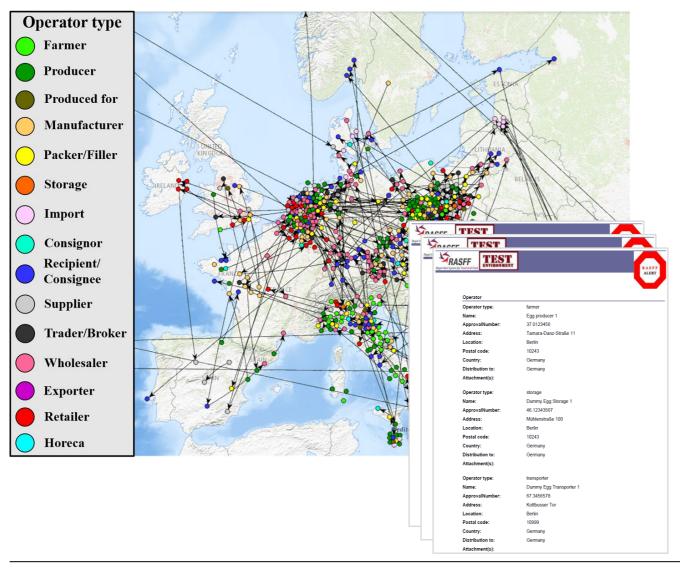


Al-supported data extraction: The Rapid Alert Supply Network EXtractor (RASNEX) tool for mining supply chain information

15.05.2025

Daria Savvateeva, Matthew Salewski & Marc C. Lorenzen

Contamination in the Supply Chain



Need to stop outbreaks as fast as possible!

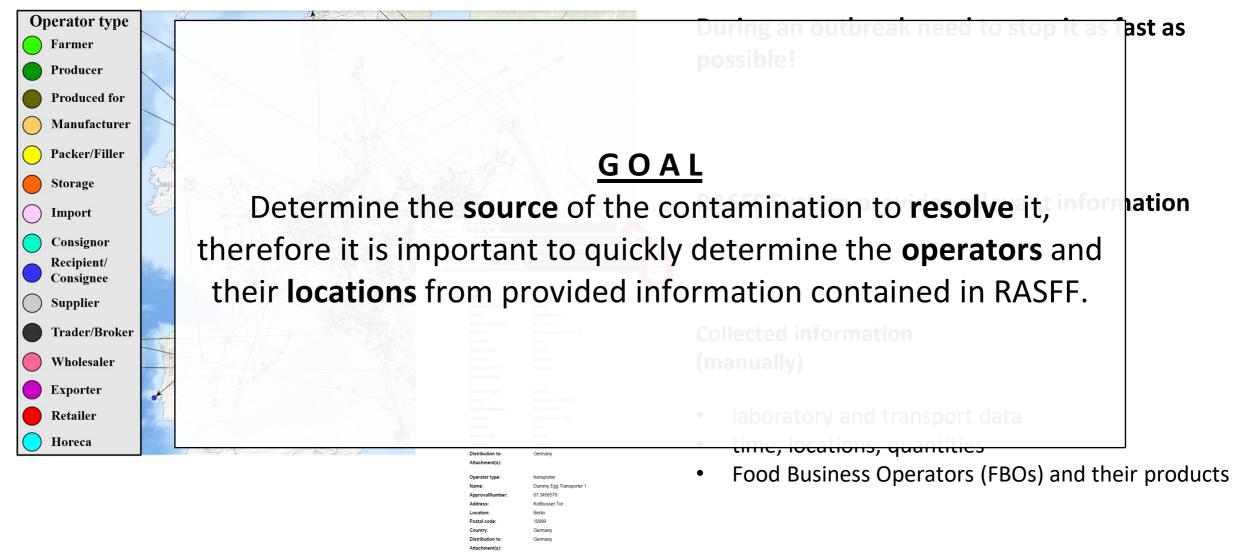
Rapid Alert System for Food and Feed (RASFF) provides relevant information

