UPDATE ON TASK FORCE ON DATA COLLECTION AND DATA MODELLING

72ND MEETING OF THE EFSA ADVISORY FORUM
03.07.2019. REYKJAVÍK

MAIN OBJECTIVE

- 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*.

OUTCOME OF THE 3RD MEETING

- Work in parallel:
  - Cluster 1 (Data collection and reporting) continued
  - Cluster 2 (Data modelling): Germany
  - Cluster 3 (Data architecture): Sweden
- Presentation to the Advisory Forum meeting in Bucharest
  - Need for more precise objectives
  - Need for interim outputs
- 1-day meeting before the 72nd AF meeting in Reykjavik, Iceland
- Publication of meeting minutes

AGENDA OF THE 4TH MEETING OF THE TF
2ND JULY 2019, REYKJAVÍK

- Discussion on Inventory of Reporting Needs (Draft report on Data collection and reporting: Member State reporting in the Food/Feed domain)
- Discussion on Data Modelling
- Discussion on Data Architecture
- Discussion on other topics connected to the work of the TF
  - IMSOC
  - RASFF
  - MANCP annual report
- Discussion on remaining topics to be addressed

AGENDA ITEM 1

DISCUSSION ON THE DRAFT REPORT ON DATA COLLECTION AND REPORTING: MEMBER STATE REPORTING IN THE FOOD/FEED DOMAIN

DATA COLLECTION AND REPORTING:
MEMBER STATE REPORTING IN THE FOOD/FEED DOMAIN

- Draft report elaborated and discussed
- Summarizing the work done so far in the area of data collection and reporting
- **Submission to Advisory Forum for adoption by the end of August ('Interim report')**
- Main conclusions shared now

DATA COLLECTION AND REPORTING:
MEMBER STATE REPORTING IN THE FOOD/FEED DOMAIN

MAIN CONCLUSIONS

- **IMSOC** is and will be one of the most important data hubs related to the safety of the food chain.
- It is a high need for EFSA and the Member States to understand better:
  - the strategic data model for IMSOC, including mechanisms for data inputs and outputs that are useful to the whole food safety community
  - the Commission's and Member States' views on its future development
  - the possible connections to other data systems

DATA COLLECTION AND REPORTING:
MEMBER STATE REPORTING IN THE FOOD/FEED DOMAIN

MAIN CONCLUSIONS

- It is an imperative for EFSA, Member States and the Commission to **collaborate on data models and reporting requirements** during legislation development.
  - This could improve interoperability between data collection systems and support the development of reports for both risk assessment and risk management purposes before finalisation.
  - This could be an important driver of improved efficiency of data exchange across Europe.

DATA COLLECTION AND REPORTING:
MEMBER STATE REPORTING IN THE FOOD/FEED DOMAIN

MAIN CONCLUSIONS

- It is recommended to initiate discussions on the way towards elaborating a common "**European food safety data model**".
  - There is a clear room for discussion and debate on the usefulness and feasibility of a fully standardized, single food safety model
    - a common model would mean common business rules, meaning highly harmonized legislation
  - However, aiming for as much standardized model as possible might be a valid objective.

DATA COLLECTION AND REPORTING:
MEMBER STATE REPORTING IN THE FOOD/FEED DOMAIN

MAIN CONCLUSIONS

- Domains **weakly interconnected** had also been identified
- Careful investigation of strengthening their connections should take place
- Notable examples:
  - Food composition
  - Food consumption
  - Traceability data

DATA COLLECTION AND REPORTING:
MEMBER STATE REPORTING IN THE FOOD/FEED DOMAIN

MAIN CONCLUSIONS

- Future discussions shall also touch **solutions other than relational databases**.
  - The data handling practices currently in place in the food chain safety domain (at Commission, EFSA and also member States) are still using mainly structured relational databases.
  - This paradigm poses increasing need for standardization of different data models.
  - A possible way forward would be to carefully explore other, modern solutions taking into account the advantages and the possible challenges as well.

