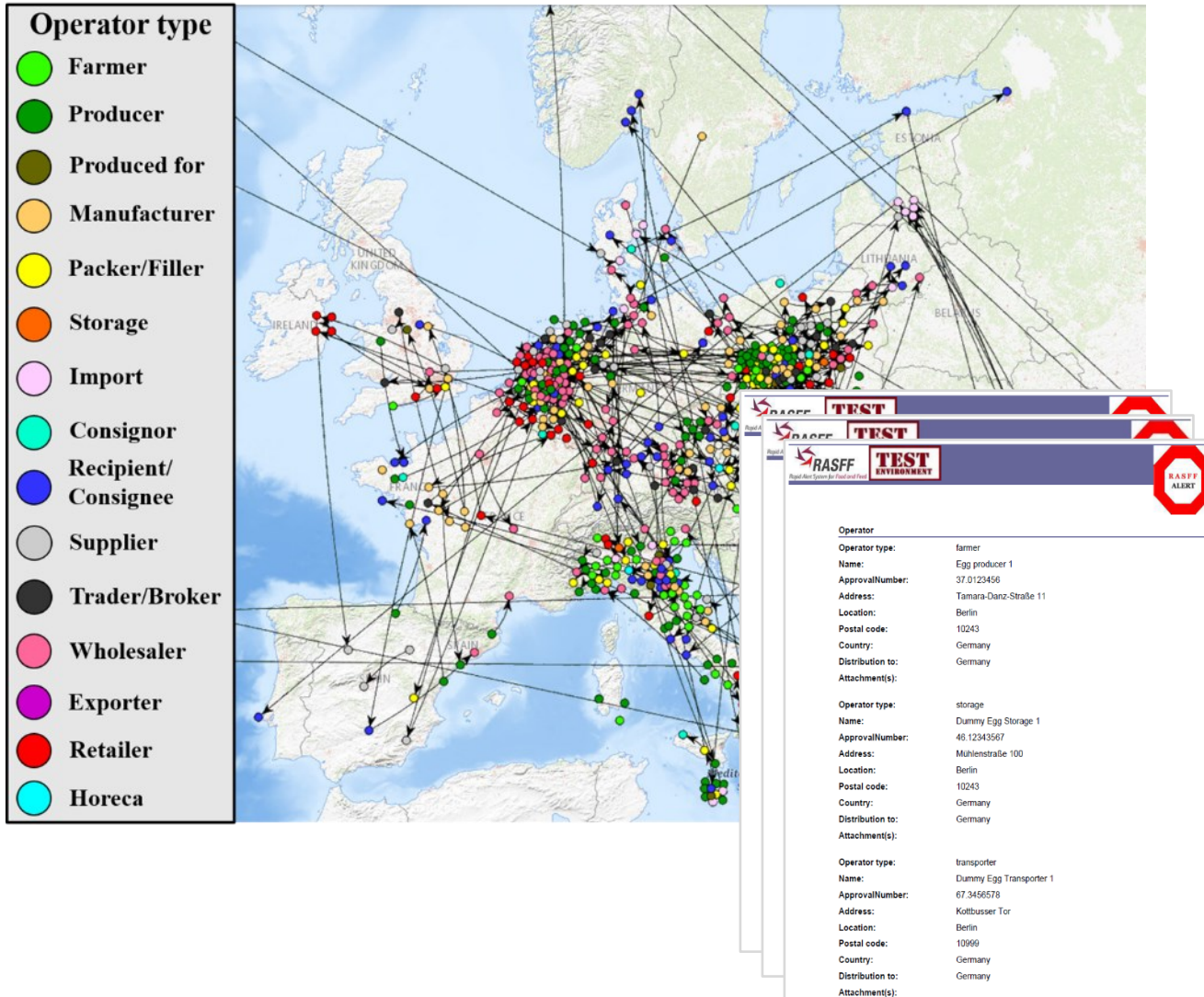


AI-supported data extraction: The Rapid Alert Supply Network EXtractor (RASNEX) tool for mining supply chain information

23.01.2025

Daria Savvateeva, Matthew Salewski & Marc C. Lorenzen

Contamination in the Supply Chain



Need to stop outbreaks as fast as possible!

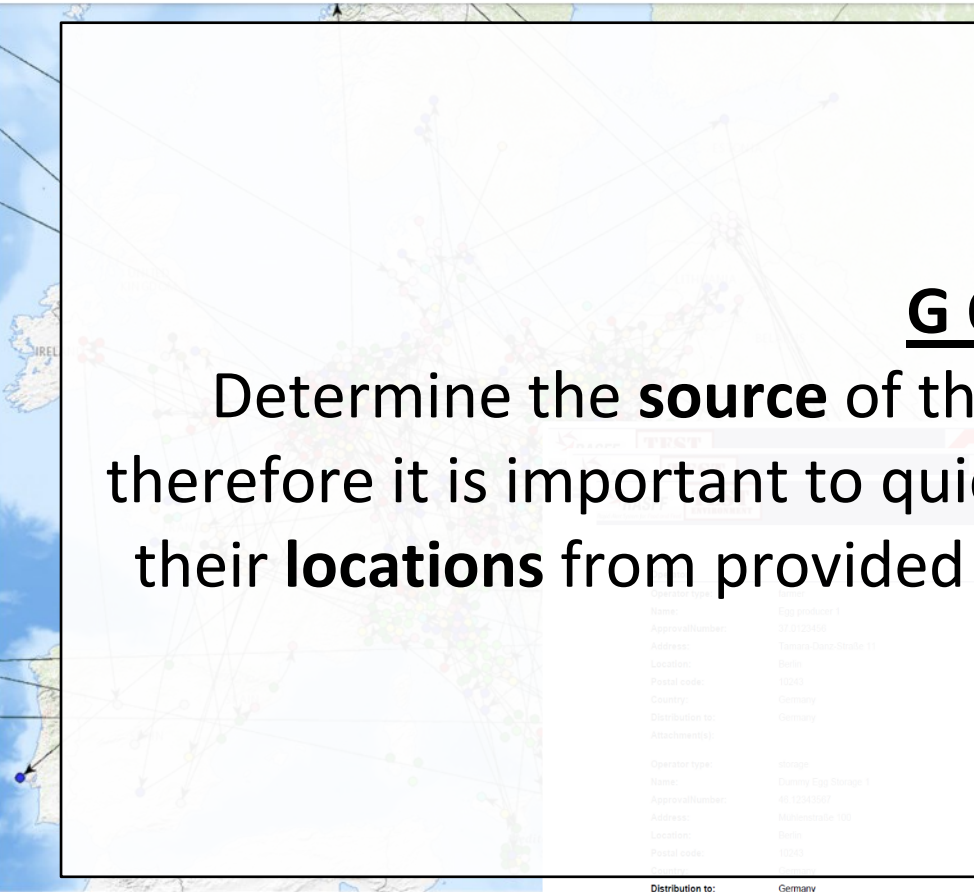
Rapid Alert System for Food and Feed (RASFF) provides relevant information

