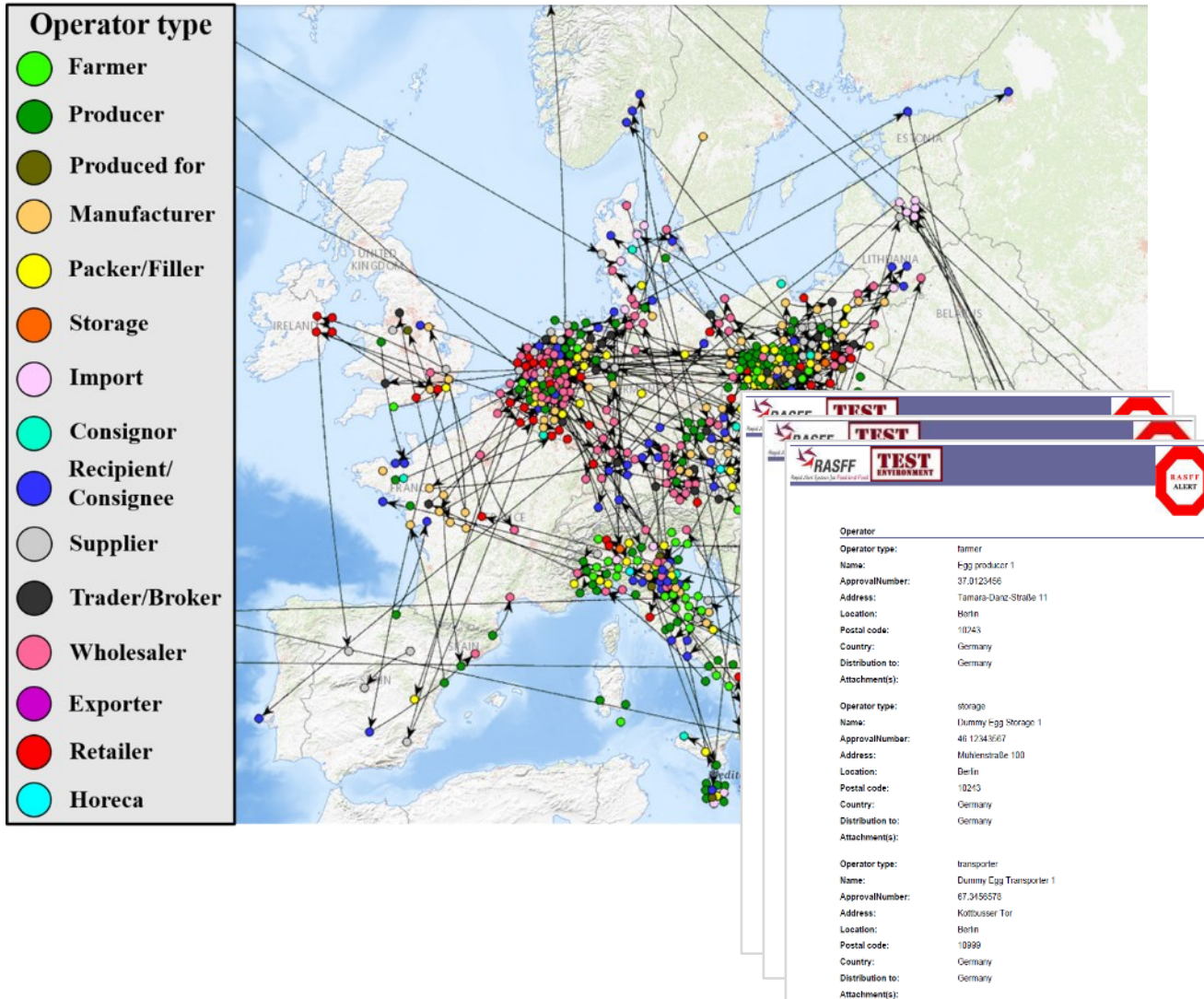


# **AI-supported data extraction: The Rapid Alert Supply Network EXtractor (RASNEX) tool for mining supply chain information**

11.02.2025

Daria Savvateeva, Matthew Salewski & Marc C. Lorenzen

# Contamination in the Supply Chain



**Need to stop outbreaks as fast as possible!**

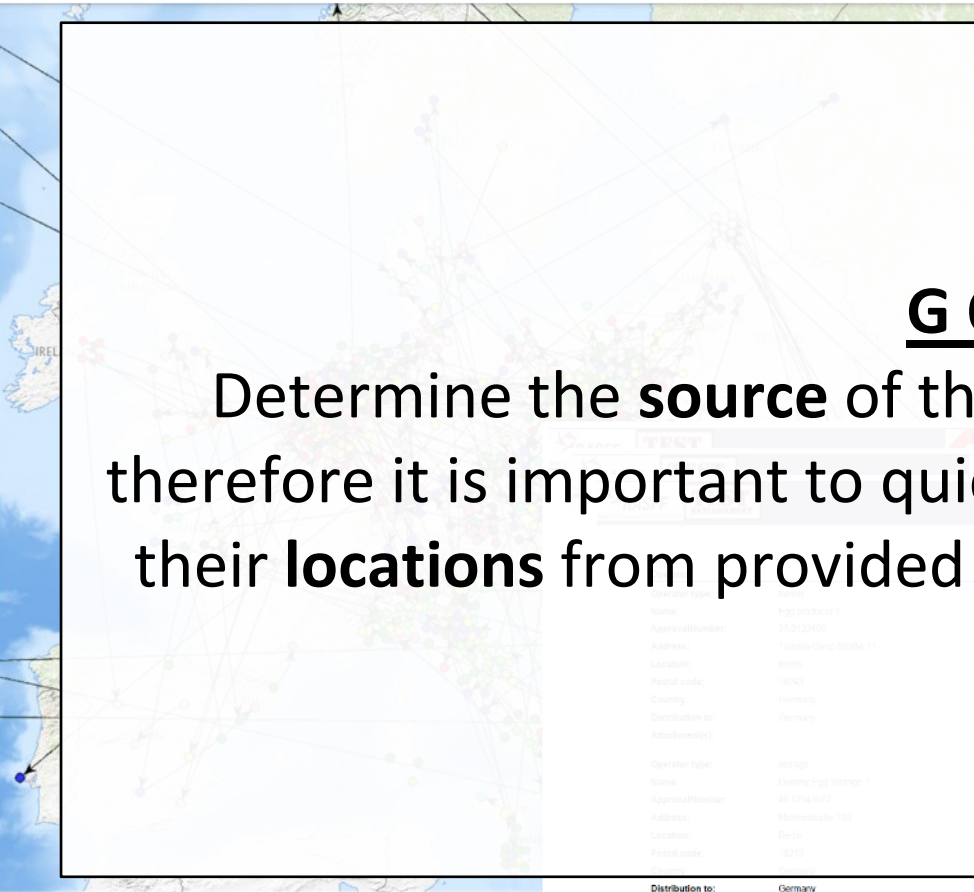
**Rapid Alert System for Food and Feed (RASFF) provides relevant information**

**Collected information (manual work)**

- laboratory and delivery data
- times, 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: Dummy Egg Transporter 1  
 ApprovalNumber: 67.3456578  
 Address: Kottbuser Tor  
 Location: Berlin  
 Postal code: 10999  
 Country: Germany  
 Distribution to: Germany  
 Attachments:



# 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**

<b>Notification number:</b>	348541
<b>Reference:</b>	
<b>Notification type:</b>	food
<b>Notification basis:</b>	company's own check
<b>Notification classification:</b>	alert notification
<b>Product name:</b>	eggs
<b>Product category:</b>	eggs and egg products
<b>Notifying country:</b>	Germany
<b>Notifying Country region:</b>	BfR
<b>CP Reference:</b>	
<b>Date of notification:</b>	15/07/2020
<b>Subject:</b>	fipronil dummy contamination
<b>Infosan Informed:</b>	No
<b>eCommerce Related:</b>	No
<b>Mutual recognition:</b>	No
<b>Notification status:</b>	DRAFT
	15/07/2020 10:31:37
<b>Follow-up status:</b>	
<b>Number of follow-ups:</b>	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)  
Document reference: 348541\_1598944581062

01/09/2020 09:16:20

Page 1 of 13

**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**

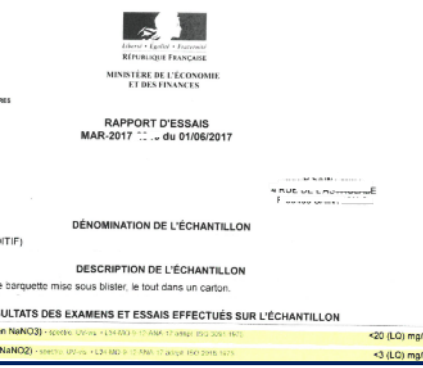
**CP Reference:** 1-4  
**Organisation / ministry:** German Ministry of Test Notification  
**Contact person:** ANTONIA ANALYSIS

**Additional information:** With reference to fup11: During the investigation of one of the recipients it was found that Freshmarket Klaus Hohn, Weißenseufer 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.

Notification reference: (348541)  
Document reference: 348541\_1693910790350

05/09/2023 12:46:30

Page 14 of 16

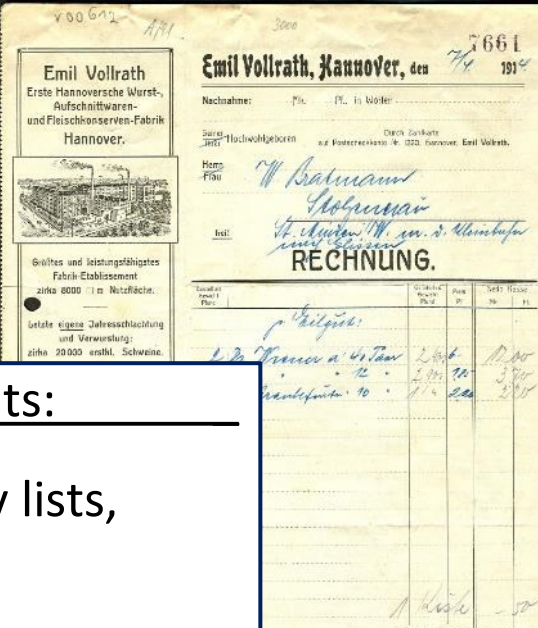


RAPPORT D'ESSAIS  
MAR-2017 ... du 01/06/2017

DÉNOMINATION DE L'ÉCHANTILLON  
(TIF)  
DESCRIPTION DE L'ÉCHANTILLON  
barquette mise sous blister, le tout dans un carton.

ULTATS DES EXAMENS ET ESSAIS EFFECTUÉS SUR L'ÉCHANTILLON

n NANO3	<20 (LC) mg/kg
n NANO2	<3 (LC) mg/kg



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

RECHNUNG.

**Attachments:**

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

Notification reference: (348541)  
Document reference: 348541\_1598944581062

01/09/2020 09:16:20

Page 1 of 13



# 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):	

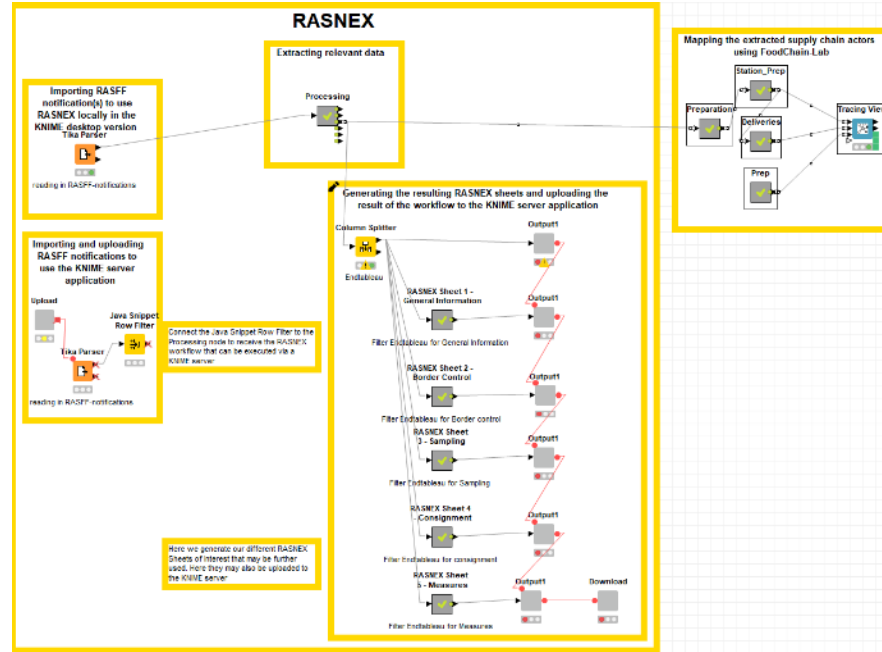
## Challenges:

- 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
- Continuous 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		RASFF ALERT
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):		



Row ID	BU logs	Specif	Maxim	Sample	Distrib	Corrig	Origin	Public	Public health certificate date	CHED n...	Other d...	Durabil...	Descr...	Descr...	PL/HL	Operat...	Approv...	Address	Location	Postal ...	Count...	Distrib...	
Sampling (Ma...)	37/2010		Absence	0_0																			
Sampling (Ma...)	37/2010		Absence mg...	0_1																			
Traceability (...)					distribution t...		Prinumber...	Germany		1:OE-1234567 and 2:OE-3456789			5000	350 kg	0_0		farmer	37.0123456	Tamara-Danz-Straße 11	Berlin	10243	Germany	Germany
Traceability (...)					distribution t...		Prinumber...	Germany		1:OE-1234567 and 2:OE-3456789			5000	350 kg	0_0	storage	46.12343567	Mühlenstraße 100	Berlin	10243	Germany	Germany	
Traceability (...)					distribution t...		Prinumber...	Germany		1:OE-1234567 and 2:OE-3456789			5000	350 kg	0_0	transporter	67.3456578	Kottbusser Tor	Berlin	10999	Germany	Germany	
Traceability (...)					distribution t...		Prinumber...	Germany		1:OE-1234567 and 2:OE-3456789			5000	350 kg	0_0	wholesaler	12.789123	Marerplatz	München	80331	Germany	Austria	
Traceability (...)					distribution t...		Prinumber...	Germany		1:OE-1234567 and 2:OE-3456789			5000	350 kg	0_0	retailer	45.123456	Prinz Eugen-Strasse 27	Wien	1030	Austria	Austria	
Traceability (...)					distribution t...		Prinumber...	Germany		1:OE-7891234			1000	70 kg	1_0	farmer	78.456789	Schloßplatz	Stuttgart	70173	Germany	Germany	
Traceability (...)					distribution t...		Prinumber...	Germany		1:OE-7891234			1000	70 kg	1_0	wholesaler	89.123456	Südstraße 1	Heilbronn	74072	Germany	Germany	
Traceability (...)					distribution t...		Prinumber...	Germany		1:OE-7891234			1000	70 kg	1_0	exporter	96.123123	Bernhard-Grüneke-Allee 1	Frankfurt am Main	60316	Germany	Spain	
Traceability (...)					distribution t...		Prinumber...	Germany		1:OE-7891234			1000	70 kg	1_0	importer	11.456456	Plaza Mayor	Madrid	28012	Spain	Spain	