AGENDA ITEM 2

DISCUSSION ON DATA MODELLING

DATA MODEL

- 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

PRINCIPLES OF THE DISCUSSION DOCUMENT

- Little redundancy with regard to data retention at European level
- Clarity regarding the locations of data storage and their (potential) availability → Data Governance
- Description of data management and storage and the connection between databases
- Unambiguous channels of data transfer
- Simple data transmission to EU organisations, exchange of data between the EU organisations, harmonisation of the data collections
- Standardised catalogues
- Less data formats

EVOLUTION AND ADOPTION OF THE SAMPLES DATA MODEL

- Sample based data format SSD2 extended to control data format also for consumer goods, cosmetics and animal feeds etc
- Differs from food consumption data and others
- Extension to all data areas
- Linkable data models to other data models that need to be linked (not all information needs to be joined)
- Reduction of available data models (unification, same vocabulary = catalogues)
- Description of interfaces and data location
- **Suggestions were made for the extension of existing formats and formation of new ones**

SUGGESTIONS FOR THE EXTENSION OF EXISTING FORMATS AND FORMATION OF NEW ONES

- SSD 2 data model for controls data
- Data model for Tracing
- Data model for Safeguard Measures
- Data model for Food Consumption
- Data model for Release Quantities or Trade of Veterinary Products/Pharmaceuticals and Plant Protection Products/Pesticides
- Data model for Registered and Approved Establishments/ Holdings
- Data model for Foodborne Outbreaks and Epizootic Diseases
- Data model for Vector Distribution
- (New) data format for MANCP reporting

FURTHER STEPS

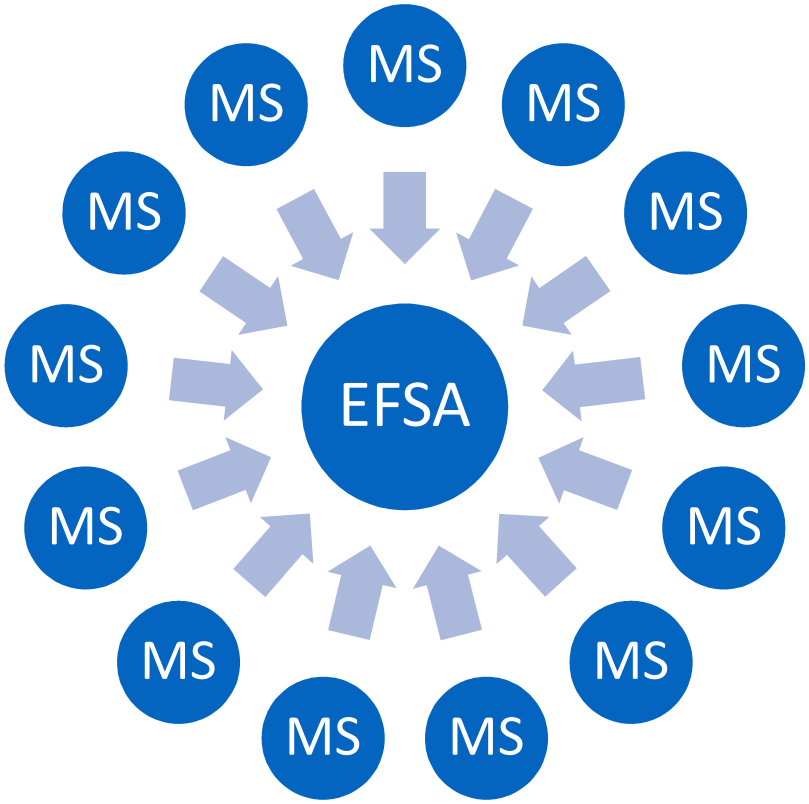
- Checking the models
- Checking the relations of the linkages
- Develop further data models ?
- Find all necessary linkages
- Define core dimensions (e.g. time, location, population, hazard)
- Finally draft data models and adopt catalogues (decide for specific catalogues)
- Describe data governance