Collected information (manual work)

- laboratory and transport data
- time, locations, quantities
- Food Business Operators (FBOs) and affected products

Contamination in the Supply Chain

- Operator type**
- Farmer
 - Producer
 - Produced for
 - Manufacturer
 - Packer/Filler
 - Storage
 - Import
 - Consignor
 - Recipient/Consignee
 - Supplier
 - Trader/Broker
 - Wholesaler
 - Exporter
 - Retailer
 - Horeca



During an outbreak need to stop it as fast as possible!

GOAL

Determine the **source** of the contamination to **resolve** it, therefore it is important to quickly determine the **operators** and their **locations** from provided information contained in RASFF.

Collected information (manually)

- laboratory and transport data
- time, locations, quantities
- Food Business Operators (FBOs) and their products

Name: Egg producer 1
 ApprovalNumber: 67 345678
 Address: Kottbusser Tor
 Location: Berlin
 Postal code: 10969
 Country: Germany
 Distribution to: Germany
 Attachment(s):

Operator type: transporter
 Name: Dummy Egg Transporter 1
 ApprovalNumber: 67 3456578
 Address: Kottbusser Tor
 Location: Berlin
 Postal code: 10969
 Country: Germany
 Distribution to: Germany
 Attachment(s):

Distribution to: Germany
 Attachment(s):

Operator type: transporter
 Name: Dummy Egg Transporter 1
 ApprovalNumber: 67 3456578
 Address: Kottbusser Tor
 Location: Berlin
 Postal code: 10969
 Country: Germany
 Distribution to: Germany
 Attachment(s):

Rapid Alert System for Food and Feed (RASFF) notifications





food - alert notification - company's own check - eggs - Germany - no definition_fipronilno definition_fipronil
 Notification Link <https://webgate.acceptance.ec.europa.eu/rasff/notification/view/348541>

General Information

Notification number: 348541
 Reference: food
 Notification type: company's own check
 Notification basis: alert notification
 Product name: eggs
 Product category: eggs and egg products
 Notifying country: Germany
 Notifying Country region: BfR
 CP Reference: 15/07/2020
 Subject: fipronil dummy contamination
 Infosan Informed: No
 eCommerce Related: No
 Mutual recognition: No
 Notification status: DRAFT
 Date of notification: 15/07/2020 10:31:37
 Follow-up status: 0
 Number of follow-ups: 0

Distribution Status

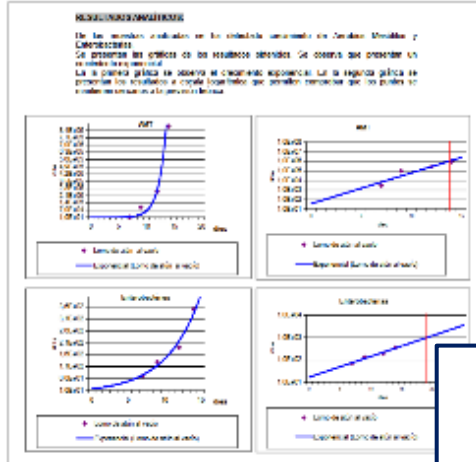
distribution to other member countries

Countries notified for the notification and associated follow-up

For follow-up:

Countries notified for the notification and associated follow-up

Notification reference: (348541) 01/09/2020 09:16:20 Page 1 of 13
 Document reference: 348541_1598944581062



REPUBLICQUE FRANÇAISE
 MINISTÈRE DE L'ÉCONOMIE ET DES FINANCES

RAPPORT D'ESSAIS
 MAR-2017... du 01/06/2017

DÉNOMINATION DE L'ÉCHANTILLON
 LONGE DE THON (ADDITIF)

DESCRIPTION DE L'ÉCHANTILLON
 Échantillon reçu dans une barquette mise sous blister, le tout dans un carton.

RÉSULTATS DES EXAMENS ET ESSAIS EFFECTUÉS SUR L'ÉCHANTILLON

1. Nitrites E251-E252 (en NaNO₂) : 100 mg/kg

2. Nitrate E251-E252 (en NaNO₃) : 20 (LO) mg/kg

Emil Vollrath
 Erste Hannoversche Wurst-,
 Aufschnittwaren- und
 Fleischkonserven-Fabrik
 Hannover.

Emil Vollrath, Hannover, den 17. 7. 1917

Nachnahme: Mk. Pt. in Wörten

Herrn
 Ffido
 W. Bahmann
 Hofmann
 H. Meyer W. an d. Klemmbuße
 RECHNUNG.

Größtes und leistungsfähigstes
 Fabrik-Etablissement
 circa 8000 qm Nutzfläche.

Letzte eigene Jahresschlachtung
 und Verwertung:
 erstkl. Schweine.

Ann-Adresse:
 th, Hannover
 Nord 429
 721

Scheckkonto:
 20, Hannover
 Meyer & Sohn,
 annover.

Quantität	Einheit	Preis	Werte
2	Pf.	12,00	24,00
1	Pf.	3,75	3,75
1	Pf.	2,25	2,25

- ## Attachments:
- Invoices, Delivery lists, Receipts
 - Product labels, Ingredients lists, Recipes
 - Lab reports, measurement slips
 - Official reports and letters from agencies

Posto	Posizione	Posto	Posizione
15	15	15	15
16	16	16	16
17	17	17	17
18	18	18	18
19	19	19	19
20	20	20	20
21	21	21	21
22	22	22	22
23	23	23	23
24	24	24	24
25	25	25	25
26	26	26	26
27	27	27	27
28	28	28	28
29	29	29	29
30	30	30	30
31	31	31	31
32	32	32	32
33	33	33	33
34	34	34	34
35	35	35	35
36	36	36	36
37	37	37	37
38	38	38	38
39	39	39	39
40	40	40	40
41	41	41	41
42	42	42	42
43	43	43	43
44	44	44	44
45	45	45	45
46	46	46	46
47	47	47	47
48	48	48	48
49	49	49	49
50	50	50	50



How to extract data from RASFF?