KNIME-based solution

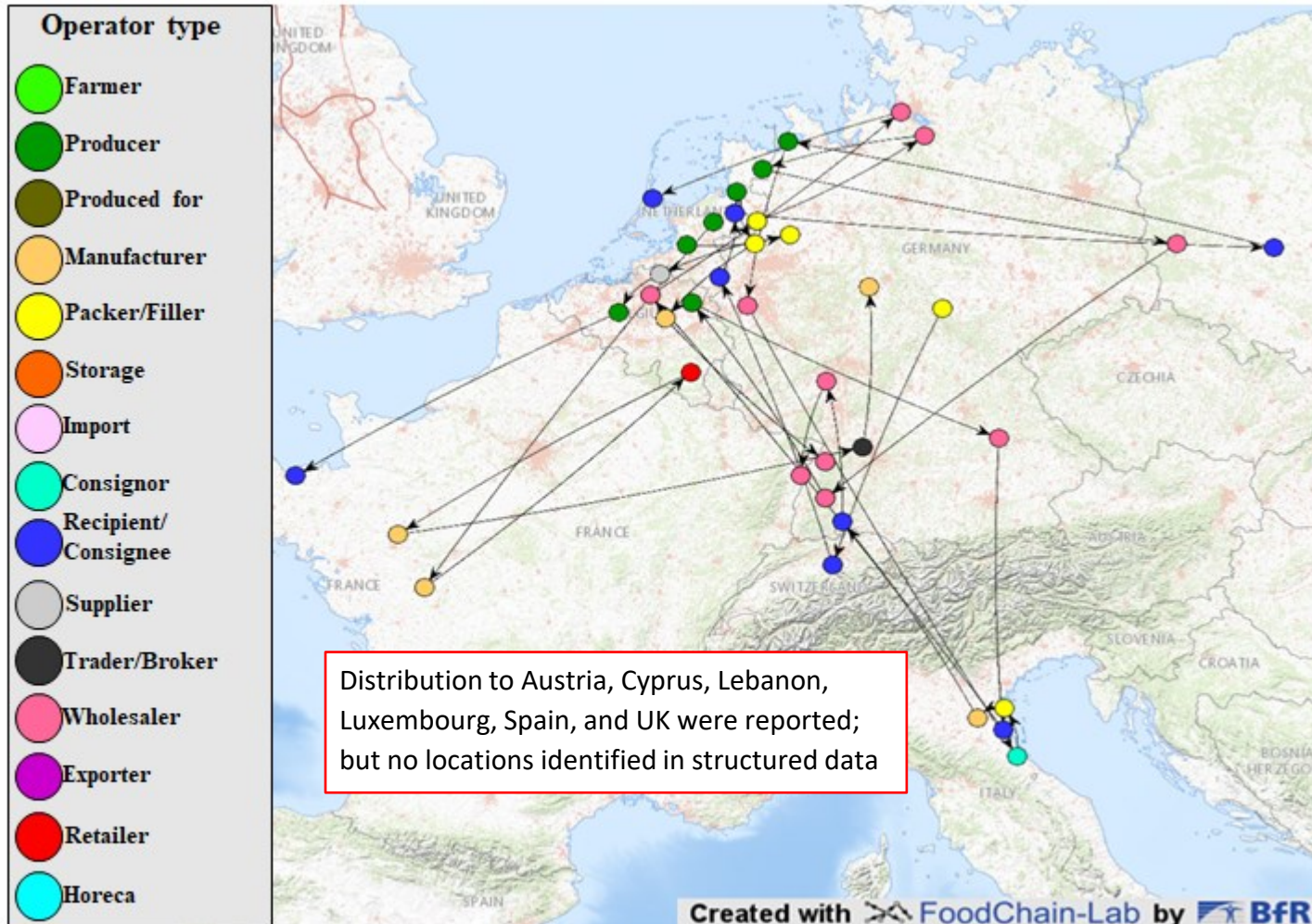
„Konstanz Information Miner“, <https://www.knime.com/>

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.

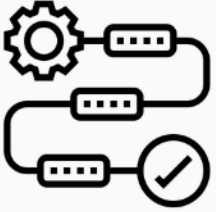
**Need more advanced tools**



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](http://www.flaticon.com)  
Icon made by Freepik from [www.flaticon.com](http://www.flaticon.com)



# RASFF notifications – attached documents

**Anlagen**

Name

- AnalyticalreportTA2016854-BT\_0.pdf
- CMR.PDF
- Invoice.pdf
- Invoice130820.pdf

**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

**RASFF**  
Rapid Alert System for

FF - Trout fillet Sliced Cold Smoked

**General Information**

**Notification Reference:**

**Notification Number:**

**Notification Date:**

**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) | Nordring-Straße 3-10 | 10559 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

Die Zeichen und Notizen von	Geplantes Datum (BfR-Notifizierung)	Terminanlass	Datum	Organisationsbereich/ Angelegenheit
SGT	15.02	12433	21.01.2025	Itzer 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 1736, 1773 und 1938 Unterlagen über die Temperaturkontrolle und Laborergebnisse vorgelegt hat.

Mit freundlichen Grüßen  
im Auftrag

Unterschrift  
Anlage(n)

In Geschäftsbereich  
Bundesinstitut für Risikobewertung (BfR)  
Nordring-Straße 3-10  
10559 Berlin  
T +49 30 18412 0  
bf@bf.bund.de  
bf.bund.de

**Standorte**  
**Berlin Jungfernheide**  
Max-Dohrn-Straße 8-10  
10588 Berlin  
T +49 30 18412 0  
F +49 30 18412-99999  
**Berlin Marienfelde**  
Diedersdorfer Weg 1  
13227 Berlin  
T +49 30 18412-0  
F +49 30 18412-99999  
**Berlin Diederichsen**  
Alte-Maienfelde 15/21  
13227 Berlin  
T +49 30 18412 0  
F +49 30 18412-99999  
**Berlin-Nahmitzer-Damm**  
Nahmitzer Damm 12  
13227 Berlin  
T +49 30 18412 0  
F +49 30 18412-99999

Deutschsprachige  
einschließlich  
Gemeinsam zu Ihren Rechten  
Beschwerden  
BfR kontaktieren  
schutz@bf.bund.de

BfR | Risiken erkennen –  
Gesundheit schützen



	Dis	Company	Country	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 two types of files: RASFF notifications and invoices. The RASFF notifications are from the German Ministry of Food and Consumer Protection, detailing food safety incidents. The invoice is a German 'RECHNUNG' (invoice) from 'Brotmannufaktur' to 'Brotmannufaktur'.

