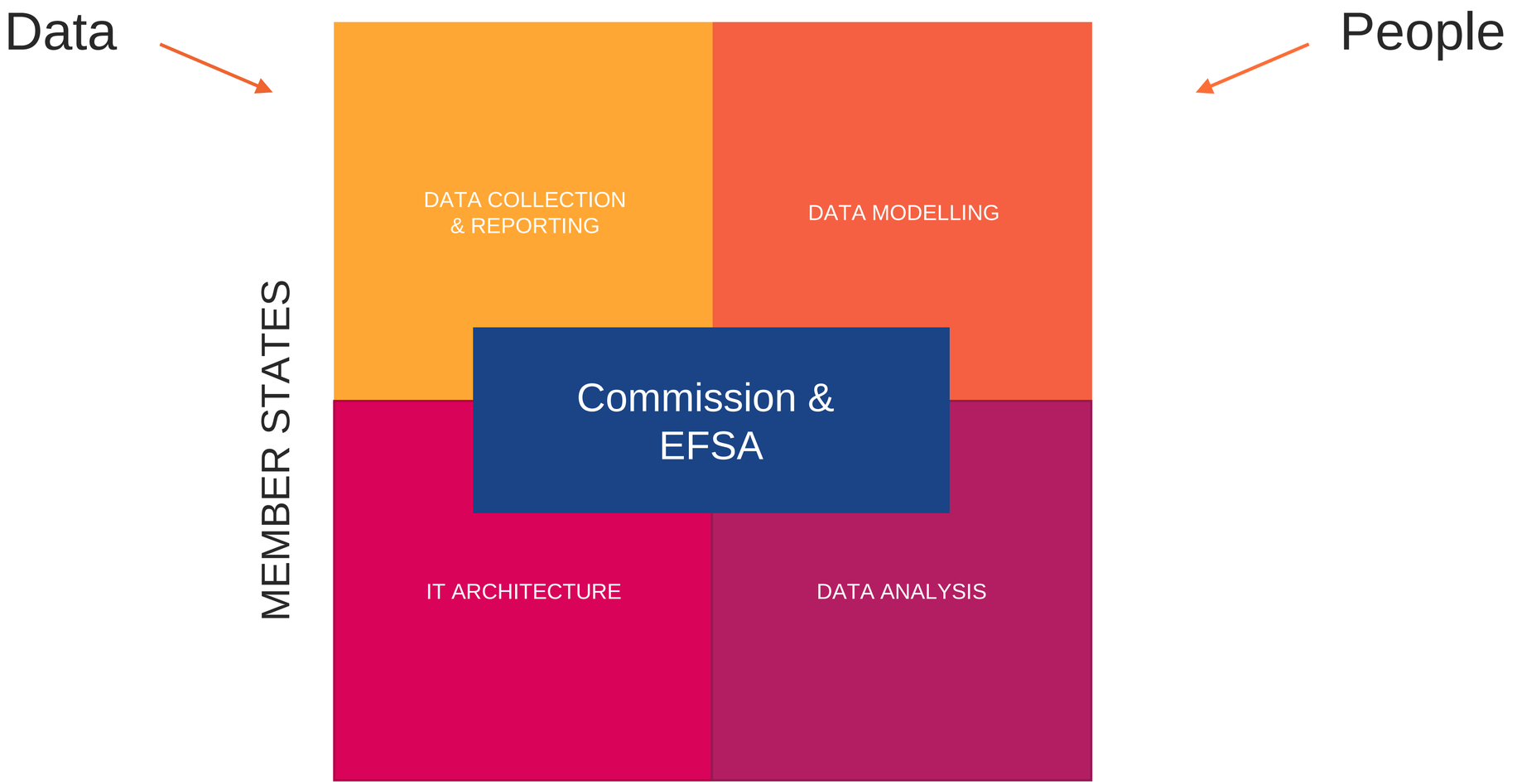


Better evidence-based **decision making?**

COMPUTATIONAL SCIENCE AS A SOLUTION

- Computational science:
  - Able to detect patterns which can not be detected by a smaller set of data
  - Those **emerging patterns** can be surprising & counter-intuitive
- 'more is different’
- **What do we need to achieve this?**