Operator	
Operator type:	farmer
Name:	Egg producer 1
ApprovalNumber:	37.0123456
Address:	Tamara-Danz-Straße 11
Location:	Berlin
Postal code:	10243
Country:	Germany
Distribution to:	Germany
Attachment(s):	
Operator type: storage	
Name:	Dummy Egg Storage 1
ApprovalNumber:	46.12343567
Address:	Mühlenstraße 100
Location:	Berlin
Postal code:	10243
Country:	Germany
Distribution to:	Germany
Attachment(s):	
Operator type: transporter	
Name:	Dummy Egg Transporter 1
ApprovalNumber:	67.3456578
Address:	Kottbusser Tor
Location:	Berlin
Postal code:	10999
Country:	Germany
Distribution to:	Germany
Attachment(s):	

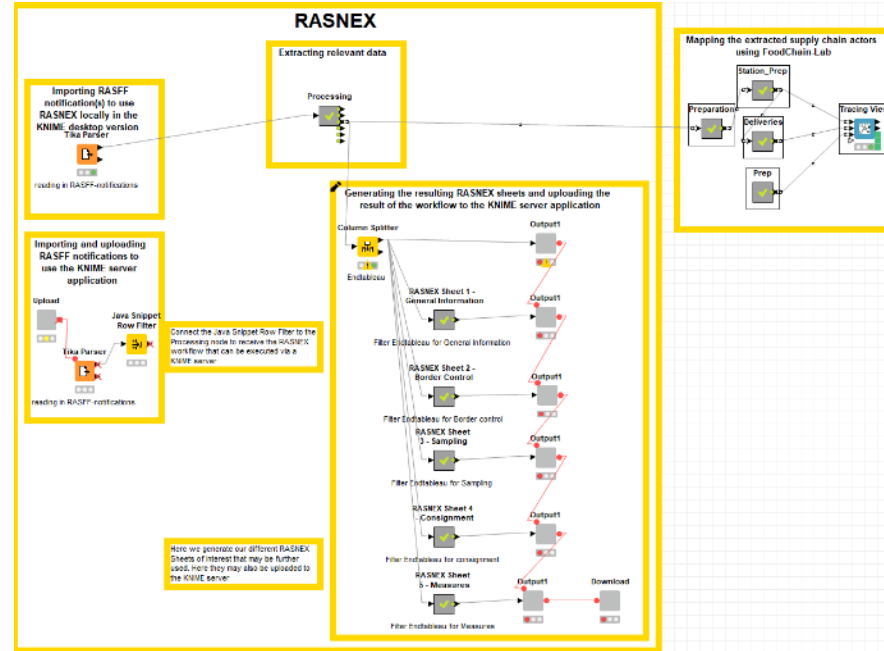
Challenges:

- Manual work that is done in multiple institutions and MS in parallel
- Non-standardised & unstructured formatting of attached documents
- Multiple languages
- Multiple address formats amongst MS
- Continious flow of information (updates on the RASFF)

→ **How can we support people working with RASFF?**

Where do we come from? RASNEX 1.0

RASFF TEST ENVIRONMENT	
Operator type:	farmer
Name:	Egg producer 1
ApprovalNumber:	37.0123456
Address:	Tamara-Danz-Straße 11
Location:	Berlin
Postal code:	10243
Country:	Germany
Distribution to:	Germany
Attachment(s):	
Operator type:	storage
Name:	Dummy Egg Storage 1
ApprovalNumber:	46.12343567
Address:	Mühlenstraße 100
Location:	Berlin
Postal code:	10243
Country:	Germany
Distribution to:	Germany
Attachment(s):	
Operator type:	transporter
Name:	Dummy Egg Transporter 1
ApprovalNumber:	67.3456578
Address:	Kottbusser Tor
Location:	Berlin
Postal code:	10999
Country:	Germany
Distribution to:	Germany
Attachment(s):	



Row ID	EU legs	Specify	Maximu	Sample	Distribu	Consp	Origin	Public	Public health certificate date	CVED n	Other d	Durabil	Descr	Descr	PL/HE	Operat	Approv	Address	Location	Postal	Counr	Distribu
Traceability (...)					distribution t...	Printnumber...	Germany		Origin: Germany 1-DE-1234567 und 2-DE-3456789			5000	350 kg	0_0	storage	46.12343567	Mühlenstraße 100	Berlin	10243	Germany	Germany	
Traceability (...)					distribution t...	Printnumber...	Germany		Origin: Germany 1-DE-1234567 und 2-DE-3456789			5000	350 kg	0_0	transporter	67.3456578	Kottbusser Tor	Berlin	10999	Germany	Germany	
Traceability (...)					distribution t...	Printnumber...	Germany		Origin: Germany 1-DE-1234567 und 2-DE-3456789			5000	350 kg	0_0	wholesaler	12.789123	Marienplatz	München	80331	Germany	Austria	
Traceability (...)					distribution t...	Printnumber...	Germany		Origin: Germany 1-DE-1234567 und 2-DE-3456789			5000	350 kg	0_0	retailer	45.123456	Pinz Eugen-Straße 27	Wien	1030	Austria	Austria	
Traceability (...)					distribution t...	Printnumber...	Germany		Origin: Germany 1-DE-7891234			1000	70 kg	1_0	farmer	78.456789	Schloßplatz	Stuttgart	70173	Germany	Germany	
Traceability (...)					distribution t...	Printnumber...	Germany		Origin: Germany 1-DE-7891234			1000	70 kg	1_0	wholesaler	89.123456	Südstraße 1	Heilbronn	74072	Germany	Germany	
Traceability (...)					distribution t...	Printnumber...	Germany		Origin: Germany 1-DE-7891234			1000	70 kg	1_0	exporter	56.123123	Bernhard-Grämek-Allee 1	Frankfurt am Main	60316	Germany	Spain	
Traceability (...)					distribution t...	Printnumber...	Germany		Origin: Germany 1-DE-7891234			1000	70 kg	1_0	importer	11.456456	Plaza Mayor	Madrid	28012	Spain	Spain	