Collected information (manual work)

- laboratory and delivery data
- times, locations, quantities
- Food Business Operators (FBOs) and affected products

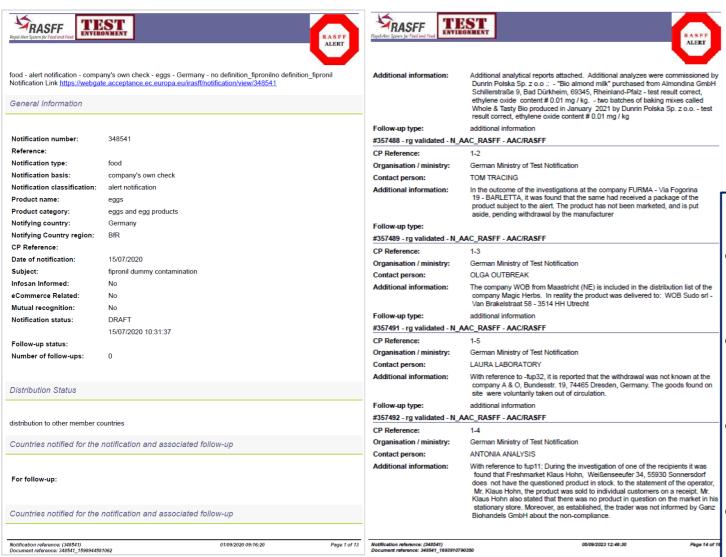


Contamination in the Supply Chain





Rapid Alert System for Food and Feed (RASFF) notifications

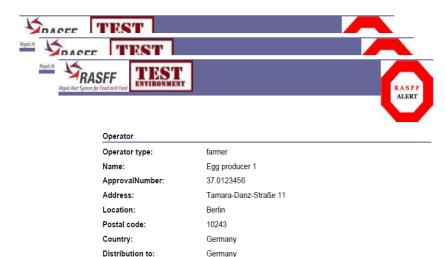


Attachments:

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



How to extract data from RASFF?



Attachment(s):

Operator type: storage

Name: Dummy Egg Storage 1
ApprovalNumber: 46.12343567

Berlin

Address: Mühlenstraße 100

Postal code: 10243
Country: Germany
Distribution to: Germany

Attachment(s):

Location:

Operator type: transporter

Name: Dummy Egg Transporter

ApprovalNumber: 67.3456578
Address: Kottbusser Tor
Location: Berlin
Postal code: 10999
Country: Germany
Distribution to: Germany

Attachment(s):

Challenges:

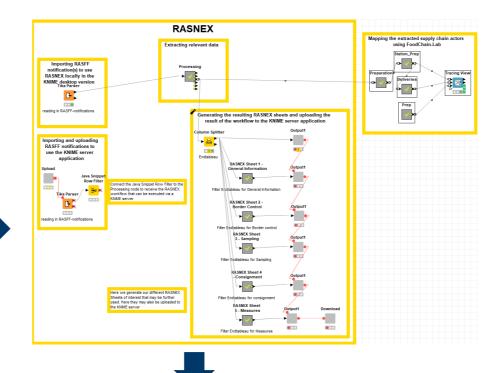
- Manual work that is done in multiple institutions and Member States in parallel
- Non-standardised & unstructured formatting of attached documents
- Multiple languages
- Multiple address formats amongst Member States
- Continious flow of information (updates on the RASFF)
- → How can we support people working with RASFF?