AGENDA ITEM 3

DISCUSSION ON DATA ARCHITECTURE

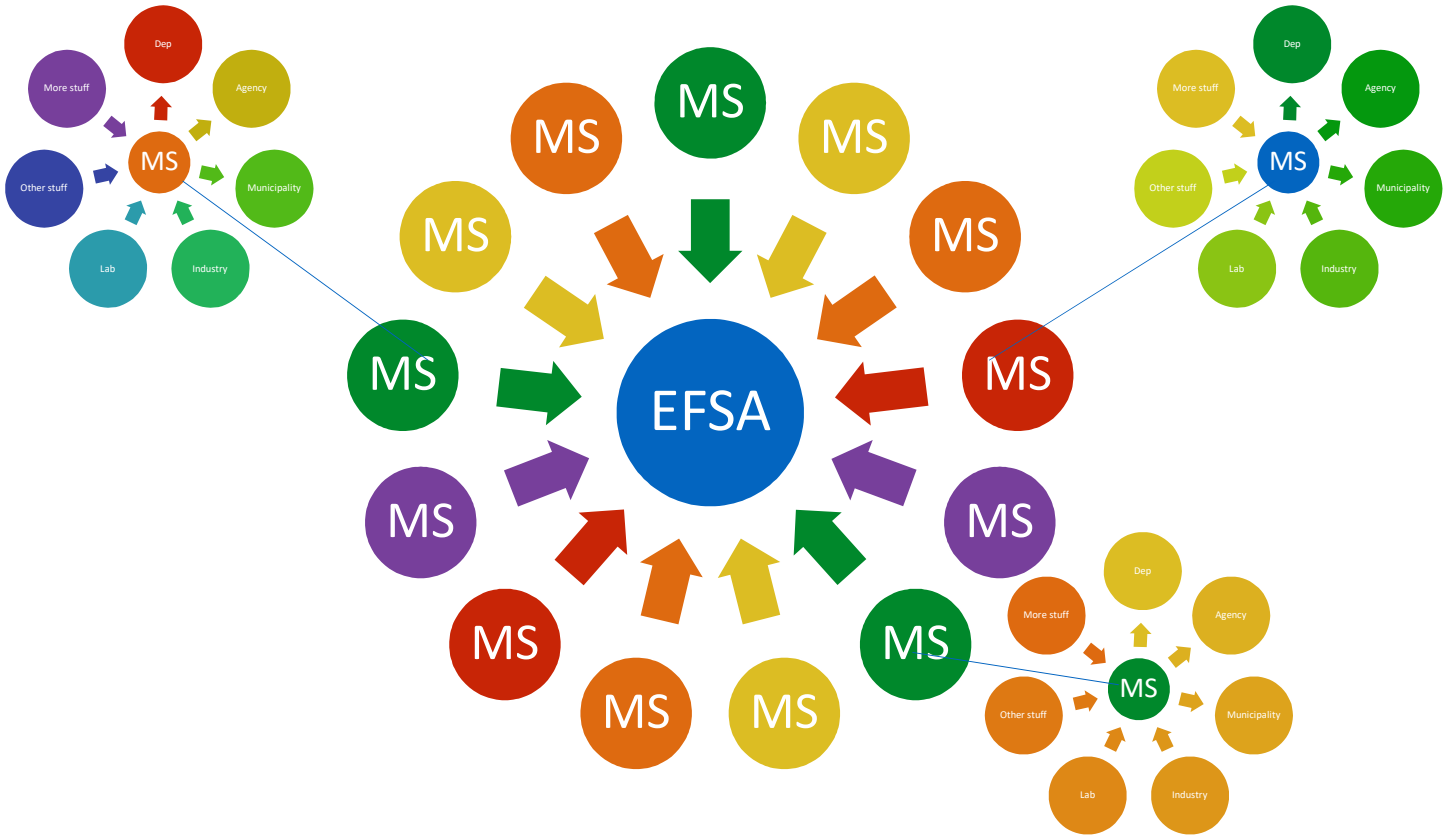
DATA EXCHANGE IN 2027

TRADITIONAL PARADIGM – GIVE ME ALL YOUR DATA



DATA EXCHANGE IN 2027