KNIME-based solution

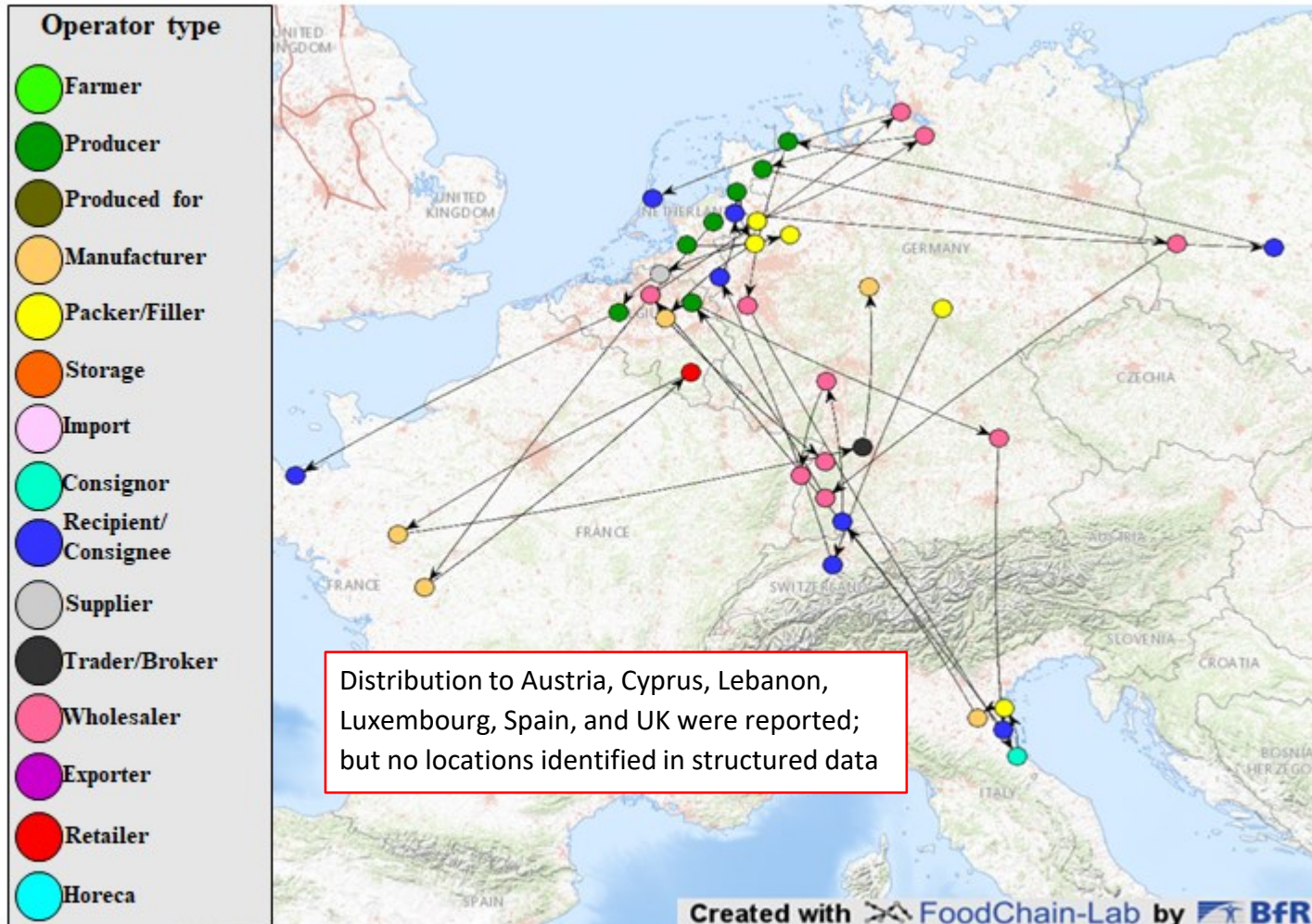
„Konstanz Information Miner“,
<https://www.knime.com/>

Low-code platform for data analytics and reporting

free and open-source, and locally hosted

Structured RASFF data to tabular format

Where do we come from? Fipronil contamination in the egg supply chain



Source: gstatic.com

Single notification containing **700** follow-ups notifications,

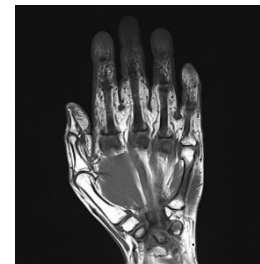
yet **only 40 locations** were identified in the structured data.

Significantly more data in the attachments.

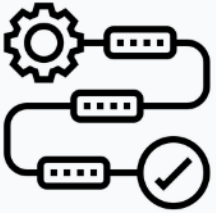
Need more advanced tools



If RASNEX 1.0 is like an X-Ray into the supply chain incident, then we need MRI



Upgrade to RASNEX 2.0 – EFSA-BfR FPA Project SA 07 & SA 10



- Text Mining through AI
- Optical Character Recognition (OCR)



1. Extract relevant information from fups & attached documents

- Extract recognized addresses
- Operators (names) of supply chains
- Intent of the fup
- Products
- Analytical results
- Date of production
- Billing dates
- Delivered quantities

2. Linking RASNEX with the FoodChain-Lab data extraction workflow (UTX format)

Icon made by Eucalyp from www.flaticon.com
Icon made by Freepik from www.flaticon.com

RASFF notifications – attached documents

The screenshot shows a RASFF notification interface. On the left, a sidebar titled 'Anlagen' (Attachments) lists four files: 'AnalyticalreportTA2016854-BT_0.pdf', 'CMR.PDF', 'Invoice.pdf', and 'Invoice130820.pdf'. A red callout box points to this list and contains the following text:

Various, non-standardised, un- or semi-structured formats

Files (PDFs, Word, Excel, etc.)