Where do we come from? RASNEX 1.0



Structured RASFF data to tabular format

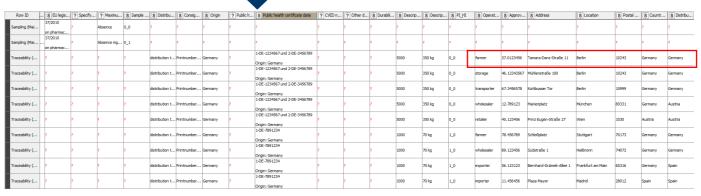


KNIME-based solution

"Konstanz Information Miner", https://www.knime.com/

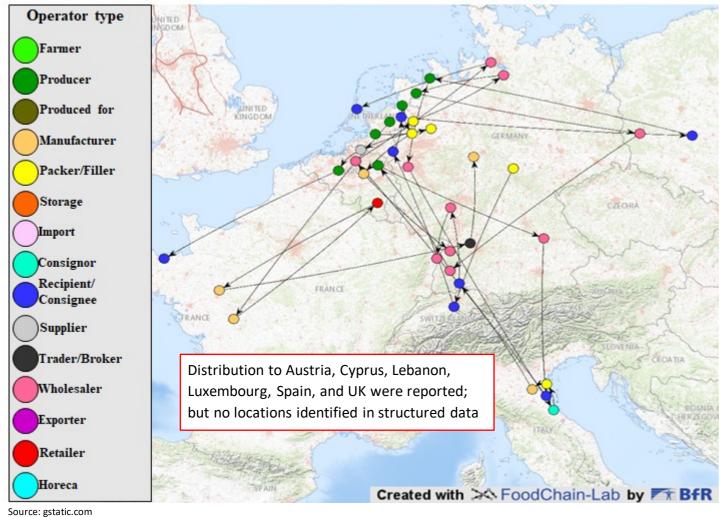
Low-code platform for data analytics and reporting

free and open-source, and locally hosted





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



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)

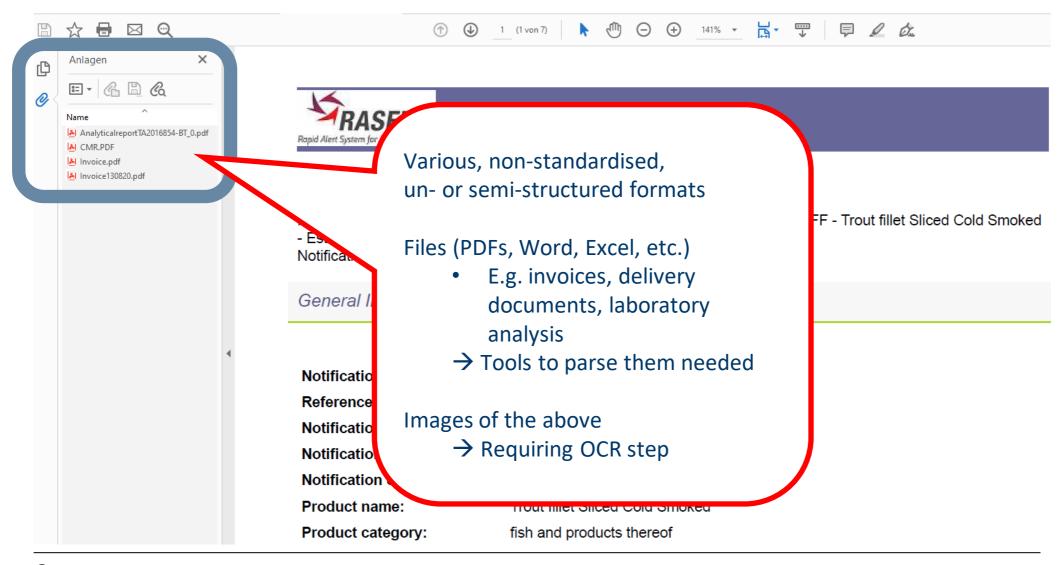


- 1. Extract relevant information from fups & attached documents
 - Extract recognized addresses
 - Operators (names) of supply chains
 - Intent of the fup
 - Products
 - Analytical results
 - Date of production
 - Billing dates
 - Delivered quantities
- 2. Linking RASNEX with the FoodChain-Lab data extraction workflow (UTX format)

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



RASFF notifications – attached documents





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

additional information

Follow-up type: #357488 - rg validated - N AAC RASFF - AAC/RASFF

CP References 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:

#357489 - rg validated - N_AAC_RASFF - AAC/RASFF

CP Reference:

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:

Organisation / ministry: German Ministry of Test Notification

LAURA LABORATORY Contact person:

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:

Organisation / ministry: German Ministry of Test Notification

ANTONIA ANALYSIS Contact person:

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

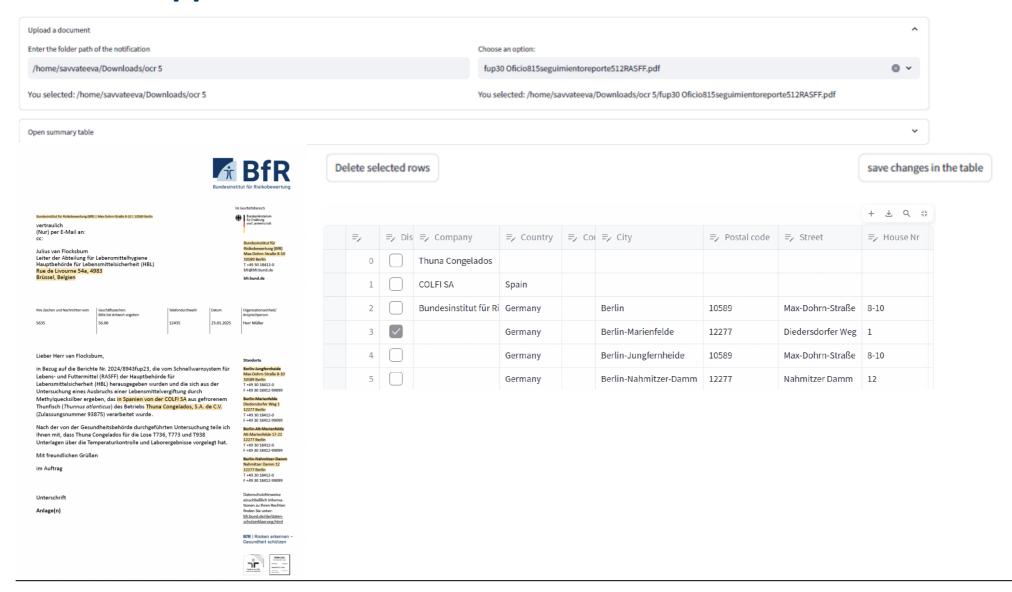
Backend methods (swappable components):

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



Notification reference: (348541)

