How data harmonization and data classification foster interoperability and data sharing

Olaf MOSBACH-SCHULZ (EFSA)

European Food Safety Authority (EFSA)
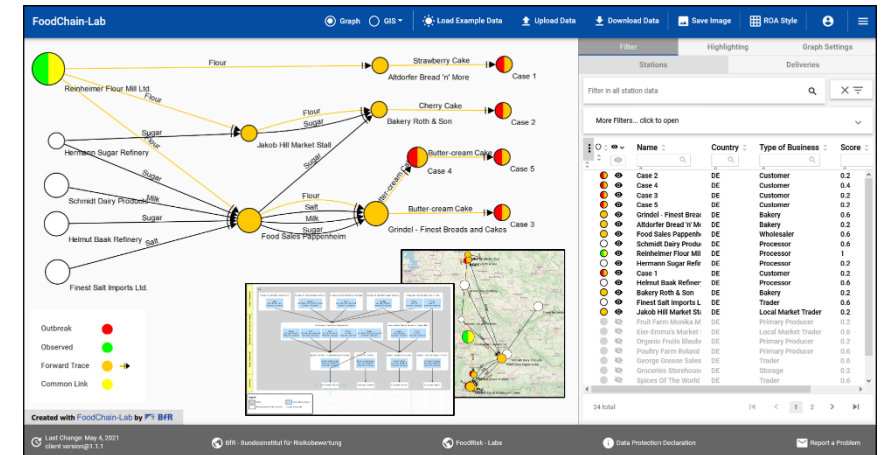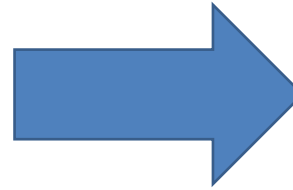Methodological and Scientific Support Unit (MESE)

Disclaimer

**The presentation was drafted
under the sole responsibility of the authors
and is not considered as an EFSA output.**

**The positions and opinions presented
are those of the author alone
and are not intended
to represent the views of EFSA.**

~~Understanding~~ ~~Reconstructing~~ the Food Supply Chain

2011

2025

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Five challenges to address in 2025

1. Increase of convenience food

2. Specialisation of production steps

3. Forced economic competition

4. Increase of disruptions

5. Information flooding

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Conclusion

Need for

- Faster data collection

- Better data exchange

- Data analysis on growing data sets

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025
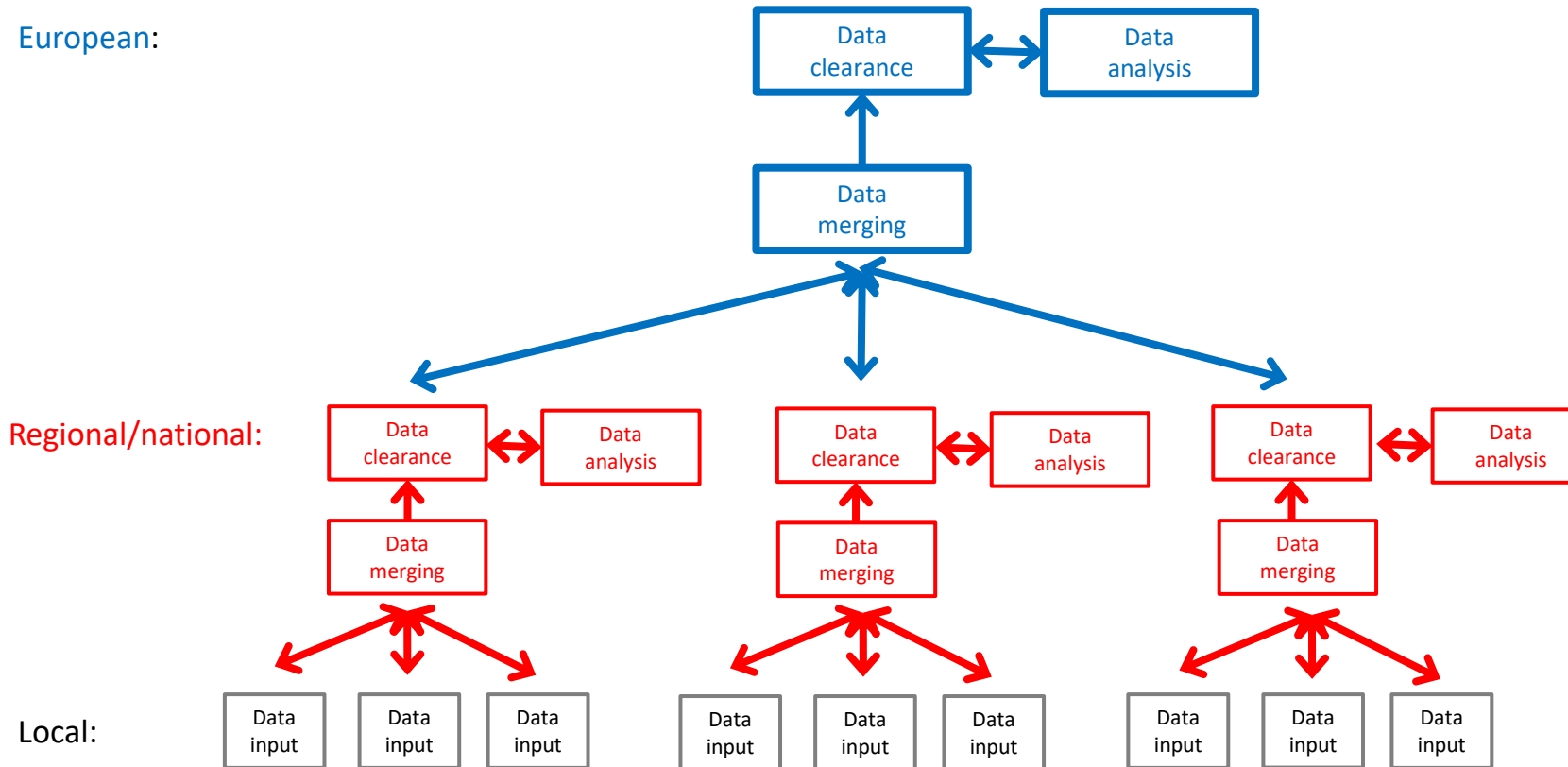
thus…