## RASNEX Dashboard

The RASNEX Dashboard interface includes an 'Upload a document' section with a file selection dropdown showing '/home/sandm/Downloads/RASFF-test\_14.pdf'. Below this is a 'Document Viewer' showing the content of the uploaded RASFF notification. To the right is an 'Extracted UTX data' table with columns for 'Display in doc', 'Company', 'Country', 'City', and 'Post'. The table lists various companies like Wittenberger Backerei, Expata Europe B.V., and others.

## Output and Analysis

The output analysis section features a map of Europe with colored data points representing different operator types. A legend on the left identifies the colors: Green for Producer, Yellow for Producer/Importer, Orange for Storage, Pink for Importer, Cyan for Importer/Exporter, Blue for Importer, Purple for Importer, Red for Importer, and Grey for Importer. Below the map is a table with columns for 'Company', 'Country', 'City', 'Postal code', 'Street', 'House no.', 'Operator type', and 'Address'. The table lists several entries, including 'Wittenberger Backerei' in Germany and 'Expata Europe B.V.' in the Netherlands.

Data input

Output



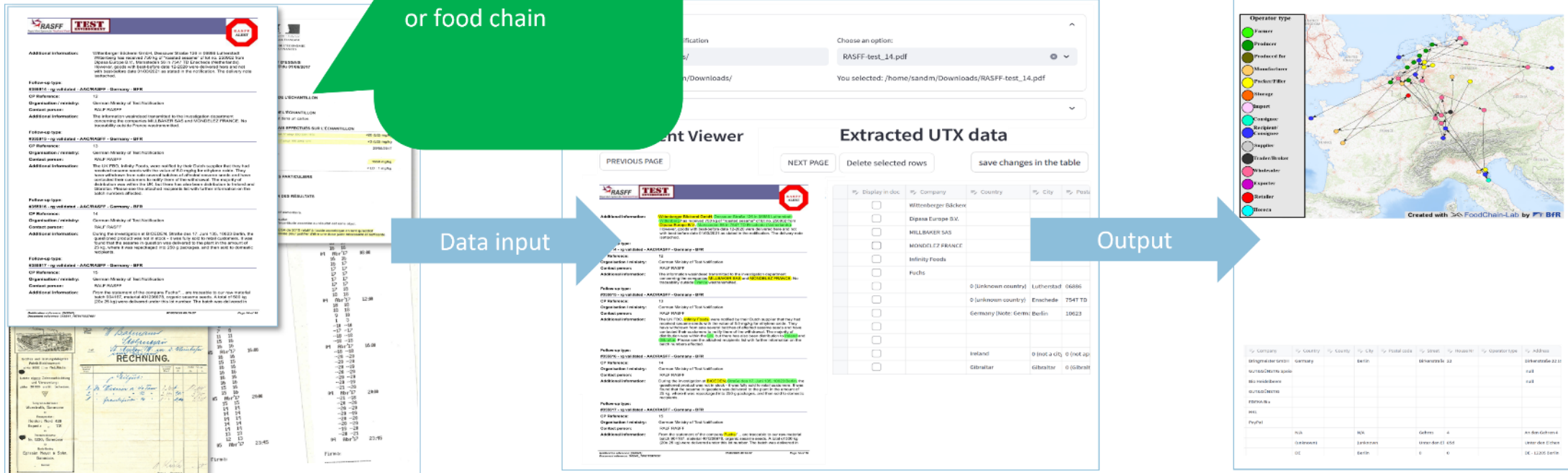
# RASNEX Workflow

## 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

The screenshot shows the RASNEX 2.0 interface. On the left, there's a 'Document Viewer' displaying a scanned document with highlighted text. On the right, there's an 'Extracted UTX data' table with columns for Company, Country, City, and Posts. A blue arrow labeled 'Data input' points from the document viewer to the dashboard.

The screenshot shows the RASNEX Dashboard. At the top, there's an 'Upload a document' section with a text input for the folder path and a dropdown for the document name. Below that, there's a 'Document Viewer' showing a scanned document with highlighted text. To the right, there's an 'Extracted UTX data' table with columns for Company, Country, City, and Posts. A blue arrow labeled 'Output' points from the dashboard to the output and analysis section.

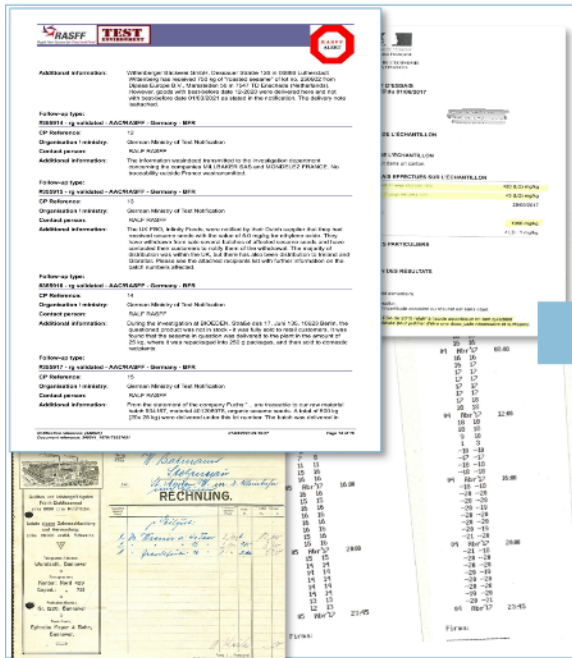
The screenshot shows the RASNEX Dashboard's output and analysis section. On the left, there's a legend for 'Operator type' with color-coded circles for various roles like Farmer, Producer, etc. On the right, there's a map of Europe with colored dots representing operator locations. Below the map, there's a table of extracted data with columns for Company, Country, City, and Posts.