DATA ANALYSIS FRAMEWORK



EFSA ADVISORY FORUM
DATA COLLECTION AND MODELLING TASK FORCE

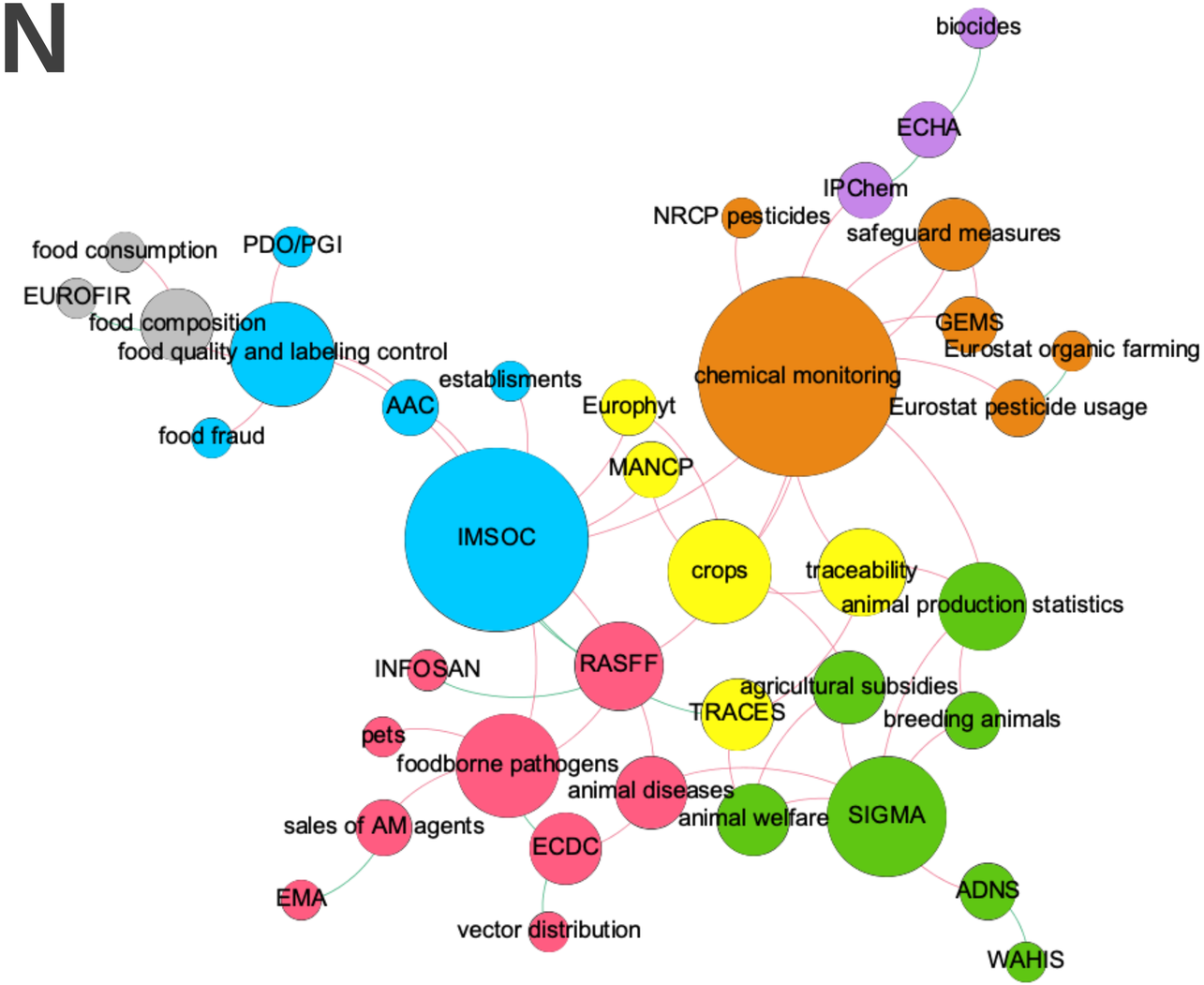
MAIN OBJECTIVE OF THE TASK FORCE

- The main objective of the Task Force is *to overview the data collection and reporting* processes and the *data model* and *IT infrastructure* used, from a ***strategic perspective***, and to formulate recommendations at a strategic level.

DATA COLLECTION

Inventory of MS reporting needs/tasks was mapped (>120)

- 'IMSOC'
- 'chemical hazards'
- 'traceability'
- 'animal health'
- 'microbiological hazards'
- 'ECHA'
- 'food composition'