We have to move

- From an adjusted data collection to a common (distributed) dataset

- From a task of an individual institution to a common group exercise

- From a single solution to an ecosystem of several tools

- From the isolated incident to a system analysis

Universal Traceability data eXchange

UTX
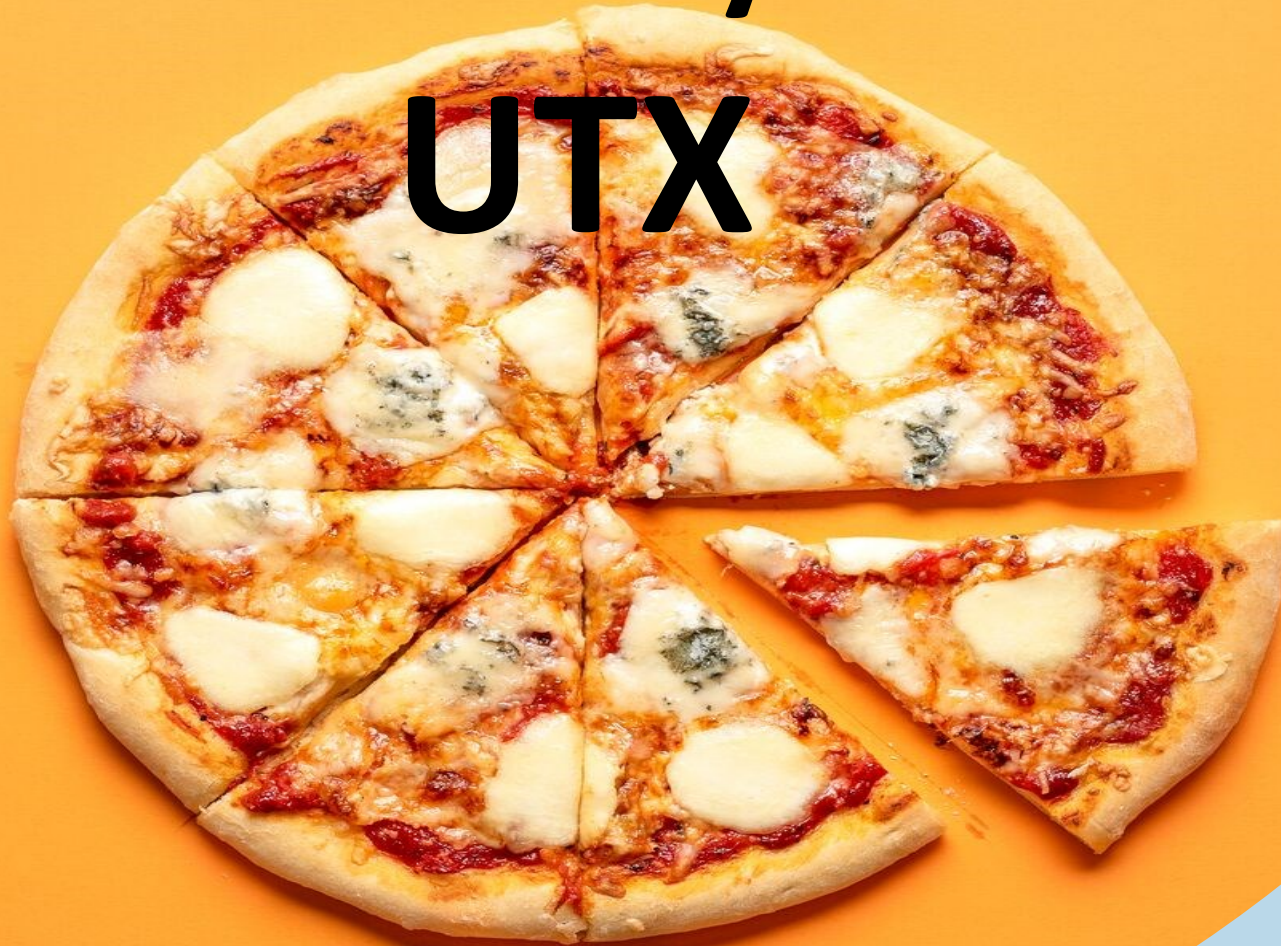