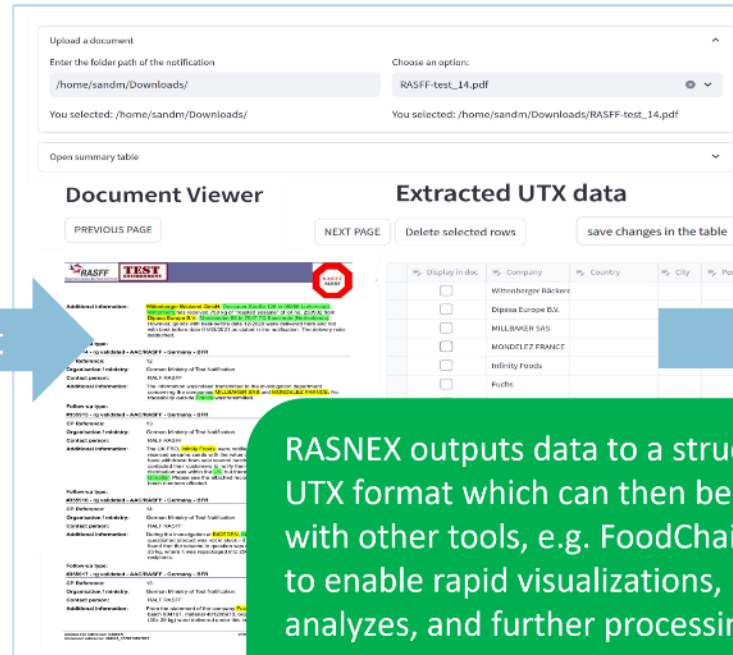


# RASNEX Workflow

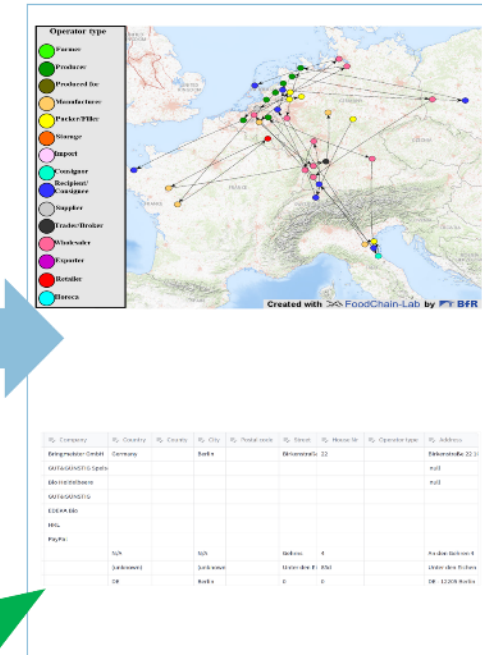
## Input documents



## RASNEX Dashboard



## Output and Analysis



# Live-Demo



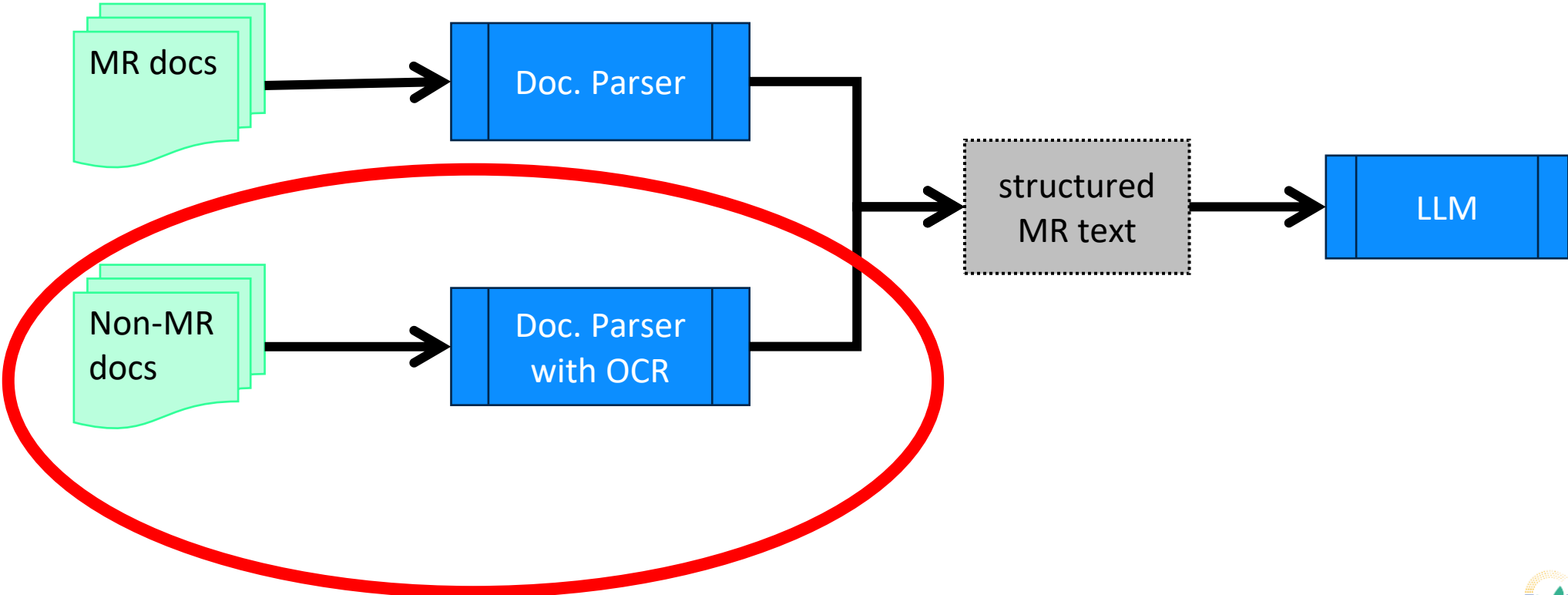
# Extraction from attachments - GPT4 (Large model) workflow: Accuracy and perspectives Live-Demo

Document parsing  
with Docling

Optical character  
recognition

Structured output of  
required info (with  
the large model)

# Extraction from attachments - GPT4 (Large model) workflow: Accuracy and perspectives Live-Demo





# Extraction from attachments: Example 1

**Institut Scientifique d'Hygiène et d'Analyse**  
S.A.S au capital de 4 100 000 euros - SIREN : 489 268 045 - APE : 7022 Z  
N° TVA intracommunautaire : FR50 489 268 045 - RCS : EVRY 489 268 045  
Rue du Moulin Blanc - CHAMONIX - 92275 LONGJUMEAU  
Tél : 01 69 79 31 50 - Fax : 01 64 48 82 49

**ISHA**  
Un laboratoire du Groupe Beta

N° dossier : CH.02010  
Code client : 040010000  
V/Ref. : COURRIER  
Date de commande : 10/04/2017  
Date réception : 13/04/2017

J'MARINE  
MONSIEUR DUFAR  
7 A 11 ALLEE DE FRIELEUSE  
BP 90231  
94539 RUNGIS CEDEX

Département : CHIMIE  
Champlan, le 21/04/2017

**RAPPORT D'ANALYSE**

Echantillon : [redacted]  
Libellé : 2.FILET DE THON DECONGELE SOUS VIDE  
DLC : 16/04/2017  
Date Fabr. :  
N° de lot / EMB : 11090

Fournisseur : PRIME  
Début des analyses : 18/04/2017  
Code article :

Analyse	Résultat	Méthode	Spécification
<b>Analyses : Chimie</b>			
Histamine	1419 mg / kg	HPLC	< 100 mg/kg

# Extraction from attachments: Example 1

**Institut Scientifique d'Hygiène et d'Analyse**  
S.A.S au capital de 4 100 000 euros - SIREN : 489 268 045 - APE : 7022 Z  
N° TVA intracommunautaire : FR50 489 268 045 - RCS : EVRY 489 268 045  
Rue du Moulin Blanc - CHAMONIXF - 92275 LONGJUMEAU  
Tél : 01 69 79 31 50 - Fax : 01 64 48 82 49

**ISHA**  
Un laboratoire du Groupe Beta

N° dossier : CH.02015  
Code client : 040910000  
V/Ref. : COURRIER  
Date de commande : 15/04/2017  
Date réception : 15/04/2017

J'MARINE  
MONSIEUR DJAFAR  
7 A 11 ALLEE DE FRIELEUSE  
BP 90231  
94539 RUNGIS CEDEX

Département : CHIMIE  
Champlan, le 21/04/2017

**RAPPORT D'ANALYSE**

Echantillon : [REDACTED]  
Libellé : 2.FILET DE THON DECONGELE SOUS VIDE  
DLC : 16/04/2017  
Date Fabr. :  
N° de lot / EMB : 11096

Fournisseur : PRIME  
Début des analyses : 18/04/2017  
Code article :

Analyse	Résultat	Méthode	Spécification
<b>Analyses : Chimie</b>			
Histamine	1419 mg / kg	HPLC	< 100 mg/kg