- With structured and non-structured parts
- E.g. invoices, delivery documents, laboratory analysis

→ Tools to parse them needed

Images of the above

→ Requiring OCR step

The background interface shows the RASFF logo and a notification for 'Trout fillet Sliced Cold Smoked'. Below the callout, the notification details are partially visible:

Notification Reference: [unreadable]

Notification: [unreadable]

Notification: [unreadable]

Notification: [unreadable]

Product name: Trout fillet Sliced Cold Smoked

Product category: fish and products thereof

RASFF notifications – data extraction methods



Additional information: Additional analytical reports attached. Additional analyzes were commissioned by Dunrin Polska Sp. z o.o. : - "Bio almond milk" purchased from Almondina GmbH Schillerstraße 9, Bad Dürkheim, 69345, Rheinland-Pfalz - test result correct, ethylene oxide content # 0.01 mg / kg. - two batches of baking mixes called Whole & Tasty Bio produced in January 2021 by Dunrin Polska Sp. z o.o. - test result correct, ethylene oxide content # 0.01 mg / kg

Follow-up type: additional information
#357488 - rg validated - N_AAC_RASFF - AAC/RASFF

CP Reference: 1-2
Organisation / ministry: German Ministry of Test Notification
Contact person: TOM TRACING

Additional information: In the outcome of the investigations at the company FURMA - Via Fogorina 19 - BARLETTA, it was found that the same had received a package of the product subject to the alert. The product has not been marketed, and is put aside, pending withdrawal by the manufacturer

Follow-up type: additional information
#357489 - rg validated - N_AAC_RASFF - AAC/RASFF

CP Reference: 1-3
Organisation / ministry: German Ministry of Test Notification
Contact person: OLGA OUTBREAK

Additional information: The company WOB from Maastricht (NE) is included in the distribution list of the company Magic Herbs. In reality the product was delivered to: WOB Sudo srl - Van Brakelstraat 58 - 3514 HH Utrecht

Follow-up type: additional information
#357491 - rg validated - N_AAC_RASFF - AAC/RASFF

CP Reference: 1-5
Organisation / ministry: German Ministry of Test Notification
Contact person: LAURA LABORATORY

Additional information: With reference to -fup32, it is reported that the withdrawal was not known at the company A & O, Bundesstr. 19, 74465 Dresden, Germany. The goods found on site were voluntarily taken out of circulation.

Follow-up type: additional information
#357492 - rg validated - N_AAC_RASFF - AAC/RASFF

CP Reference: 1-4
Organisation / ministry: German Ministry of Test Notification
Contact person: ANTONIA ANALYSIS

Additional information: With reference to fup11: During the investigation of one of the recipients it was found that Freshmarket Klaus Hohn, Weißenseeufer 34, 55930 Sonnersdorf does not have the questioned product in stock. to the statement of the operator, Mr. Klaus Hohn, the product was sold to individual customers on a receipt. Mr. Klaus Hohn also stated that there was no product in question on the market in his stationary store. Moreover, as established, the trader was not informed by Ganz Biohandels GmbH about the non-compliance.

Challenges:

- Legal:
 - Sensitive data
→ Only local, open source models (no ChatGPT)
- Technical:
 - Various formats and data sources
 - Processing capabilities

Frontend methods:

- Dashboard: streamlit

Backend methods (swappable components):

- LLM Engine: llama3.2 with RAG: LangChain
- Document-parsing: Docling with OCR (Easy OCR, Tesseract)

RASNEX App: Dashboard

Upload a document ^

Enter the folder path of the notification

You selected: /home/savvateeva/Downloads/ocr 5

Choose an option:

You selected: /home/savvateeva/Downloads/ocr 5/fup30 Oficio815seguimientoreporte512RASFF.pdf

Open summary table v



Delete selected rows

save changes in the table

Bundesinstitut für Risikobewertung (BfR) | Max-Dohrn-Straße 8-10 | 10589 Berlin

vertraulich
(Nur) per E-Mail an:
cc:

Julius van Flocksbum
Leiter der Abteilung für Lebensmittelhygiene
Hauptbehörde für Lebensmittelsicherheit (HBL)
Rue de Livourne 54a, 4983
Brüssel, Belgien

Ihre Zeichen und Nachrichten vom	Geschäftszweck: Bitte bei Antwort angeben	Telefondurchwahl	Datum	Organisationsinheit/ Ansprechpartner
5635	56.00	12435	23.01.2025	Herr Müller

Lieber Herr van Flocksbum,

in Bezug auf die Berichte Nr. 2024/8943fup23, die vom Schnellwarnsystem für Lebens- und Futtermittel (RASFF) der Hauptbehörde für Lebensmittelsicherheit (HBL) herausgegeben wurden und die sich aus der Untersuchung eines Ausbruchs einer Lebensmittelvergiftung durch Methylquecksilber ergeben, das in Spanien von der COLFI SA aus gefrorenem Thunfisch (*Thunnus atlanticus*) des Betriebs Thuna Congelados, S.A. de C.V. (Zulassungsnummer 93875) verarbeitet wurde.

Nach der von der Gesundheitsbehörde durchgeführten Untersuchung teile ich Ihnen mit, dass Thuna Congelados für die Lose T736, T773 und T938 Unterlagen über die Temperaturkontrolle und Laborergebnisse vorgelegt hat.

Mit freundlichen Grüßen

im Auftrag

Unterschrift

Anlage(n)

Im Geschäftsbereich
Bundesinstitut für Risikobewertung (BfR)
Max-Dohrn-Straße 8-10
10589 Berlin
T +49 30 18412-0
bfr@bfr.bund.de
bfr.bund.de

Standorte
Berlin-Jungfernheide
Max-Dohrn-Straße 8-10
10589 Berlin
T +49 30 18412-0
F +49 30 18412-99099
Berlin-Marienfelde
Diedersdorfer Weg 1
12277 Berlin
T +49 30 18412-0
F +49 30 18412-99099
Berlin-Alt-Marienfelde
Alt-Marienfelde 17/21
12277 Berlin
T +49 30 18412-0
F +49 30 18412-99099
Berlin-Nahmitzer-Damm
Nahmitzer Damm 12
12277 Berlin
T +49 30 18412-0
F +49 30 18412-99099

Datenschutzhinweise
einschließlich Informa-
tionen zu Ihren Rechten
finden Sie unter:
bfr.bund.de/de/daten-
schutz/faq-datenschutz.html