Information workflow of tracing

European:

Regional/national:

Local:

Similar tasks on regional, national, and European level

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Actors

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025
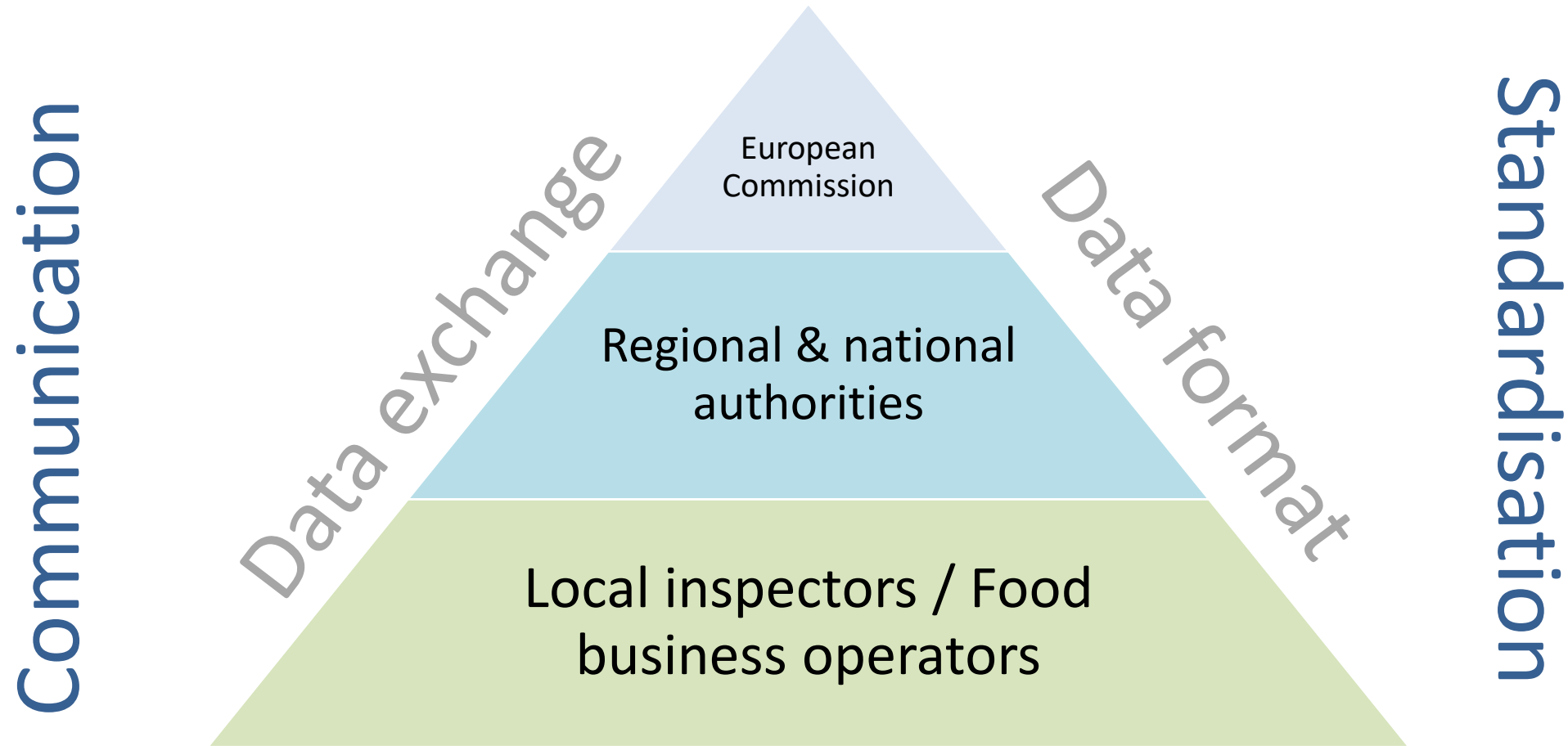
Similar tasks on regional, national, and European level