# Extraction from attachments: Example 1



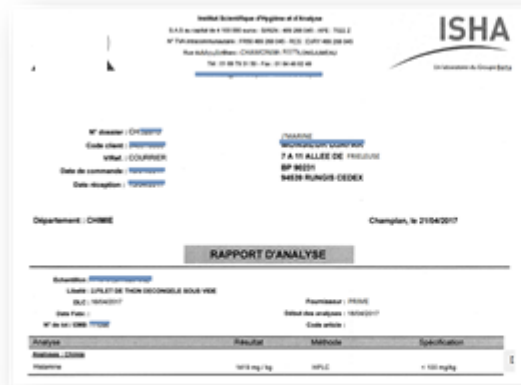
When the OCR/Doc-parsing works, Markdown preserves the original structure of the document, tables, captions, etc.

```
<!-- image -->N° EMA intracommunautaire : FR50 839 268 025 - ROS :
EURY 439 258 075 Rue du Moulin Blanc - CHAMONIX - F-92275
LONGJUMEAU Tél : 04 39 79 33 50 - Fax : 01 56 48 32 79
service@analyse.fr - www.analyse.fr<!-- image -->Institut Scientifique d
Hygiène et d Analyse insutut Scienüngue a Fiygtene et a Analyse S.A.S au
capital de 5 100 000 euros - SRIEN : 489 635 095 - APT : 7832 Zbet, 0469
/9 31 50 - Fax : 01 64 48 82 49 A Un laboratoire du Groupe BetaN°
dossier : CH.52815 J'MARINE Date de commande : 13/04/2017 BP
30231Code client : 560912300 MONSIEUR DJUFOR VIRef. : COURRIER TA
11 ALLEE DE FRIELEUSE Date réception : 13/08/2017Département :
CHIMIE Champlan, le 21/04/2017Echantillon : T 4010
{CH.56515.2}Echantillon : T 4010 {CH.56515.2)Fournisseur : PRIME Date
Fabr. : Debut des analyses : 19/04/2017Code article : | Analvses : |
| |-----|-----| | | nistamine 1419
mg/kg HPLC < 100 mg/kg |<!-- image -->
```



name	country	region	city	postal_code	street	house_nr
institut Scientifique d Hygiène et d Analyse	France	None	Longjumeau	92275	Rue du Moulin Blanc	None

# Extraction from attachments: Example 1

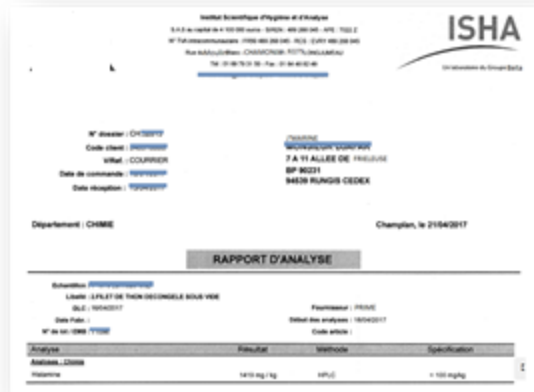


<!-- image -->N° EMA intracommunautaire : FR50 839 268 025 - ROS : EURY 439 258 075 Rue du Moulin Blanc - CHAMONIX - F-92275 LONGJUMEAU Tél : 04 39 79 33 50 - Fax : 01 56 48 32 79 service@analyse.fr - www.analyse.fr<!-- image -->institut Scientifique d Hygiène et d Analyse insutut Scienüngue a Fiygtene et a Analyse S.A.S au capital de 5 100 000 euros - SRIEN : 489 635 095 - APT : 7832 Zbet, 0469 /9 31 50 - Fax : 01 64 48 82 49 A Un laboratoire du Groupe BetaN° dossier : CH.52815 J'MARINE Date de commande : 13/04/2017 BP 30231 Code client : 560912300 MONSIEUR DJUFOR VIREf. : COURRIER TA 11 ALLEE DE FRIELEUSE Date réception : 13/08/2017 Département : CHIMIE Champlan, le 21/04/2017 Echantillon : T 4010 {CH.56515.2} Echantillon : T 4010 {CH.56515.2} Fournisseur : PRIME Date Fabr. : Debut des analyses : 19/04/2017 Code article : | Analyses : | |-----|-----| | | nistamine 1419 mg/kg HPLC < 100 mg/kg <!-- image -->



name	country	region	city	postal_code	street	house_nr
institut Scientifique d Hygiène et d Analyse	France	None	Longjumeau	92275	Rue du Moulin Blanc	None

# Extraction from attachments: Example 1



<!-- image -->N° EMA intracommunautaire : FR50 839 268 025 - ROS : EURY 439 258 075 Rue du Moulin Blanc - CHAMONIX - F-92275 LONGJUMEAU Tél : 04 39 79 33 50 - Fax : 01 56 48 32 79 service@analyse.fr - www.analyse.fr<!-- image -->Institut Scientifique d Hygiène et d Analyse insutut Scienüngue a Fiygtene et a Analyse S.A.S au capital de 5 100 000 euros - SRIEN : 489 635 095 - APT : 7832 Zbet, 0469 /9 31 50 - Fax : 01 64 48 82 49 A Un laboratoire du Groupe BetaN° dossier : CH.52815 J'MARINE Date de commande : 13/04/2017 BP 30231Code client : 560912300 MONSIEUR DJUFOR VIREf. : COURRIER TA 11 ALLEE DE FRIELEUSE Date réception : 13/08/2017Département : CHIMIE Champlan, le 21/04/2017Echantillon : T 4010 {CH.56515.2}Echantillon : T 4010 {CH.56515.2}Fournisseur : PRIME Date Fabr. : Debut des analyses : 19/04/2017Code article : | Analvses : | | |-----|-----| | | nistamine 1419 mg/kg HPLC < 100 mg/kg |<!-- image -->

SCORE:  
1 out of 2



name	country	region	city	postal_code	street	house_nr
institut Scientifique d Hygiène et d Analyse	France	None	Longjumeau	92275	Rue du Moulin Blanc	None



# Extraction from attachments: Example 2


**KAWAŁKI Z POŁĘDWICY Z TUŃCZYKA  
ŻÓŁTOPŁETWEGO**  
*Thunnus albacares, YFT*

Składniki: Tuńczyk Żółtopłety, antyoksydanty E-300, E-301, E331.  
Pochodzenie: Pacyfik FAO 77  
Metoda łowienia: Złwione  
Sposób łowienia: Sieci okrążające i podrywki, Okrężnice.  
Sposób pakowania: Próżniowo.  
Przechowywanie: W temp. Od 0 do +3°C  
Towar wcześniej zamrożony, nie zamrażać ponownie.