**BfR | Risiken erkennen –
Gesundheit schützen**



	☰	☰ Dis	☰ Company	☰ Country	☰ Co	☰ City	☰ Postal code	☰ Street	☰ House Nr
	0	<input type="checkbox"/>	Thuna Congelados						
	1	<input type="checkbox"/>	COLFI SA	Spain					
	2	<input type="checkbox"/>	Bundesinstitut für Ri	Germany		Berlin	10589	Max-Dohrn-Straße	8-10
	3	<input checked="" type="checkbox"/>		Germany		Berlin-Marienfelde	12277	Diedersdorfer Weg	1
	4	<input type="checkbox"/>		Germany		Berlin-Jungfernheide	10589	Max-Dohrn-Straße	8-10
	5	<input type="checkbox"/>		Germany		Berlin-Nahmitzer-Damm	12277	Nahmitzer Damm	12



RASNEX Workflow

Input documents

The input documents section shows a RASFF notification (RASFF 2023-01-14) regarding a food safety issue with organic sesame seeds. It includes details about the product, the manufacturer (Dipasa Europe B.V.), and the investigation process. Below the notification is a French invoice (RECHNUNG) for the same product, showing quantities and prices.

RASNEX Dashboard

The RASNEX Dashboard interface shows a document upload section where a file named 'RASFF-test_14.pdf' has been selected. Below this is the 'Document Viewer' showing the RASFF notification content. To the right is the 'Extracted UTX data' table, which lists the extracted data points from the document.

Company	Country	City	Postal
<input type="checkbox"/>	Wittenberger Bäckers		
<input type="checkbox"/>	Dipasa Europe B.V.		
<input type="checkbox"/>	MILLBAKER SAS		
<input type="checkbox"/>	MONDELEZ FRANCE		
<input type="checkbox"/>	Infinity Foods		
<input type="checkbox"/>	Fuchs		
<input type="checkbox"/>	0 (Unknown country)	Lutherstad	06886
<input type="checkbox"/>	0 (unknown country)	Enschede	7547 TD
<input type="checkbox"/>	Germany (Note: Germu Berlin)		10623
<input type="checkbox"/>	Ireland	0 (not a city)	0 (not apt)
<input type="checkbox"/>	Gibraltar	Gibraltar	0 Gibraltar

Output and Analysis

The output and analysis section features a map of Europe with colored dots representing different operator types. A legend identifies the operator types: Farmer (green), Producer (orange), Produced for (yellow), Manufacturer (red), Packer/Filler (purple), Storage (blue), Import (pink), Unlabeled (grey), Recipient (light blue), Supplier (dark blue), Trader/Broker (brown), Wholesaler (light green), Exporter (dark red), and Blocca (light purple). Below the map is a table listing the extracted data points.

Company	Country	City	Postal code	Street	House Nr	Operator type	Address
Birngemaster GmbH	Germany	Berlin		Birkenstraße	22		Birkenstraße 22 14
gutsMuths+paris							null
Bio-Hochfläche							null
GUTSGUTTING							
EDERA Bio							
H&L							
PayPal	N/A	N/A		Gebrens	4		An den Gehren 4
	(unknown)	(unknown)		Unter den Eichen			Unter den Eichen
DE	Berlin		0	0			DE-12205 Berlin

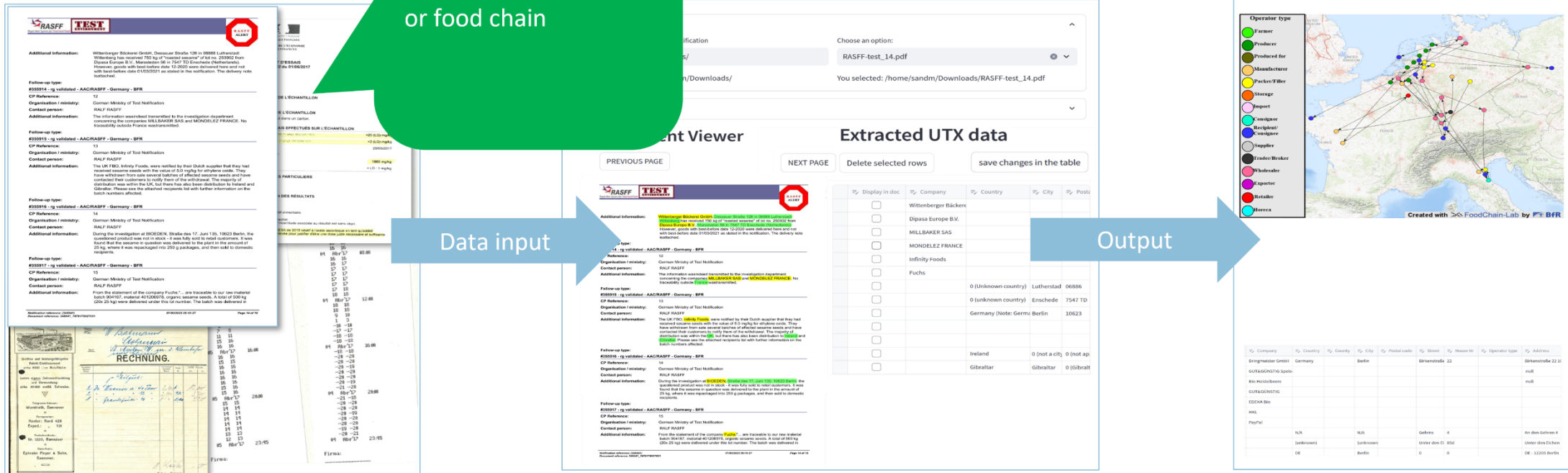
RASNEX Workflow

Input documents

RASFF Notifications contain multiple pages of structured and unstructured text, and can be updated during an incident in the feed or food chain