Similar standards:

For data collection (WHAT?)

For data classification (HOW?)

For data exchange (WHICH FORMAT?)
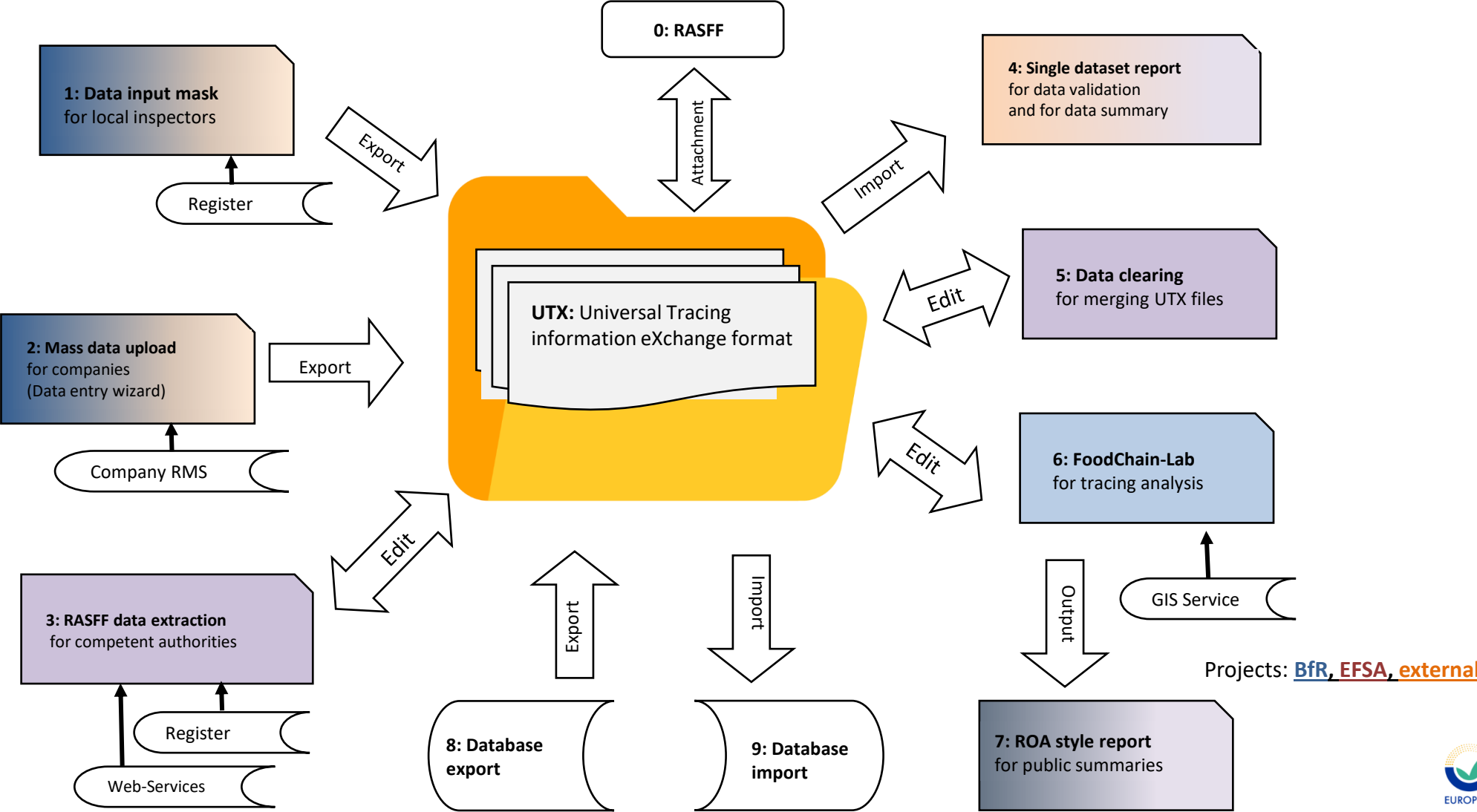
Similar tools:

For data input, e.g. consistency check

For data clearance, e.g. conflict solving

For data analysis, e.g. reporting, FCL

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Similar standards: UTX

0: RASFF

1: Data input mask
for local inspectors

Register

Export

Attachment

4: Single dataset report
for data validation
and for data summary

Import

2: Mass data upload
for companies
(Data entry wizard)

Export

Company RMS

UTX: Universal Tracing
information eXchange format

Edit

5: Data clearing
for merging UTX files

Edit

6: FoodChain-Lab
for tracing analysis

GIS Service

Edit

3: RASFF data extraction
for competent authorities

Register

Web-Services

Export

Import

8: Database export

9: Database import

Output

Projects: **BfR**, **EFSA**, **external**

7: ROA style report
for public summaries

efsa
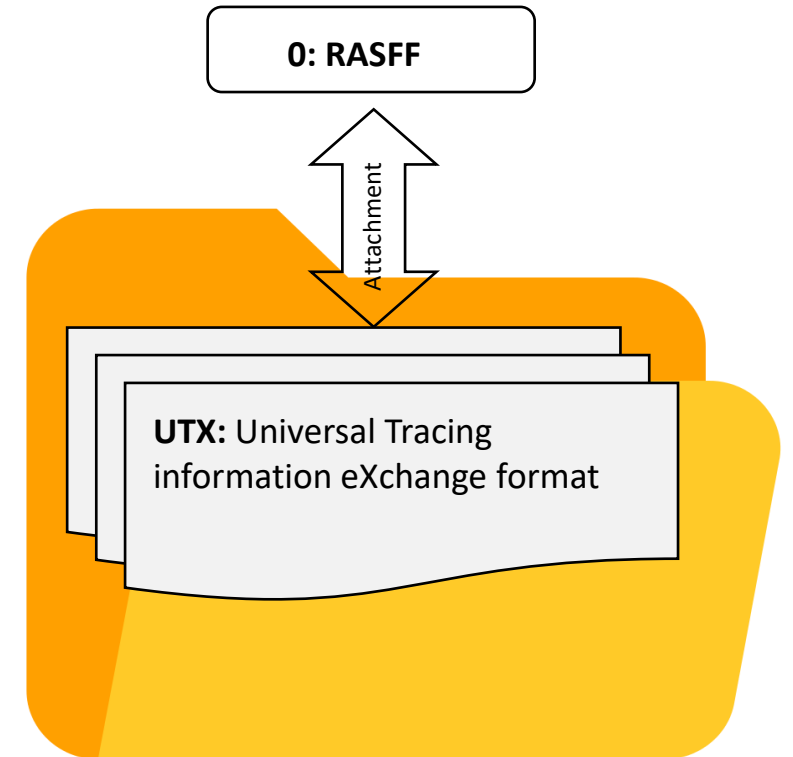EUROPEAN FOOD SAFETY AUTHORITY

BfR

Data exchange

All exchange of tracing data
will be made via attachments
within the RASFF system:

Structured data in json format (UTX file)

Initial, additional and summary files

Versioning of information

Possibility to combine several RASFF notifications,
e.g. outbreaks of similar genotype of the past

0: RASFF

Attachment

UTX: Universal Tracing information eXchange format

UTX main principles

Shareable

Structured

Lightweight

Open

Transparent

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Main principle: Shareable

— UTX serves tracing of items in the food and feed supply chain

— UTX allows co-working with several tools

— UTX parameters and their formats are interoperable to existing tracing systems on European, national, commercial, or food area specific level.

— UTX uses existing standards for tracing

— UTX contains clear-text, as xml or json

— UTX has no data protection

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Main principle: Structured

− UTX is organised in core tables

− UTX core tables are:

  − **Investigation** to collect information

  − **Product** under investigation

  − **Station** in the supply chain

  − **Activity** in the supply chain

− UTX core tables are interoperable

− UTX core tables have hierarchical sub-tables (1:n relations):
e.g. Product → Lot → Tracing unit (TRU)

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Main principle: Lightweight

- UTX stores information on highest level:
  - e.g. Product: "Band name=Coca Cola"
  - e.g. Lot: "Expiry date=31/01/2023"
  - e.g. TRU: "Package unit=1 litre(bottle)"

- UTX uses relations between tables

- UTX stores all necessary information to perform a tracing exercise
  (minimal requirement of information)

- UTX can store all available information,
  which is collected during an investigation
  (maximal extent of information)