INFORMACJA O WARTOŚCI ODŻYWCZEJ	
Na 100 g produktu	
Energia:	336 kJ/81 Kcal
Tłuszcz:	0,85 g
w tym:	
kwasy tłuszczowe nasycone:	0,12 g
kwasy tłuszczowe jednonienasycone:	0,22 g
kwasy tłuszczowe wielonienasycone:	0,51 g
Omega 3:	0,02 g
Węglowodany:	0,12 g
w tym:	
Cukry:	<0,1g
Białko:	20,74 g
Sól:	0,38 g

Data produkcji: 26/05/2017  
Należy spożyć do: 05/06/2017  
Nr partii: 1029 V10146  
Masa NETTO: 1 Kg  
Prime, S.A., MCP Casilla 80, 0940, Hiszpania

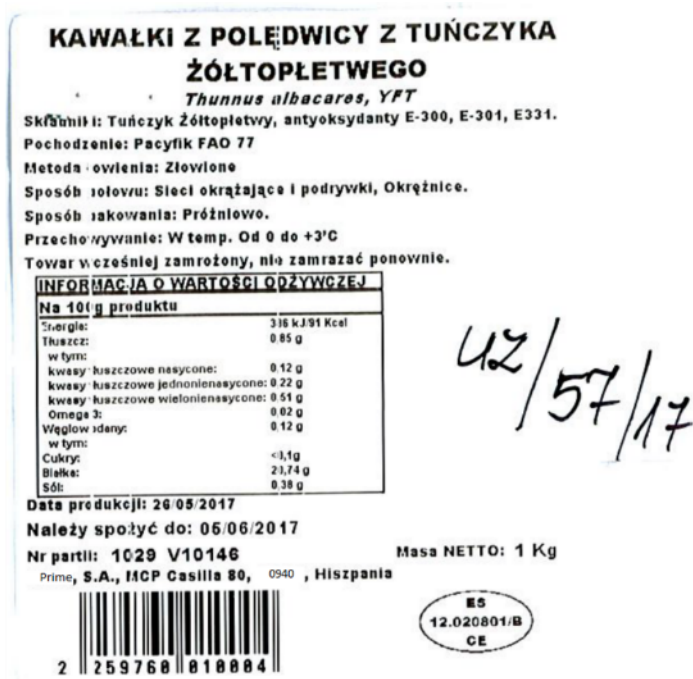
42/57/17



2 259768 818884



# Extraction from attachments: Example 2



## KAWAŁKI Z POLEDWICY Z TUNCZYKA ŻÓŁTOPLETWEGO  
Thunnus albacares, YFT  
SkfSunib 1: Tuńczyk Żółtopletwy, antyoksydanty E-300, E-301, E331.  
Pochodzenie: Pacyfik FAO 77  
Metoda owłoczenia:  
Sposób solowania: Sieci okrężające i podrywki, Okrężnice.  
Sposób pakowania  
Przechowywanie; W temp. Od 0 do +3°C  
Towar wcześniej zamrożony, nie zamrażać ponownie.  
Należy spożyć do: 06.06 2017  
- 1029 V10148 Masa NETTO: 1 Kg  
MCP Casilla 80, 0940, Hiszpania  
<!-- image -->

# Extraction from attachments: Example 2



Uz/57/17



```
## KAWALKI Z POLEDWICY Z TUNCZYKA ZOLTOPLETWEGO
Thunnus aulbacares, YFT
SkfSunib 1: Tuńczyk 2éitopletwy, antyoksydanty E-300, E-301, E331.
Pochodzenie: Pacyfik FAO 77
Motoda ovtlen |s:
Sposéh 10otovsu: Sieci okratajace 1! podrywski, Okrqznice.
Sposób rkowania
Prrechoywenne; W temp. Od Y do +46
Tovear wczéeiniel zamrotony,tg
Naley spoiyc do: 06.06 2017
- 1029 V10148 Mass NETTO: 1 Kg
MCP Casiita $0, 0940, Hiszpania
<!-- image -->
```

SCORE:  
1 out of 1



name	country	region	city	postal_code	street	house_nr
MCP Casiita	Hiszpania	None	None	0940	None	None

# Extraction from attachments: Example 2

Liberté • Égalité • Fraternité  
RÉPUBLIQUE FRANÇAISE  
MINISTÈRE DE L'ÉCONOMIE  
ET DES FINANCES

**RAPPORT D'ESSAIS**  
MAR 2017 - 2070 00 0700/2017

SERVICE COMMUN DES LABORATOIRES  
DE LA DCCRF ET DE LA DGDJ  
LABORATOIRE DE MARSEILLE  
145 TRAVERSE MIEUX SUNISI  
13743 MARSEILLE CEDEX 13

SIVEP SAINT MALO  
12 RUE DE LE KOSMOLOGIE  
F-38350 SAINT MALO

**DÉNOMINATION DE L'ÉCHANTILLON**

LONGE DE THON ( ADDITIF)

**DESCRIPTION DE L'ÉCHANTILLON**

Echantillon reçu dans une barquette mise sous blister, le tout dans un carton.

**RÉSULTATS DES EXAMENS ET ESSAIS EFFECTUÉS SUR L'ÉCHANTILLON**

L34	Nitrates E251-E252 (en NaNO3) - spectre UV-vis • L34-MD-9-12-ANA-17 adapt ISO 2051-1675	<20 (LQ) mg/kg
L34	Nitrites E249-250 (en NaNO2) - spectre UV-vis • L34-MD-9-12-ANA-17 adapt ISO 2918-1675	<3 (LQ) mg/kg
L67	Date de début d'analyse	29/05/2017
L67	Acides ascorbique et isoascorbique - CLHP UV • L67-IN-60-ANA-60	
L67	Acide ascorbique (E300-303)	1960 mg/kg
L67	Acide isoascorbique (Acide érythorbique) (E315-318)	< LD : 1 mg/kg

**COMMENTAIRES PARTICULIERS**

Température à la réception : +14.6°C


**INTERPRÉTATION DES RÉSULTATS**

Les analyses et essais ont été réalisés en regard des :

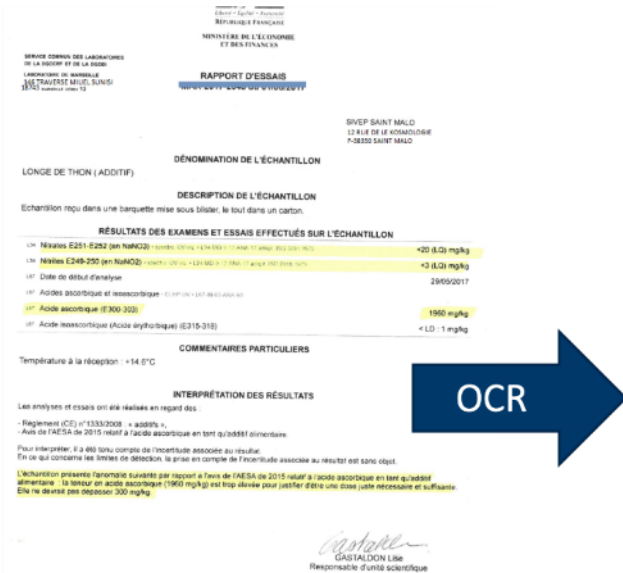
- Règlement (CE) n°1333/2008 : « additifs »,
- Avis de l'AESA de 2015 relatif à l'acide ascorbique en tant qu'additif alimentaire.

Pour interpréter, il a été tenu compte de l'incertitude associée au résultat.  
En ce qui concerne les limites de détection, la prise en compte de l'incertitude associée au résultat est sans objet.

L'échantillon présente l'anomalie suivante par rapport à l'avis de l'AESA de 2015 relatif à l'acide ascorbique en tant qu'additif alimentaire : la teneur en acide ascorbique (1960 mg/kg) est trop élevée pour justifier d'être une dose juste nécessaire et suffisante. Elle ne devrait pas dépasser 300 mg/kg.