RASNEX App: Dashboard

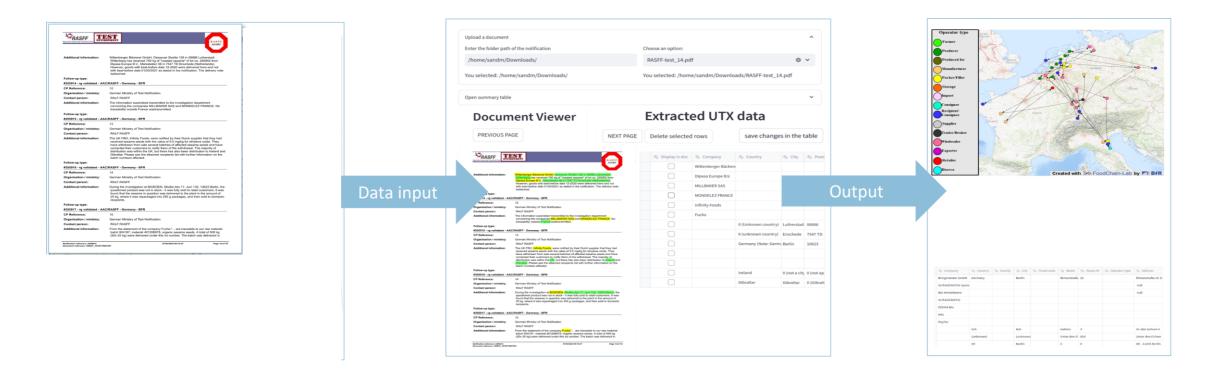




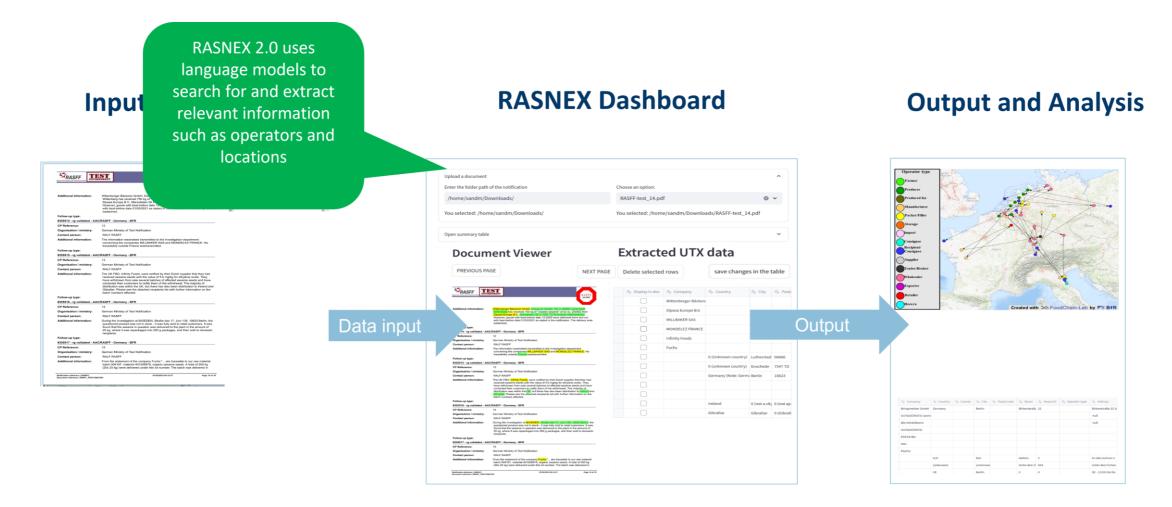
Input documents

RASNEX Dashboard

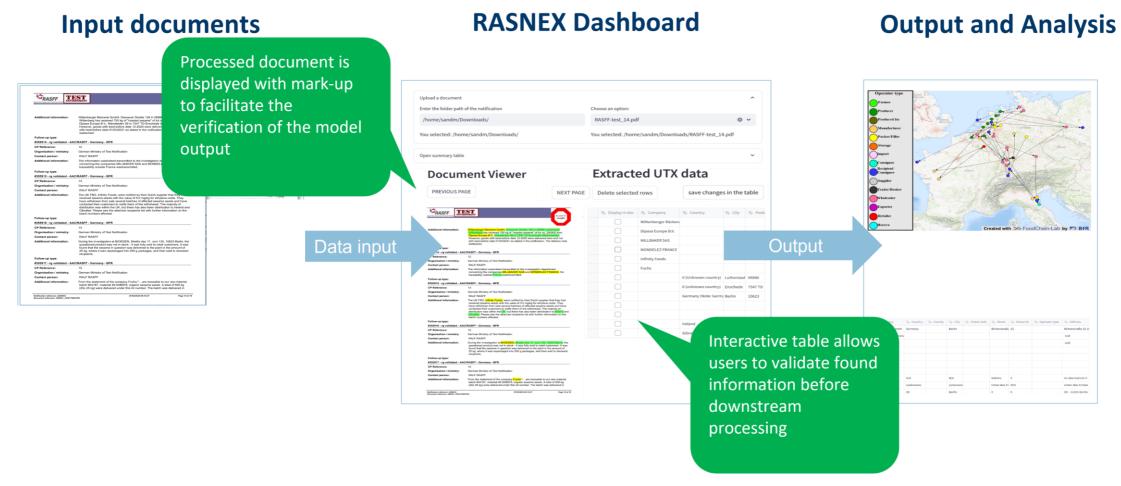
Output and Analysis









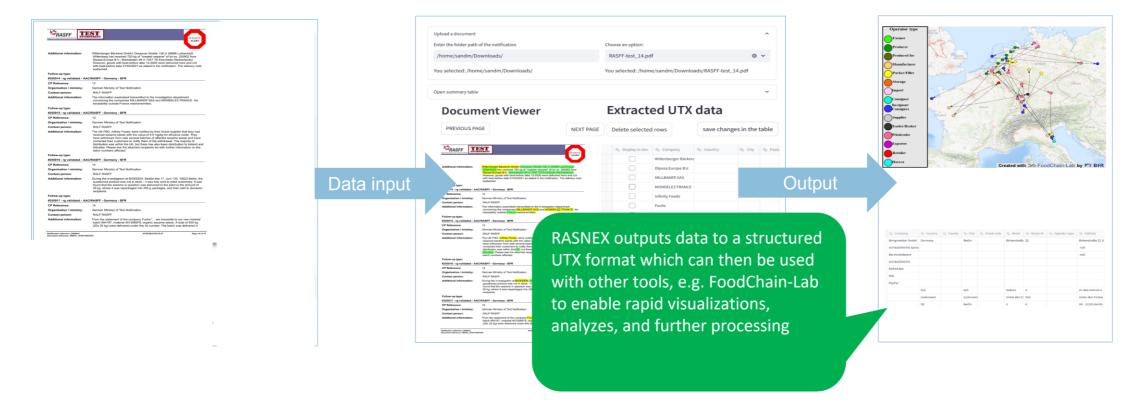




Input documents

RASNEX Dashboard

Output and Analysis





Live-Demo



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

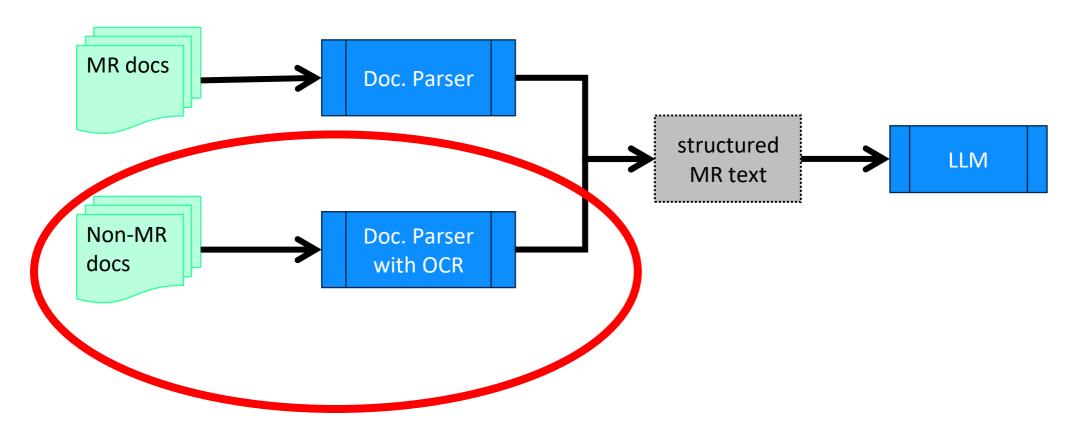
Document parsing with Docling

Optical caracter recognition

Structured output of required info (with the large model)



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





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, document parsers based on Computer Vision (CV) and Optical Character Recognition (OCR)



Group discussion

Limitations of the proof-of-principle:

- 1. data quality -> scans, etc.
- 2. processing capabilities
- -> small models with limited accuracy, showing promising results
- -> server solution tested currently with dummy data
- 1. How are you dealing with data extraction?
- 2. Usage of the RASFF System?

- 3. Where could you see advantages of RASNEX in your daily work?
- 4. What would you need to be able to apply RASNEX?
- 5. Limitations for using AI for outbreak investigation in the UK?





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)