DATA COLLECTION

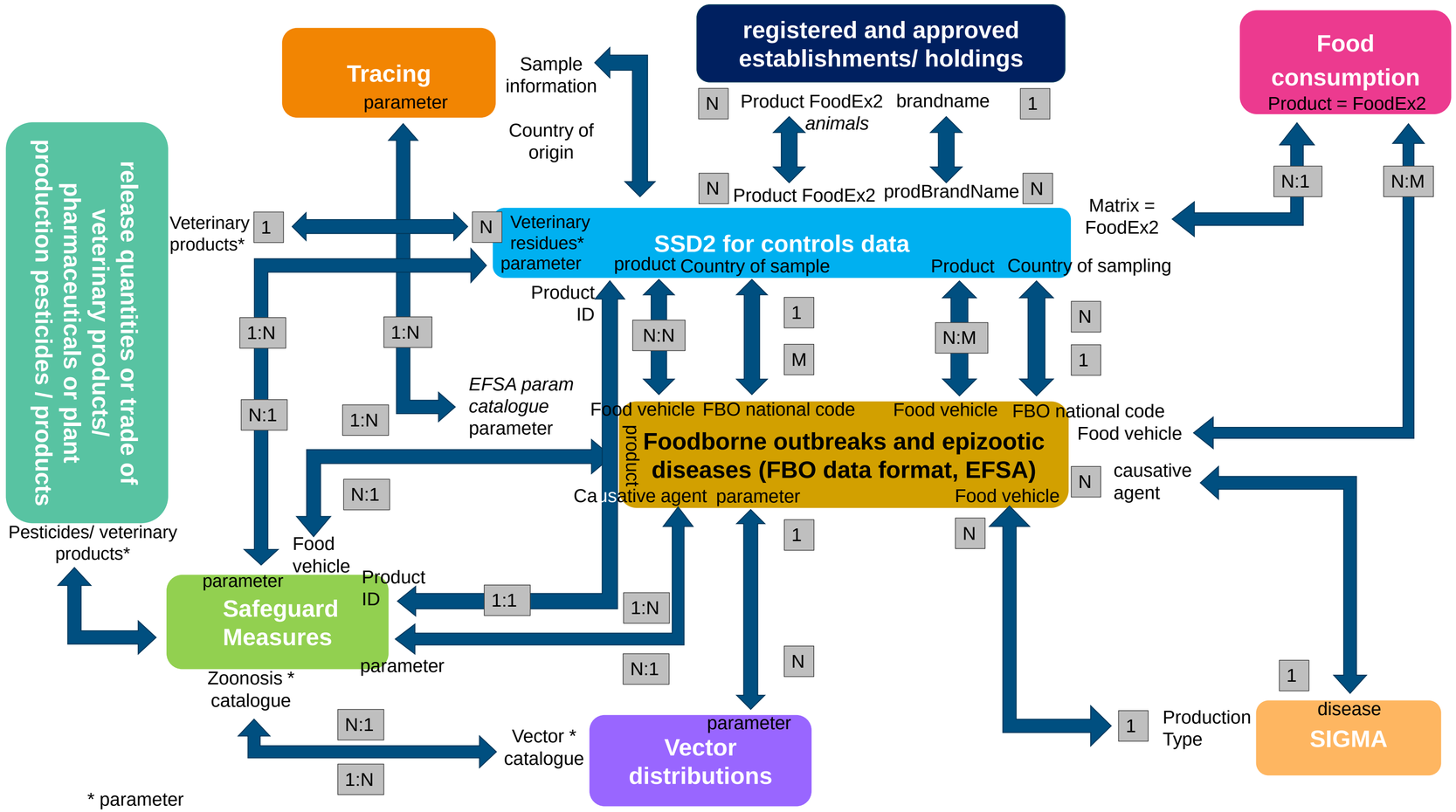
- Important hubs identified:
  - IMSOC, SIGMA, chemical monitoring, foodborne pathogens
- Important modules identified:
  - food composition
  - traceability
- Many similar/overlapping/parallel report flows
- Many connections could/should be improved

DATA MODELLING

- A data model is an abstract model that organizes elements of data and standardizes how they relate to one another and to properties of the real world entities.
- Data models are specified in a data modelling notation, which is often graphical in form.
- Why data modelling is important?
  - Data is an asset of your organization
  - Needs to be understood to be managed
  - Don’t need to look at the detail right away (or sometimes ever)
  - An aid to understanding
  - Provides a common vocabulary