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Main principle: Open

UTX specifications are freely accessible

requires information on different levels:

— mandatory → rejection when missing

— recommended → warning when missing

— optional → no action when missing

UTX allows tool specific extensions

UTX can contain extra-core tables

— managed only by specific applications

— non standardised

UTX supports different languages

uses codes, if possible

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Main principle: Transparent

UTX table entries are linked to an information source

UTX allows merging of files

double entries will be cleaned

additional information will be merged

conflicting information will be solved

UTX records all actions on entry

data input, merging, changes

includes versioning of entries

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Data collection tool

UTX will be provided with a specific editor for the end user

The UTX editor has two access modes:

- Read-only → free access
- Editing → information source needed
    - Open a specific entry for editing
    - Submit changes
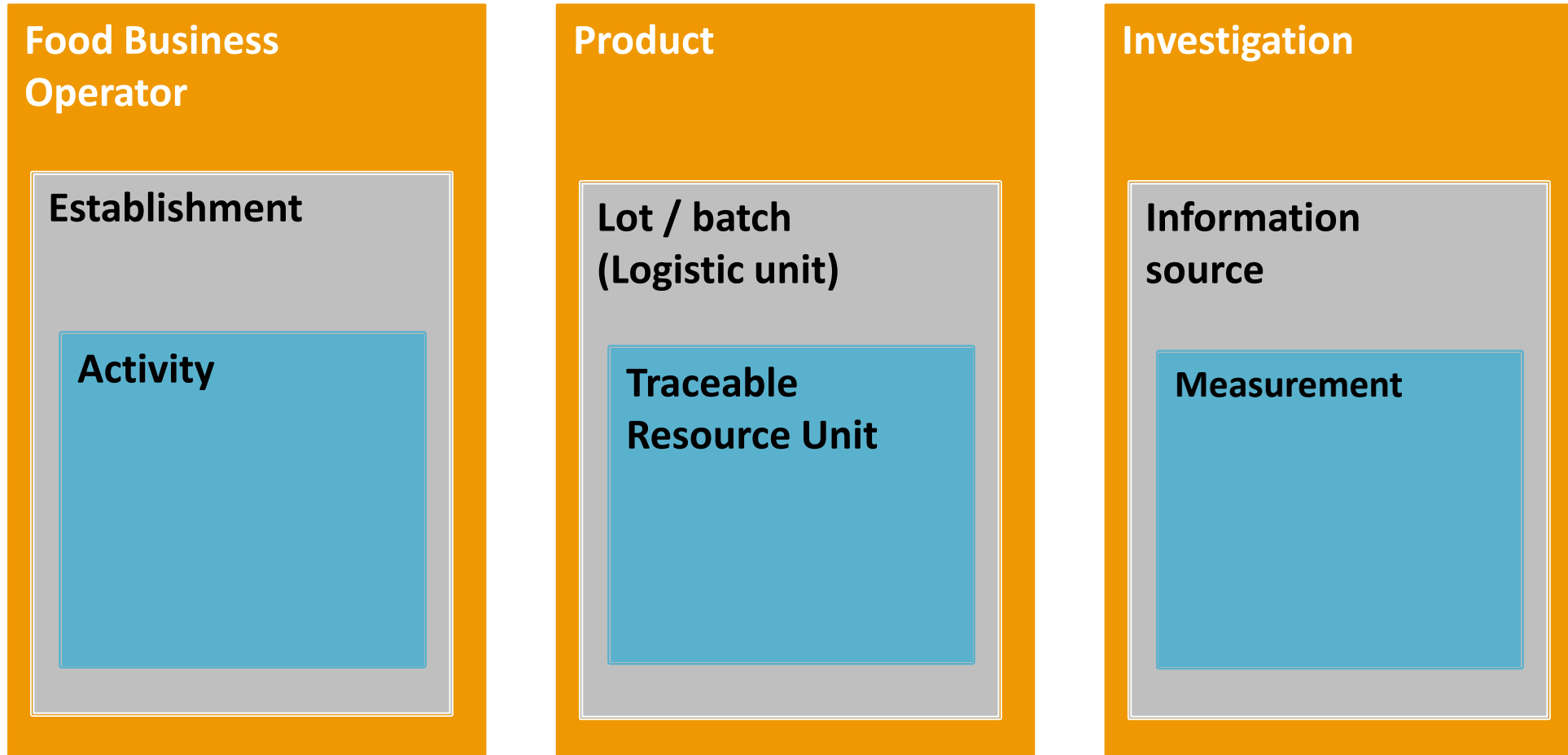
(