THIS PATTERN IS REPEATED IN EVERY STAGE

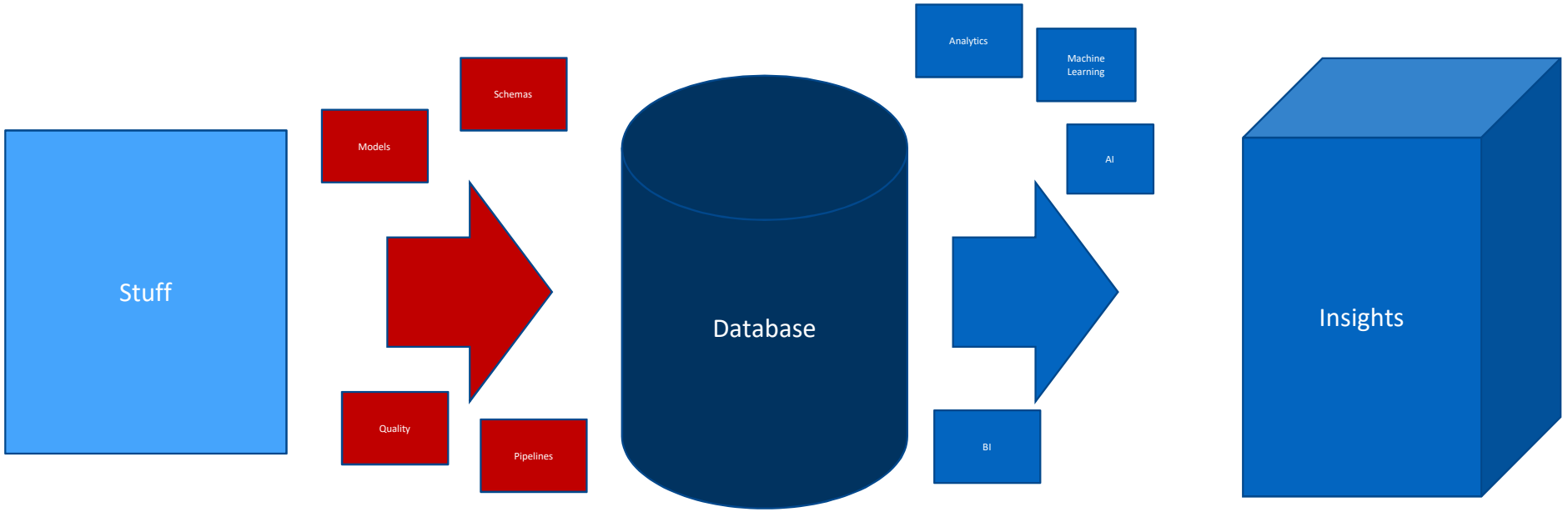


CHALLENGES OF THE CURRENT PARADIGM

- 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…)