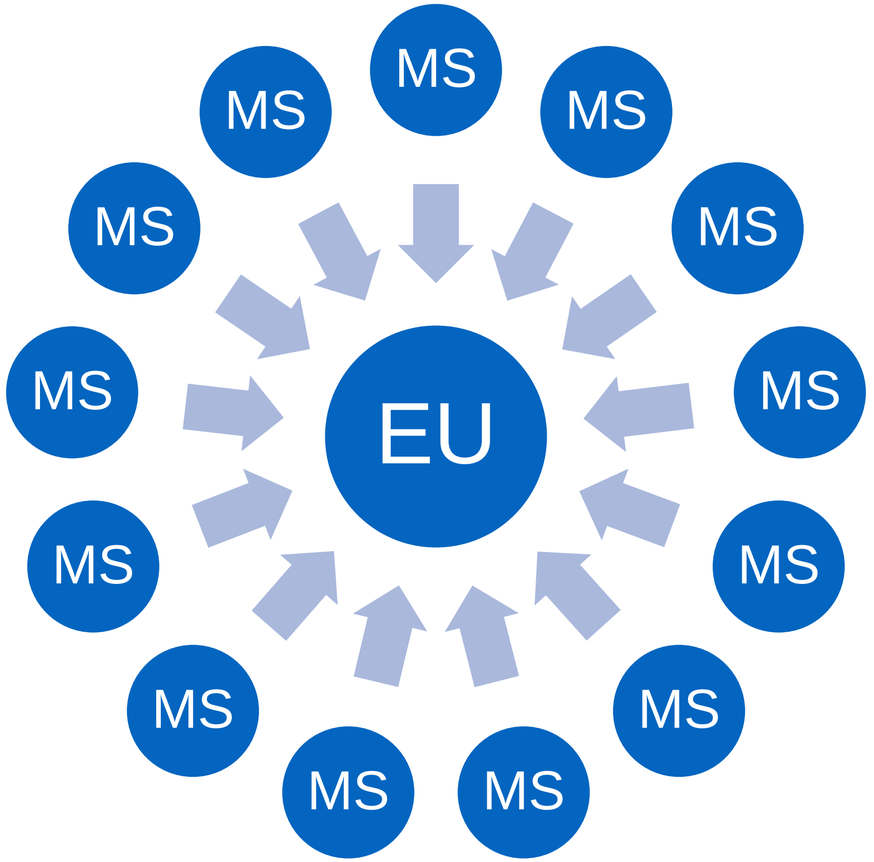
DATA MODELLING

Suggestions for the extension of existing formats and formation of new ones



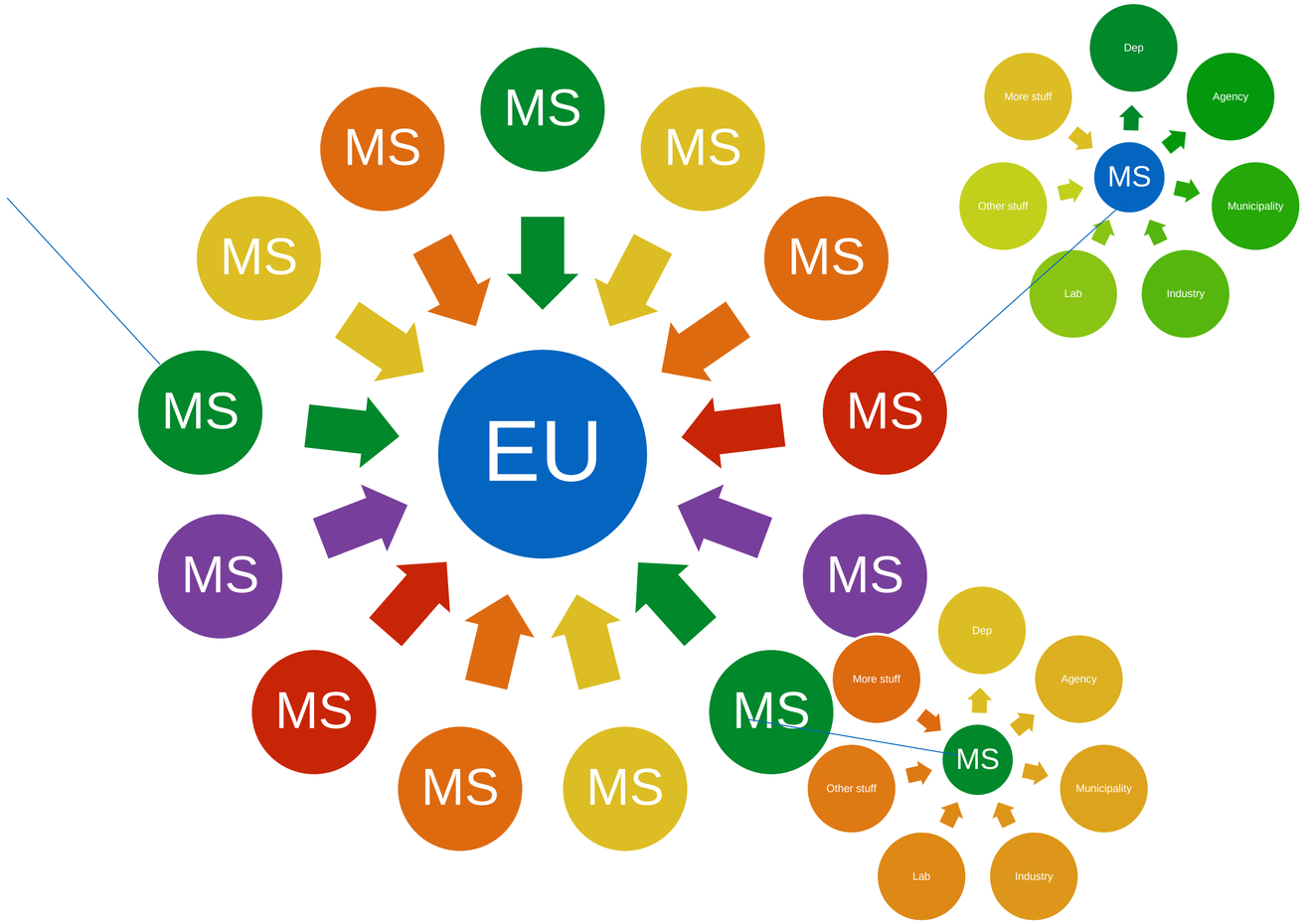
DATA ARCHITECTURE

Traditional paradigm: give me all your data



DATA ARCHITECTURE

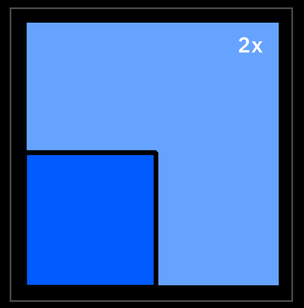
This pattern is repeated in every stage



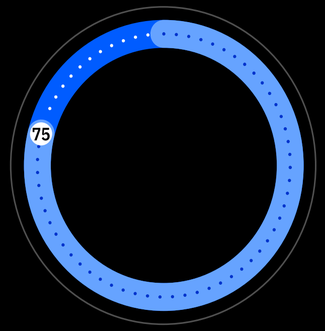
DATA ARCHITECTURE

- Too many data models (pipelines)
  - Data models become obsolete
  - High cost of redesigning
  - Added complexity
- No added value
  - Reporting data becomes a task with no added value
  - Data submitted is a subset of data already processed
- Inefficient use of resources
  - The data is repackaged, wasting resources