  
GASTALDON LiSE  
Responsable d'unité scientifique

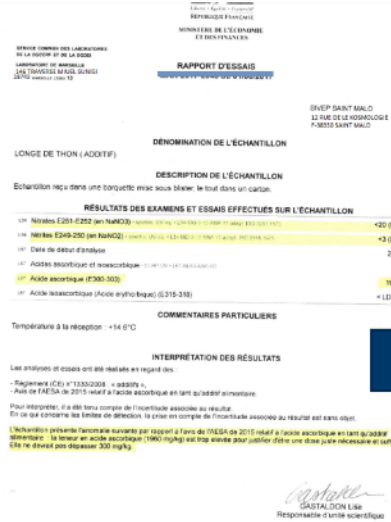
# Extraction from attachments: Example 2



**SERVICE COMMUN DES LABORATOIRES## LABORATOIRE DE MARSEILLE RAPPORT D'ESSAIS 16, TRAVERSE MIGUEL SUNISI** Ta ne CHARLES Sus MAR-2017-2048 du 01/06/201718738 marseille cedex 43## DENOMINATION DE L'ECHANTILLONLONGE DE THON ( ADDITIF)## DESCRIPTION DE L'ÉCHANTILLONEchantillon reçu dans une barquette mise sous blister. le tout dans un carton## RÉSULTATS DES EXAMENS ET ESSAIS EFFECTUÉS SUR L'ECHANTILLON | 2 Nurites 249-290 (en NaNO2) : spectre U-vis + L34 MO 12 ANA 17 adant ISO 2918 1974 <3 (LQ) ma/kc ||-----|| L67 Date de début d'analyse 29/05/2017 || 67 AClides ascorbique et isoascorbique : « ve IN. || Acide isoascorbique (Acide érythorbique) (E315-318) < LD : 1 mg/kg ## COMMENTAIRES PARTICULIERSTempérature à la réception : +14 6°C## INTERPRÉTATION DES RÉSULTATSLes analyses et essais ont été réalisés en regard des :- Règlement (CE) n°1743/2008 : « additifs ».- AVIS de l'AESA de 2015 relatif à l'acide ascorbique en tant qu'additif alimentairePour interpréter, il a été tenu compte de l'incertitude associée au résultatEn ce qui concerne les limites de détection, la prise en compte de l'incertitude associée au résultat est sans objet.<!-- image -->GASDONNE MarieResponsable d'unité scientifiqueA<!-- image -->## ÈRE DE L'ÉCONOMIESIVEP SAINT MALO 12 RUE DE LE KOSMOLOGIE F-38350 SAINT MALO## RAPPORT D'ESSAIS MAR-2017-2938 du 01/06/2017 (Suite)12 RUE DE LE KOSMOLOGIE r-38350 SAINT MALO Les résultats ne se rapportent qu'à l'échantillon soumis aux examens et essais ] : essai(s) effectué(s) au SCL Laboratoire de Montpellier [L67] : essai(s) effectué(s) au SCL Laboratoire de StrasbourgLa reproduction de ce rapport, qui comporte 2 pages, n'est autorisée que sous sa forme intégrale. MINISTESERVICE COMMUN DES LABORATOIRES Bordereau de transmission du rapport Le 01/06/2017 DE LA DGCCRF ET DE LA DGDD!146, TRAVERSE MIGUEL SUNISI 18743 MARSEILLE CEDEX 13Suivi par : RELUNDE Jackline ~ LMé! LAB016@scl.finances.gouv.fr F-35870 SAINT MALOTéléphone : 04 923 56 235 00 OIVEP SAINT MALO Télécopie : 049161 8219 4 RUE DE LE KOSMOLOGIENos références | Vos référencesRapport MAR-2017-2048 du 01/06/2017Demande reçue le 19/05/2017Suivi par LANDURE Jackline Demande Mamanda ramia la ANINEMNAT DossierDemande reçue le 19/05/2017Dénomination : LONGE DE THON ( ADDITIF)



# Extraction from attachments: Example 2



**SERVICE COMMUN DES LABORATOIRES**## LABORATOIRE DE **MARSEILLE** RAPPORT **D'ESSAIS 16, TRAVERSE MIGUEL SUNISI** Ta ne CHARLES Sus MAR-2017-2048 du 01/06/201718738 marseinte cedex 43## DENOMINATION DE L'ECHANTILLONLONGE DE THON ( ADDITIF)## DESCRIPTION DE L'ECHANTILLONEchantillon reçu dans une barquette mise sous blister. le tout dans un carton## RÉSULTATS DES EXAMENS ET ESSAIS EFFECTUÉS SUR L'ECHANTILLON | 2 Nurites 249-290 (en NaNO2) : spectre U-vis + L34 MO 12 ANA 17 adant ISO 2918 1974 <3 (LQ) ma/kc ||-----|| L67 Date de début d'analyse 29/05/2017 || 67 AClides ascorbique et isoascorbique : « ve IN. || Acide isoascorbique (Acide érythorbique) (E315-318) < LD : 1 mg/kg ## COMMENTAIRES PARTICULIERSTempérature à la réception : +14 6°C## INTERPRÉTATION DES RÉSULTATSLes analyses et essais ont été réalisés en regard des :- - Règlement (CE) n°1743/2008 : « additifs ».- - AVIS de l'AESA de 2015 relatif à l'acide ascorbique en tant qu'additif alimentairePour interpréter, il a été tenu compte de l'incertitude associée au résultatEn ce qui concerne les limites de détection, la prise en compte de l'incertitude associée au résultat est sans objet.<!-- image -->GASDONNE MarieResponsable d'unité scientifiqueA<!-- image -->## ÊRE DE **L'ÉCONOMIESIVEP SAINT MALO 12 RUE DE LE KOSMOLOGIE F-38350 SAINT MALO**## RAPPORT D'ESSAIS MAR-2017-2938 du 01/06/2017 (Suite)12 **RUE DE LE KOSMOLOGIE r-38350 SAINT MALO**Les résultats ne se rapportent qu'à l'échantillon soumis aux examens et essais ] : essai(s) effectué(s) au SCL Laboratoire de Montpellier [L67] : essai(s) effectué(s) au SCL Laboratoire de StrasbourgLa reproduction de ce rapport, qui comporte 2 pages, n'est autorisée que sous sa forme intégrale. MINISTESERVICE COMMUN DES LABORATOIRES Bordereau de transmission du rapport Le 01/06/2017 DE LA DGCCRF ET DE LA DGDD!**146, TRAVERSE MIGUEL SUNISI 18743 MARSEILLE CEDEX** 13Suivi par : RELUNDE Jackline ~~~ LMél LAB016@scl.finances.gouv.fr F-35870 SAINT MALOTéléphone : 04 923 56 235 00 OIVEP SAINT MALO Télécopie : 049161 8219 4 RUE DE LE KOSMOLOGIENos références | Vos référencesRapport MAR-2017-2048 du 01/06/2017Demande reçue le 19/05/2017Suivi par LANDURE Jackline Demande Mamanda ramia la ANINEMNAT DossierDemande reçue le 19/05/2017Dénomination : LONGE DE THON ( ADDITIF)

SCORE:  
2 out of 2



name	country	region	city	postal_code	street	house_nr
SERVICE COMMUN DES LABORATOIRES	None	None	Marseille	18743	TRAVERSE MIGUEL SUNISI	146
SIVEP SAINT MALO	None	None	SAINT MALO	38350	RUE DE LE KOSMOLOGIE	12



# Conclusion

## Limiting factors:

- accuracy of doc. parsing
- accuracy of OCR
  - doc/image resolution
- Text to Markdown
- size of the LLM
  - accuracy LLM

## Additional concerns:

- Privacy
  - Local models can provide the same accuracy but require similar infrastructure

# 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)

# 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](http://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)