Data
Input
Mask

Data Collection VS. Data Extraction

Data collection:

- Targeted data retrieval

- Minimal set of required information

- Correct data format

- On-side data validation

- Guided data collection

Data extraction:

- Use of secondary data

- Minimal rejection of information

- Missing information

- Substitute information

- Conflicting information

- Guided data extraction

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Revised Data Structure: 9 tables

Food Business Operator

Establishment

Activity

Transformation

Product

Lot / batch (Logistic unit)

Traceable Resource Unit

Status

Investigation

Information source

Measurement

Information

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

First main table: Transformation

Station hierarchy:

- **FBO: Administration responsible for the activity, e.g. company headquarter**

- **Station: Establishment, where the activity is performed, e.g. production site, production line. In case of transportation this could be a lorry or a container.**

Activity (no hierarchy): Any action changing the status of an item in the food supply chain, e.g. processing, transport, storage, retail. Trading is considered as activity as it is changing the owner (FBO) of the item.

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Second main table: Status

Product hierarchy:

- **Product:** Description of the kind of item in the food supply chain

- **Lot:** Description of the production process of the item. As special case a logistic unit is counted as lot (collation of items for the purpose of transport)

- **TRU:** Entity of a lot/item

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Third main table: Information

Information hierarchy:

- **Investigation:** Description of a collection of information sources, e.g. an investigation (RASFF notification), a register of companies

- **Information source:** Description of the source of information included in the UTX file, e.g. an inspection (investigation job), a document, a decision/change/ correction (in case of data cleaning)

- **Measurement:** All additional information on the status of a TRU, which is not added in other tables (not necessary for tracing), e.g. measurements of contaminations, temperature etc.

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Detailed UTX description

	A	B	C	D	E	
1	**Group**	**Data element**	**UTX Variable**	**UTX Type**	**UTX Constraints**	**UTX**
2	**Identification**	Activity-ID	id	PK: Text	Unique	
3	**Surrogate parameters** allowing the identification of a station, when the Station-ID is unknown	Description of the activity	activityDescription	Text		
4	**Surrogate parameters** allowing the identification of a station, when the Station-ID is unknown	Station / establishment	stationId	FK: Station.id		
5	**Surrogate parameters** allowing the identification of a station, when the Station-ID is unknown	Start of the activity	activityStart	DateTime		
6	**Surrogate parameters** allowing the identification of a station, when the Station-ID is unknown	End of the activity	activityEnd	DateTime		
7	**Classification**	Type of primary activity	activityPrimaryType	(CV: ActivityPrimaryType AND OR Text)[]	Multiple selection	
8	**Sub-table: Input per status of traceable resource units**	List of possible input TRUs of one product, which are not distinguished in the activity	truInput[i].truIds	FK: TRU.id[]		
	Sub-table: Input per status of traceable resource units	quantity	truInput[i].quantityNumber	Number		

Tabs: Activity | Station | FBO | TRU | Lot | LogisticUnit | Product | InformationSource | Investigation | ...
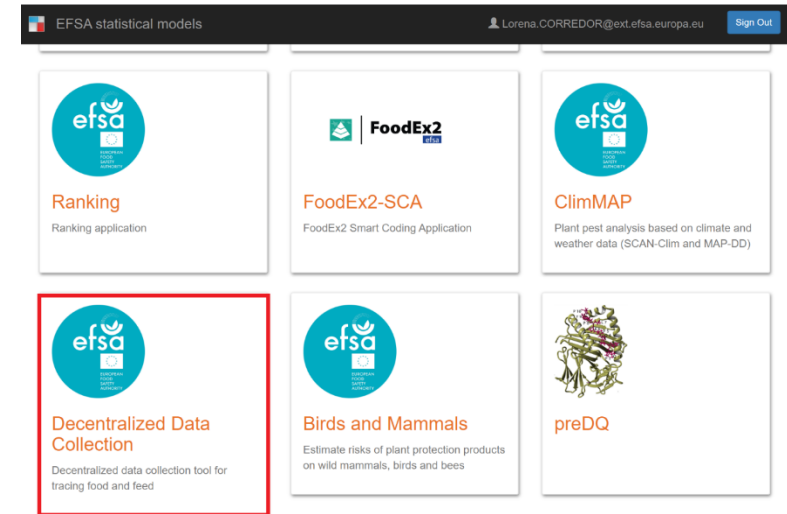
Data Collection tool within the r4eu platform

Easy-to-use data collection

Users: MSs and EFSA

The tool:

- will replace the existing data extractions forms **by a relational database**, resolving most issues of data cleaning and consistency checks

- will work directly on the **"Universal Tracing eXchange" (UTX)** format to make data extraction **interoperable**

- will be able to **directly interchange with FCL software via UTX,** enabling EFSA to perform rapid analyses to identify the source of contamination; incl. full documentation.