  - Clunky reporting mechanism (Sending files. Trying to understand error messages…)

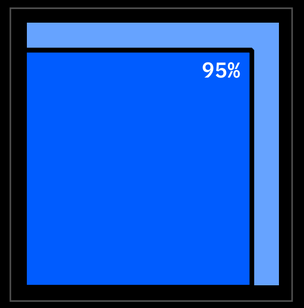
FURTHER **CHALLENGES**



2x increase:
Growth in number of enterprise respondents with over 100 TB of unstructured data between 2016 and 2017.[1]



By 2019, 75 percent
of analytic solutions will incorporate 10 or more exogenous data sources from second-party partners or third-party providers.[2]



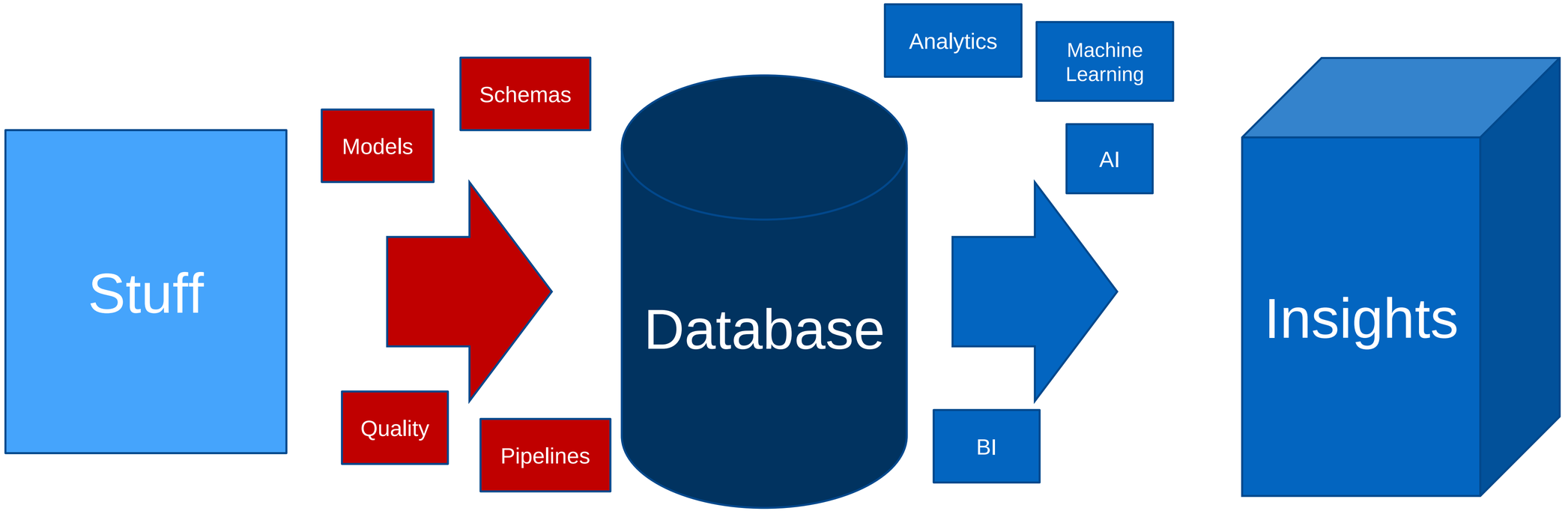
By 2025, real-time
IoT data will make up more than 95%
of real-time data.[3]

[1] Forrester, "Predictions 2018: The Honeymoon for AI Is Over," November 9, 2017.
[2] Crowdflower, "2016 Data Science Report."
[3] IDC for Seagate, "Data Age 2025: The Evolution of Data to Life-Critical," April 2017.