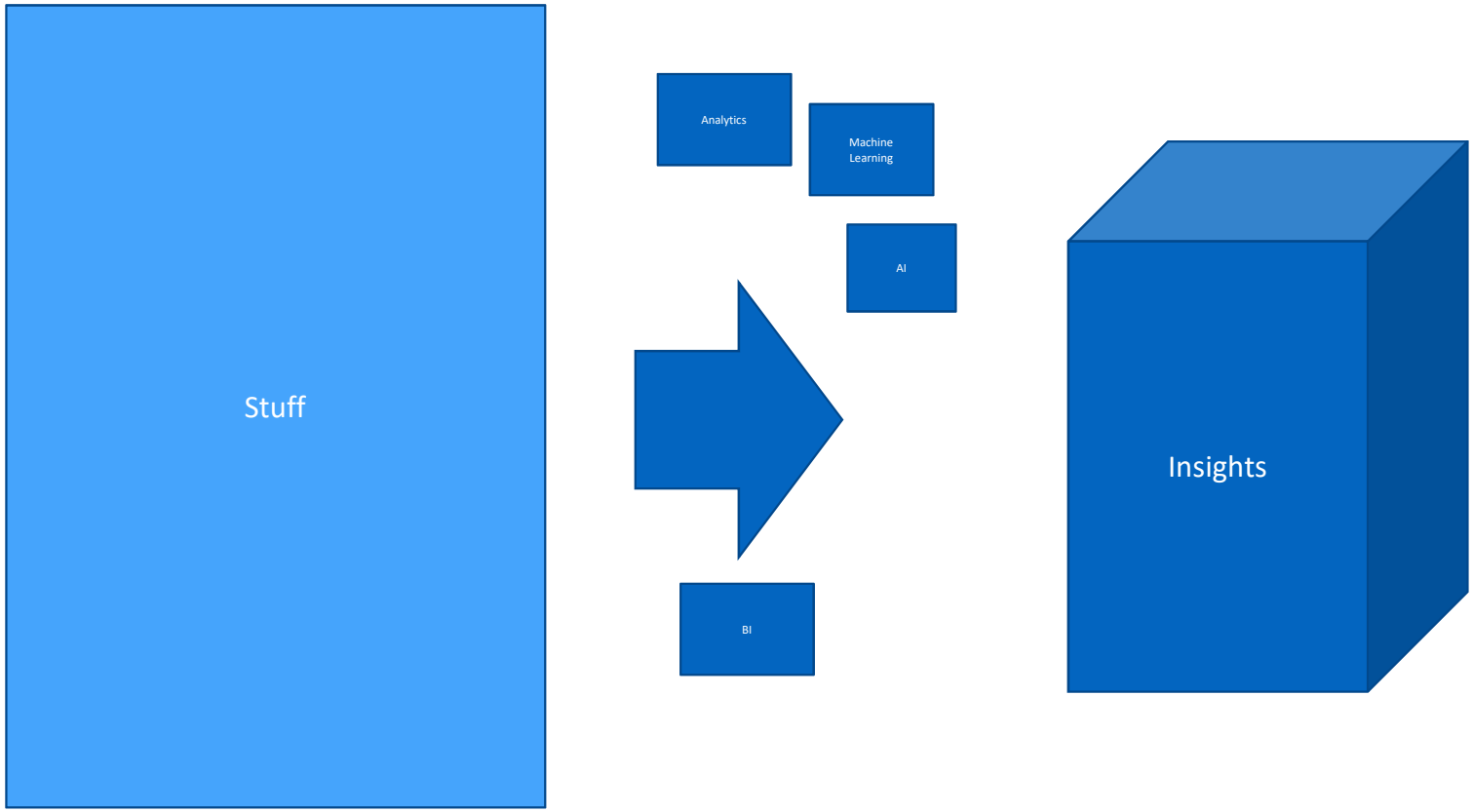
DATA EXCHANGE IN 2027

CURRENT PARADIGM



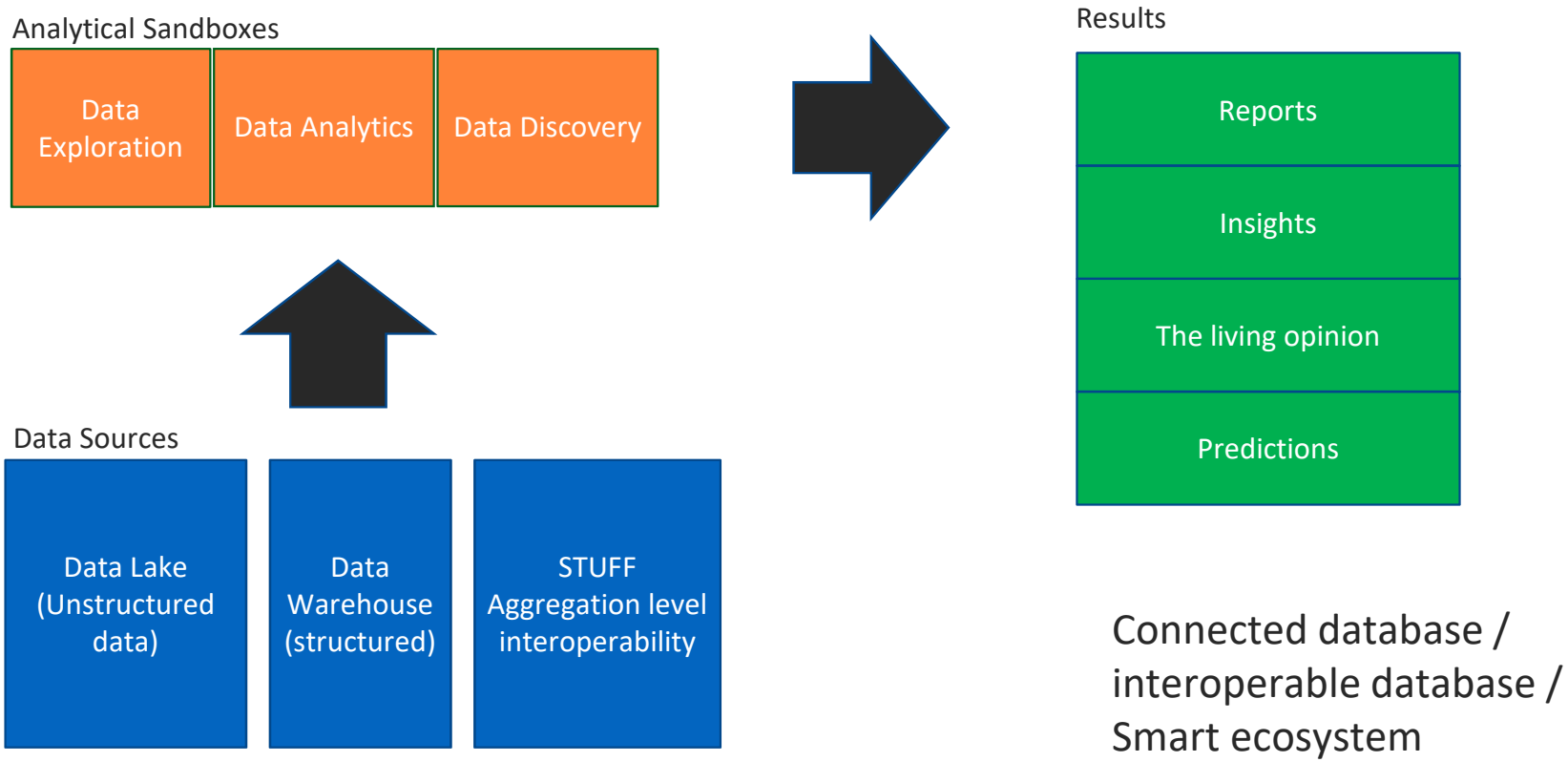
DATA EXCHANGE IN 2027

SO, WHAT IF...



DATA EXCHANGE IN 2027

(NEAR) FUTURE



Connected database / interoperable database / Smart ecosystem

DIFFERENCES BETWEEN CURRENT PARADIGM AND CONNECTED DATABASE

- A schema as used in traditional reporting is tightly bound. Our box will be filled with exactly what we have described.
- In a **connected database**, the schema does not exist until data is read.
- A schema is defined on read, we tell the machine to look for what thing and we (usually) tell it what that thing is.
- This description can be tight or loose.
  - If tight, the results will be better but some items can be missed and it will be time consuming.
  - If loose, the result might contain some other items, but it will be much faster.
- The system can also define it's own schema for finding things by learning what are the characteristics of the things we wanted. This is machine learning.

RECOMMENDATION FOR THE FUTURE ARCHITECTURE OF DATA COLLECTION AS A PROVIDER

- Store as much data as you can
- **Give other parties access to your data**
- Stop reporting/Start connecting
- Let the consumer (e.g. EFSA) handle schemas and model
- Focus on making sense of your own data
- **Let machines do the heavy lifting**
- **More data scientists are needed**

IMPORTANT CONSIDERATIONS

- In a connected database we still need structured data
- **However, full standardization/harmonization is not possible**
- Connecting different databases / data lakes / etc. with different data models could be done through the reading phase