RASNEX Dashboard

Output and Analysis



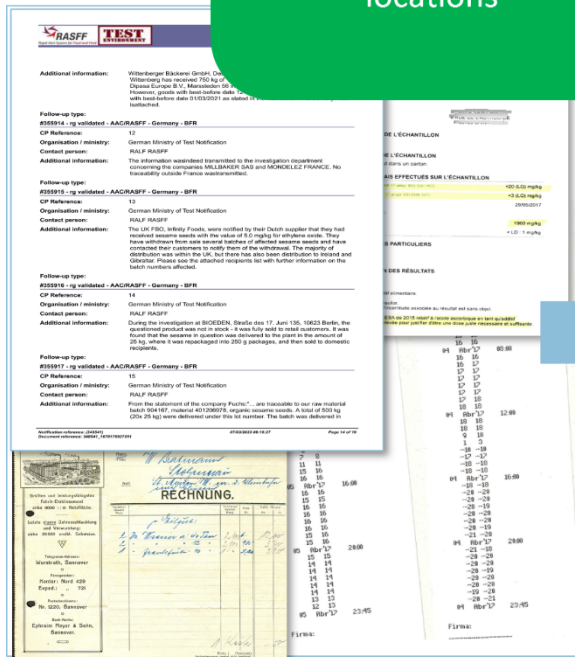
RASNEX Workflow

Input

RASNEX 2.0 uses language models to search for and extract relevant information such as operators and locations

RASNEX Dashboard

Output and Analysis



Data input

The dashboard interface includes an 'Upload a document' section with a file path input and a 'Document Viewer' showing a scanned RASFF notification. Below the viewer is an 'Extracted UTX data' table with columns for display, company, country, and city. A 'Data input' arrow points from the input documents to this dashboard.

Display in doc	Company	Country	City	Post
<input type="checkbox"/>	Wittberger Bäckerei			
<input type="checkbox"/>	Dipsas Europe B.V.			
<input type="checkbox"/>	MILLBAKER SAS			
<input type="checkbox"/>	MONDELEZ FRANCE			
<input type="checkbox"/>	Infinity Foods			
<input type="checkbox"/>	Fuchs			
<input type="checkbox"/>	0 (Unknown country)	Luterdad		06886
<input type="checkbox"/>	0 (Unknown country)	Enschede		7547 TD
<input type="checkbox"/>	Germany (Note: Germ. Berlin)			10623
<input type="checkbox"/>	Ireland	0 (not a city)	0 (not a city)	
<input type="checkbox"/>	Gibraltar	Gibraltar	0 (Gibraltar)	

Output

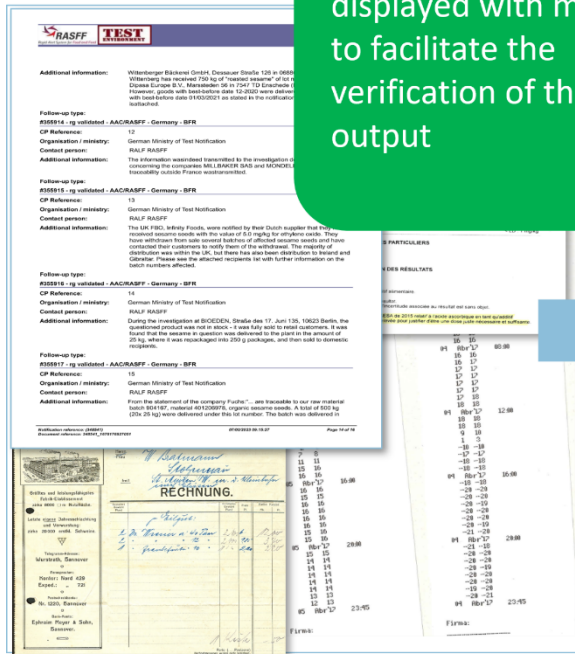
The output section features a map of Europe with colored dots representing operator locations. A legend titled 'Operator type' lists categories like Partner, Producer, and Manufacturer. Below the map is a table with columns for company, country, city, and address, listing various operators such as 'Bisminger GmbH' and 'GUTKUNSTIG'.

Company	Country	City	Postal code	Street	House No.	Operator type	Address
Bisminger GmbH	Germany	Berlin		Berlinerstraße 22			Berlinerstraße 22 10
GUTKUNSTIG							mail
Bisminger GmbH							mail
GUTKUNSTIG							mail
EDUCA Bio							
HML							
FayPi	N/A	N/A		Definis	4		An den Gärten 4
0 (Unknown)	(Unknown)	(Unknown)		Unter den Eichen			Unter den Eichen
VE	Berlin				6		DE 10205 Berlin

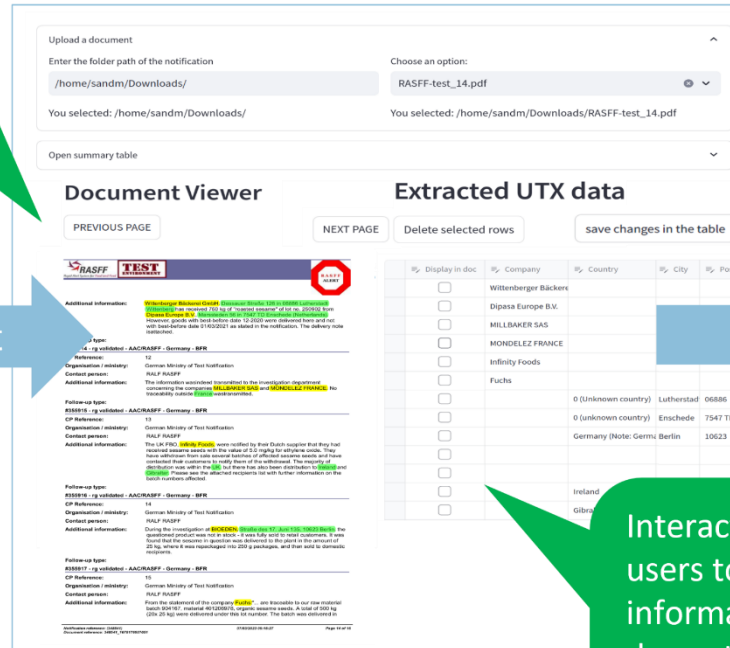
RASNEX Workflow

Input documents

Processed document is displayed with mark-up to facilitate the verification of the model output