Source: IBM

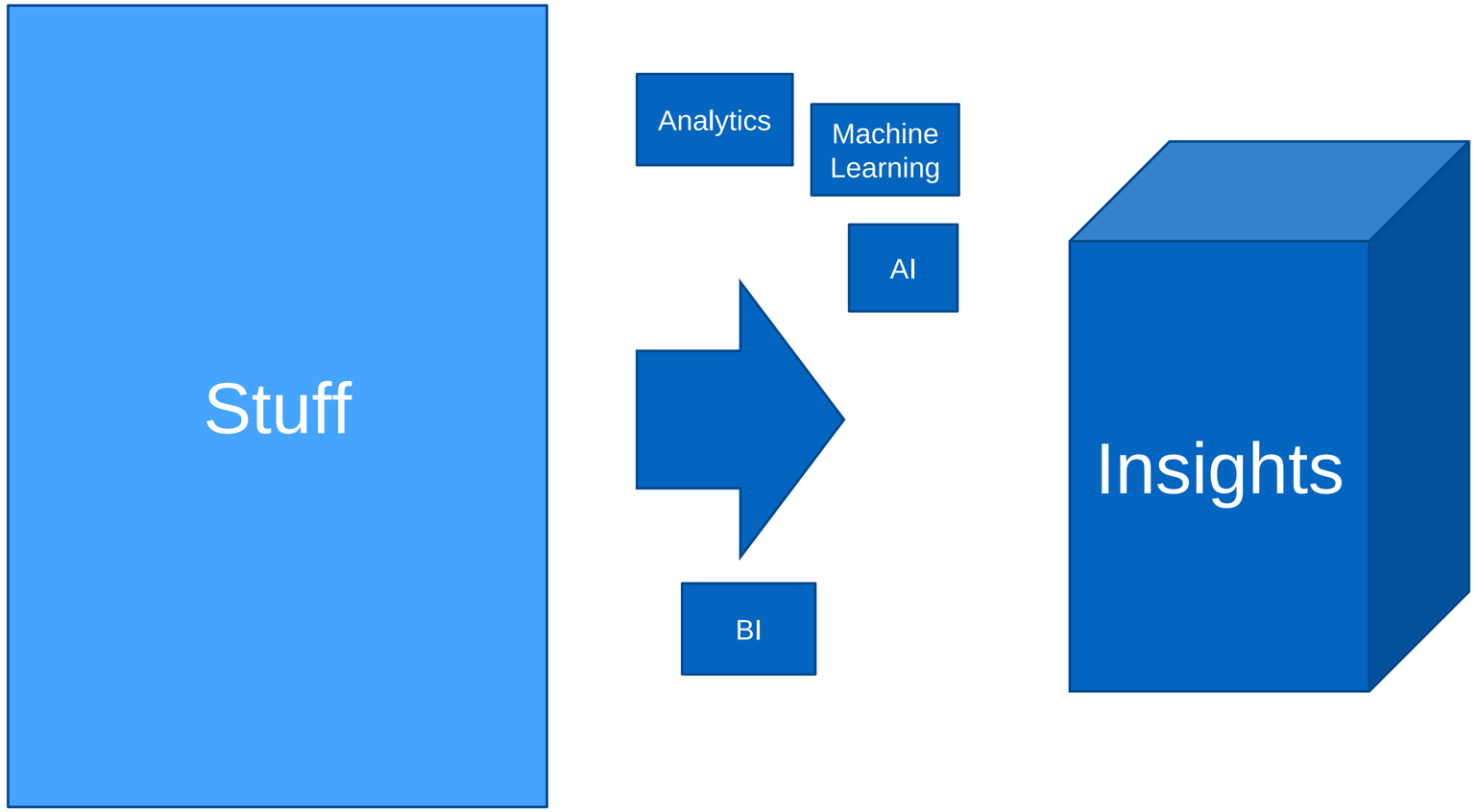
DATA ARCHITECTURE

Current paradigm



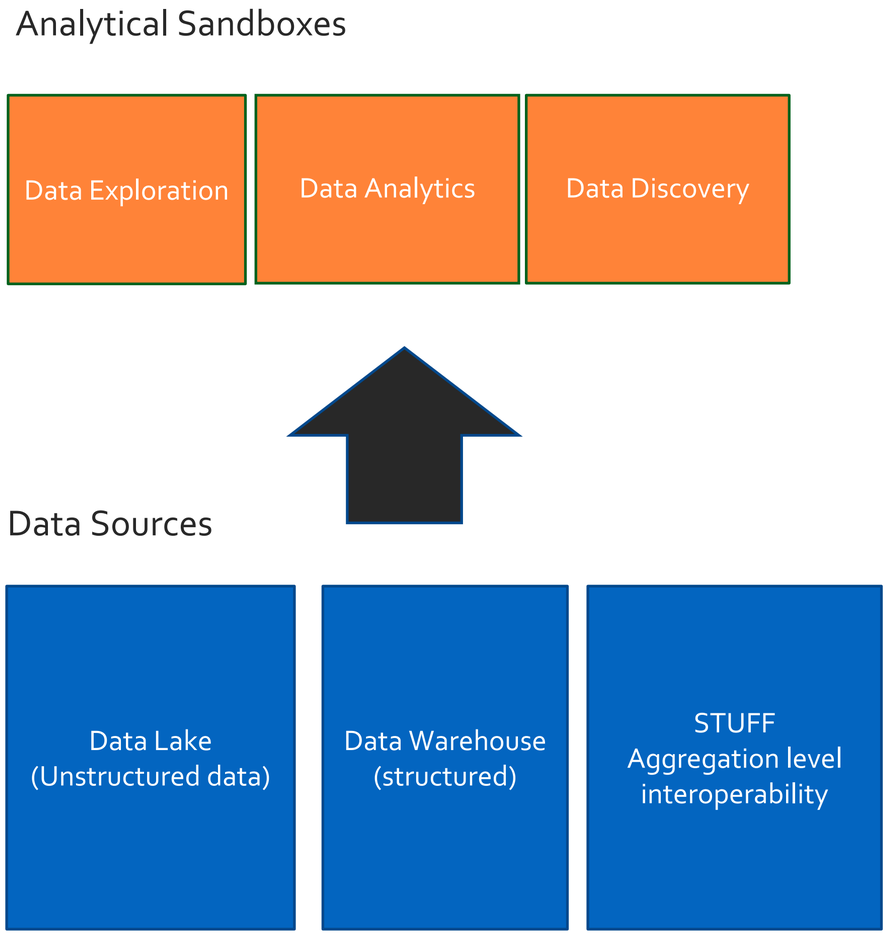
DATA ARCHITECTURE

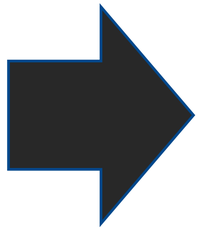
So, what if…

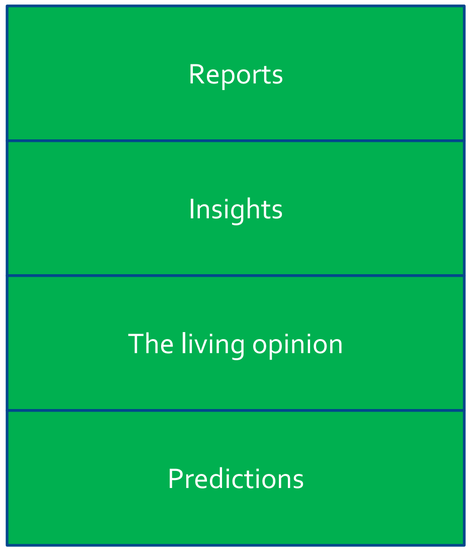


DATA ARCHITECTURE

So, what if…







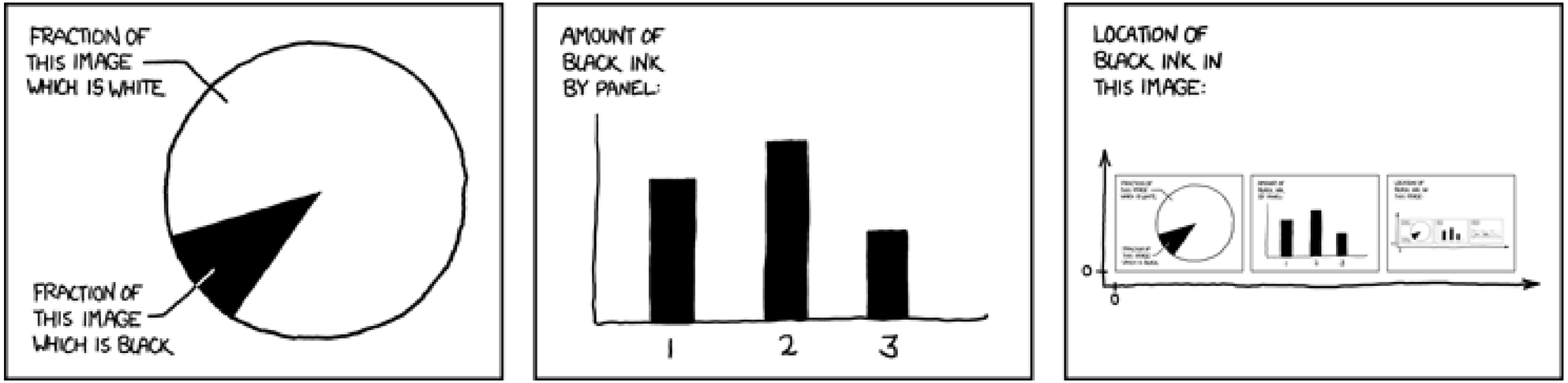
”Smart Ecosystem” or
Connected database

DATA DRIVEN ORGANIZATIONS

DATA DRIVEN ORGANIZATION

- Data is in the core of business activities
- It drives the strategy
- Organizational, procedural, capacity building changes
  - data-informed culture, agile working, room for experiments…