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification | ANSES | Paris, 23.01.2025

Food Classification

Adoption of existing standards: existing tools

FoodEx2 efsa

EFSA catalogues

FoodEx2 Smart Coding Application (SCA)

Standard Sample Description 2.0 (SSD2)

International standards
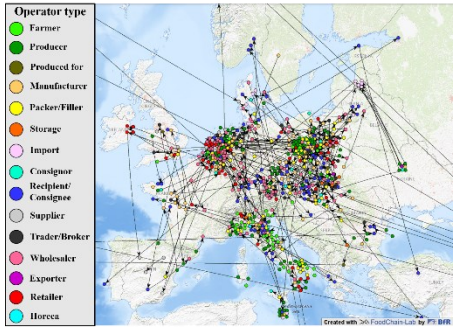
Rapid Alert Supply Network Extractor (RASNEX)

Criteria for the classification system

- Automatic minimizing of errors

- Automatic categorisation of food

- Automatic connection of relations

- Automatic collection of one step forward /
  one step backwards traceability

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Coding example: Case / Client

The case reports:

„Capuccino with oat milk and sugar"

FoodEx2: A068Y „Hot drinks or similar"

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Coding example: Coffee bar

The bar reports the recipe:

	A068Y	
Boiling:	#F28.A07GL	
Drinking water:	$F04.A03DK	
Coffee ground:	$F04.A03GN	
Oat milk:	$F04.A03TL	Producer: „Oatly"
White sugar:	$F04.A032J	

Total code: A068Y(#F28.A07GL
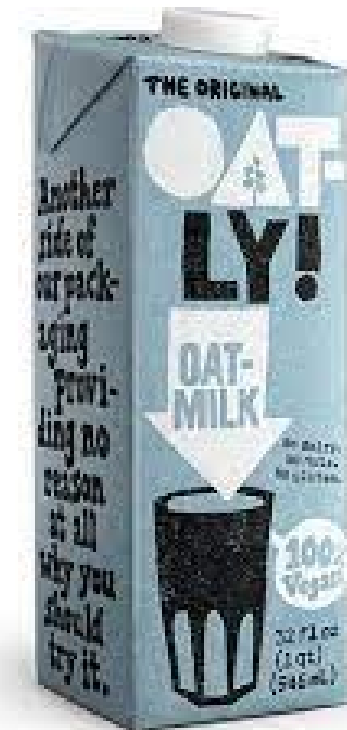 $F04.A03DK$F04.A03GN$F04.A03TL$F04.A032J)

Coding example: Producer „Oatly"

The producer reports the process, ingredients, and package:

	A03TL
Milling:	#F28.A0C03
Drinking water:	$F04.A03DK
Oat grain:	$F04.A000G
Rape seed oil:	$F04.A036V
Salt:	$F04.A042P
Vitamin B12:	$F04.A0EXP
Regulator (dipotassium phosphate):	$F04.A048C
Laminate:	$F19.A07PS

Coding example: Tracing code

Total code:

A068Y   (#F28.A07GL

$F04.A03DK $F04.A03GN $F04.A03TL

(#F28.A0C03

$F04.A03DK$F04.A000G$F04.A036V$F04.A042P

$F04.A042P$F04.A0EXP$F04.A048C$F19.A07PS)

$F04.A032J)

The initial code „A068Y" is twice enriched by the actual recipe of the bar and the producer.

Enrichment will be done by the tracing software, when additional information is available.

Advantages

- Automatic minimizing of errors → Known taxonomy

- Automatic categorisation of food → FoodEx2 tools available

- Automatic connection of relations → Integrated relations (Fazets)

- Automatic collection of one step forward / → Integrated enrichment
  one step backwards traceability

One concept for the full food chain (primary to complex products)

Inclusion of processing

Inclusion of additives, service materials, packaging

Olaf Mosbach-Schulz | FCL workshop: Data harmonisation & classification| ANSES | Paris, 23.01.2025

Thank you
for your attention

European Food Safety Authority (EFSA)
Methodology and Scientific Support Unit (MESE)
Olaf Mosbach-Schulz   olaf.mosbach-schulz@efsa.europa.eu

BfR
Bundesinstitut für Risikobewertung

(Thanks to an unknown lady who permitted this photograph of her tattoo, 2016, photograph by Olaf Mosbach-Schulz)