- For this, better analytical sandboxes are needed
- And more data scientists at food safety agencies

AGENDA ITEM 4

DISCUSSION ON OTHER TOPICS CONNECTED TO THE WORK OF THE TF

IMSOC

- Scope - Use and application of IMSOC
- 4 components: iRASFF (RASFF, AAC & Food Fraud), EUROPHYT, TRACES and ADIS
- Details on each component
- Data quality: timeliness and completeness
- Practicalities of using or interacting with the “system”?
- Possible impacts on MS resources, budgets, existing systems etc.
- **More discussion is needed between MS, COM and EFSA on the role and future of IMSOC**

MANCP ANNUAL REPORT

- Template under development for MANCP annual report
- Data on inspections but also data on samples
- Data already collected and available at MS and at EFSA level

RASFF

- Could serve as a case study for the scope of the Task Force
- Many issues identified at risk assessment and risk management domains both at EU and MS level
- Future objectives:
  - Harmonised data capture at the point of sampling
  - Focus on data already collected by MS and FBO (interoperability)
  - Increase the quality of data reported (controlled terminologies and validation rules)
- **Test the applicability of the AF task force recommendations to develop a RASFF data model – towards elaborating a common "*European food safety data model*".**

NEXT STEPS

NEXT STEPS

- Interim report to be finalised by the end of August and to be sent to the Advisory Forum for adoption
- Work on data modelling continues: drafting a written report
- Work on data architecture continues: drafting a written report
- 1-day meeting before the 73rd AF meeting in Helsinki, Finland

THANK YOU!

CONTACT
Ákos Józwiak
jozwiaka@nebih.gov.hu