  - more expertise on data is needed → education

PEOPLE



Source: https://xkcd.com/688/

- Creation and development of (big) databases is not only an IT problem
- The ability of analysis and evaluation of input data and results: high-level knowledge of food chain science is needed enabling interpretation and validation.

DATA LITERACY

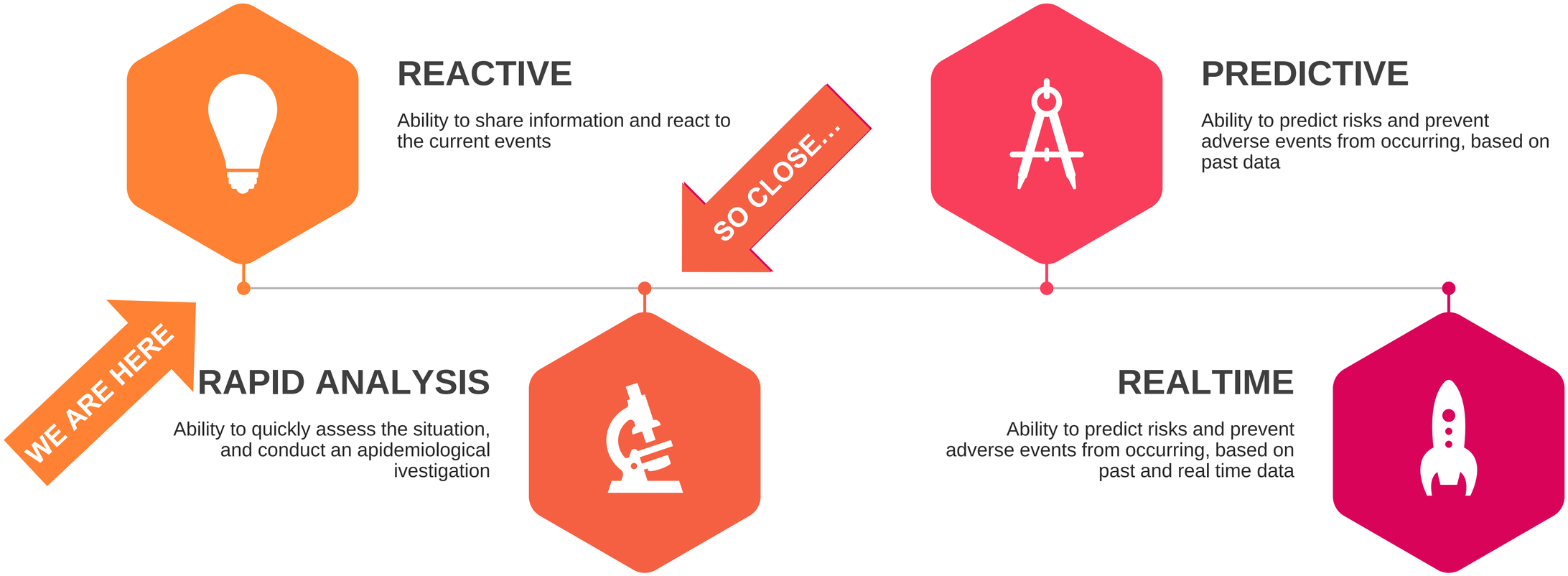
- We need data literate people
- Data literacy: spectrum of related skills
  - MIT: the ability to
    - read
    - work with
    - analyse
    - and argue with data

ARE WE READY?

ARE WE **READY?**

- We need to invest in data sharing and exploring capacities
- We need careful strategic planning for multiple timelines
- Building the future may destroy some of the current investments/achievements
- IT systems become obsolete after 7-10 years
  → build from scratch is better than patching
- Striving for "full" standardization vs Connected Databases
- Expectation Management & Change Management

ARE WE READY **FOR THE FUTURE?**



FOOD SCIENTISTS WILL NOT BE REPLACED BY AI…

...THEY WILL BE REPLACED BY FOOD SCIENTISTS USING AI

CONTACT
Ákos Józwiak
jozwiaka@nebih.gov.hu