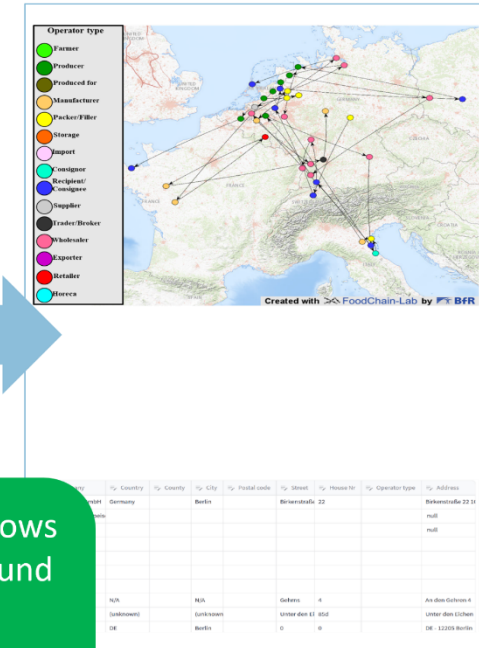
RASNEX Dashboard



Data input

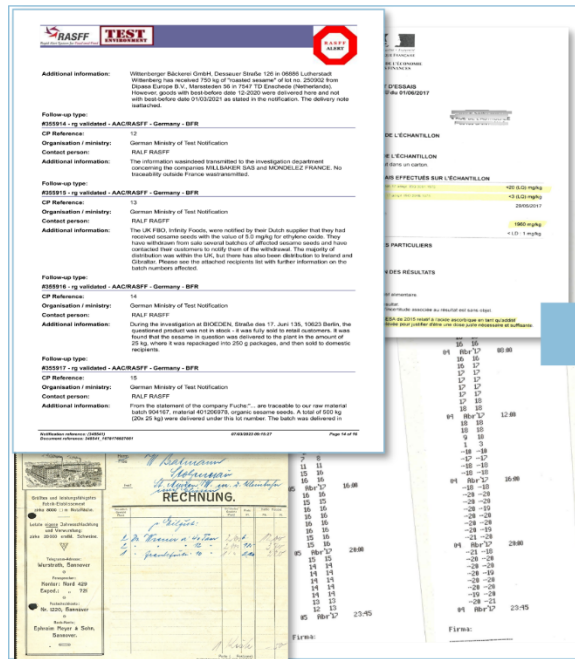
Output

Output and Analysis

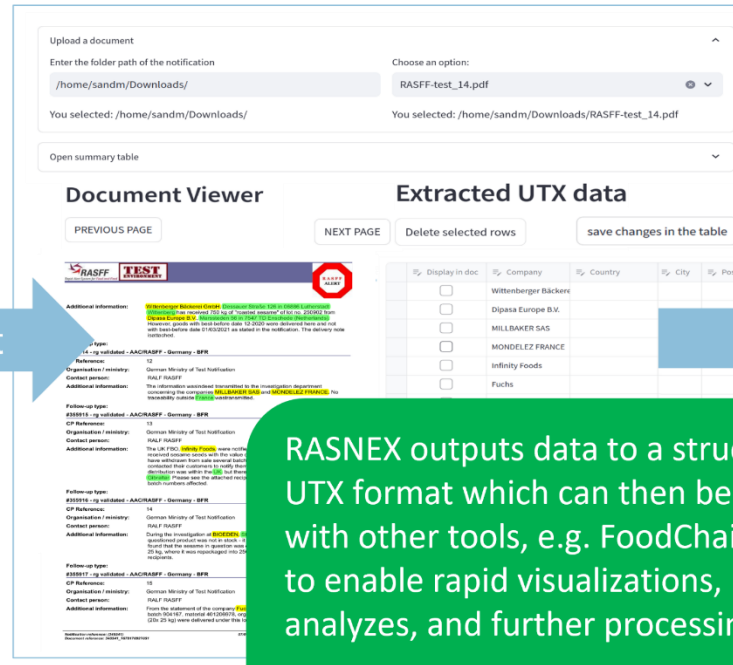


RASNEX Workflow

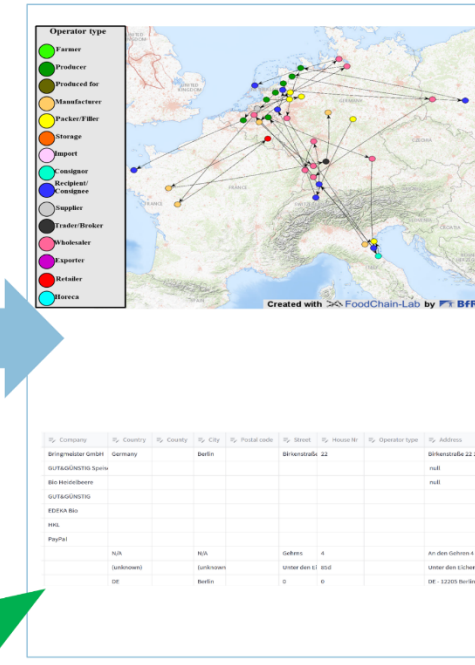
Input documents



RASNEX Dashboard



Output and Analysis



Data input

Output

RASNEX outputs data to a structured UTX format which can then be used with other tools, e.g. FoodChain-Lab to enable rapid visualizations, analyzes, and further processing

Live-Demo



Rapid advances in technology during RASNEX development

RASNEX 1.0: deterministic extraction of structured text

RASNEX 2.0

When	What	How
2021	only FUP report text	NER-based
2022	only FUP report text	NER-based with pattern-matching, supervised fine-tuning
2023	FUP report text machine readable attachments (.doc, .xls, some PDFs)	LLM-based, prompting
2024	FUP report text most attachments (.doc, .xls, some PDFs, Images)	LLM-based, prompting plus RAG, document parsers based on Computer Vision (CV) and Optical Character Recognition (OCR)

Final remark: If you are interested in collaborating and cooperating, please do not hesitate to contact us.

Thank you for your attention!

RASNEX Team: Daria Savvateeva, Matthew Salewski & Marc C. Lorenzen

German Federal Institute for Risk Assessment (BfR)

Max-Dohrn-Str. 8-10 □ 10589 Berlin

Tel. 030 - 184 12 - 28314 □ Fax 030 - 184 12 - 47 41

marc.lorenzen@bfr.bund.de □ www.bfr.bund.de

Sources

Source Slide 4: <https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcRLkLS-C3wWyFiFxTi9hBpGBb-pZAUdxAKXXQ&usqp=CAU> (accessed 21.01.2025)