Intégration et développement d'approches de fouille de textes pour la veille épidémiologique en santé animale

Mathieu Roche

Cirad, UMR TETIS, Montpellier

Montpellier, le 18 septembre 2025

Context

Text-Mining in Agriculture

[Drury and Roche, CEA'2019]

PADI-web pipeline

Google News

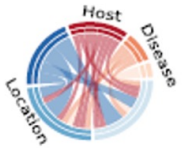
PADI WEB

Text-Mining

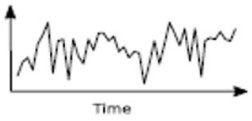
- Disease
- Host
- Location
- Date
- Number of cases
- Symptom

- AFRICAN SWINE FEVER
- ANTIMICROBIAL RESISTANCE
- AVIAN INFLUENZA
- BLUETONGUE
- BRUCELLA
- CANINE DISEASE
- CHIKUNGUNYA
- CHOLERA
- CLASSICAL SWINE FEVER
- CRIMEAN-CONGO HAEMORRHAGIC FEVER
- DENGUE
- DISEASES OF BEES
- FELIN DISEASE
- FOOT AND MOUTH DISEASE
- LEGIONELLA
- LEPTOSPIROSIS
- LUMPY SKIN DISEASE
- MPOX
- NIPAH
- ONE HEALTH
- PESTE DES PETITS RUMINANTS
- RIFT VALLEY FEVER
- SCHMALLENBERG VIRUS
- SMALL RUMINANTS
- TICK-BORNE ENCEPHALITIS
- TROPILAELAPS
- TULARAEMIA
- UNSPECIFIED DISEASE
- USUTU
- WEST NILE VIRUS
- ZIKA


Spatial Dimension



Thematic Dimension



Temporal Dimension


Source Dimension

[Arinik *et al.*, IEEE Access'2023]

PADI-web input

Language and country labels

Chile: Mysterious mass extinction of guano cormorants

DE → EN Show source language CL

ID: 8QTA0BGS1Q

Jun 4, 2023 · Jun 5, 2023 · Visit page

Source: latina-press.com

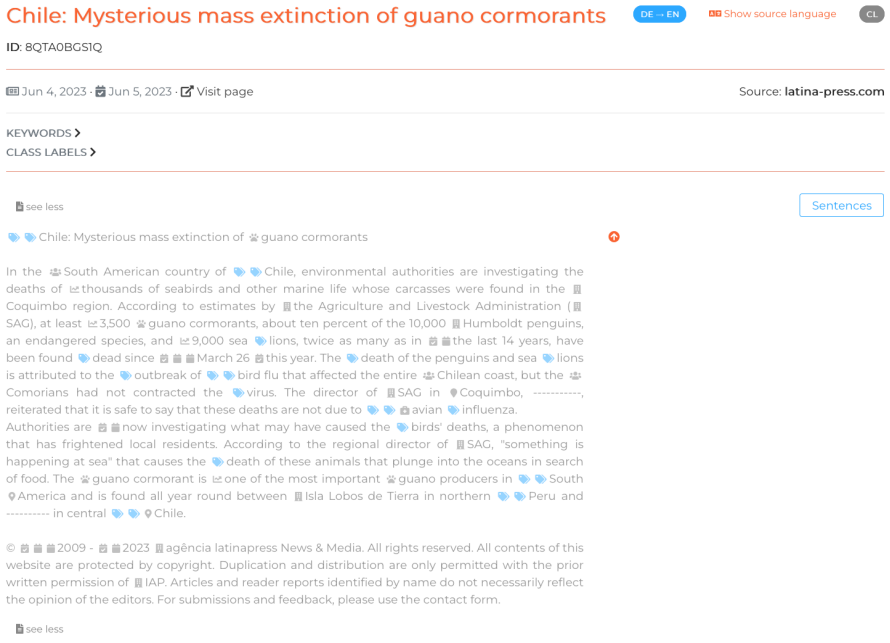
Chile: Mysteriöses Massensterben von Guano-Kormoranen

Die Behörden untersuchen nun, was den Tod der Vögel verursacht haben könnte, ein Phänomen, das die Anwohner verängstigt hat (Foto: Agência Ambiental Pick-upau)

Datum: **04. Juni 2023**
Uhrzeit: **15:49 Uhr**
Ressorts: **Chile, Natur & Umwelt**
Leserecho: **0 Kommentare**
Autor: **Redaktion**

Im südamerikanischen Land Chile untersuchen die Umweltbehörden den Tod von Tausenden von Seevögeln und anderen Meeresbewohnern, deren Kadaver in der Region Coquimbo aufgefunden wurden. Nach Schätzungen der Landwirtschafts- und Viehzuchtbehörde (SAG) sind seit dem 26.

PADI WEB



Google News

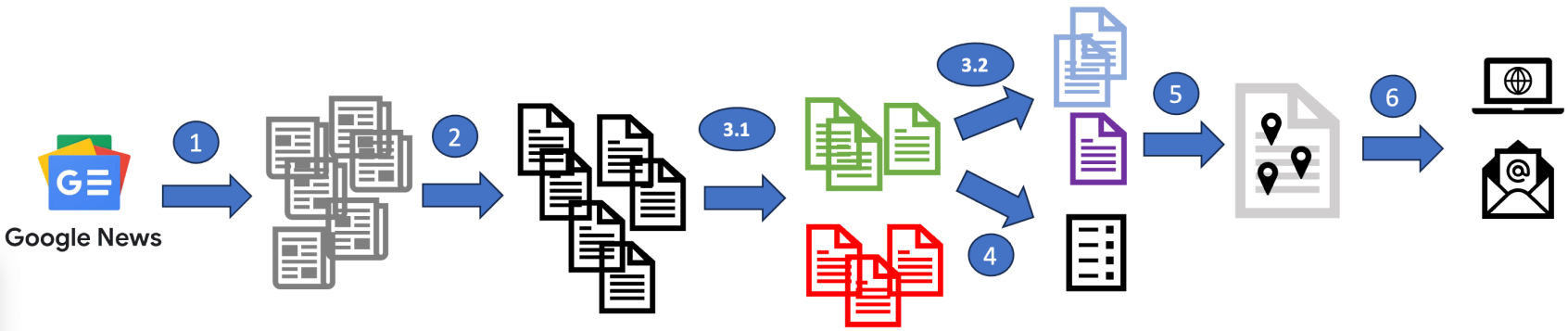
PADI-web pipeline

[Valentin *et al.*, One Health 2021]

PADI-web *Animal Health*

PADI-web *Plant Health*

PADI-web *Public Health*



Step 1: Data collection
- Multlingual articles

Step 2: Processing
- Text cleaning
- Language detection
- Translation into English

Step 3: Document classification
- **Step 3.1:** Relevant/irrelevant documents
- **Step 3.2:** Topic of documents

Step 4: Sentence classification
- Topic of sentences

Step 5: Information extraction
- Disease
- Location
- Host
- Symptom
- etc.

Step 6: Output for end-users
- Visualisations
- Notifications

PADI-web process

Chile: Mysterious mass extinction of guano cormorants

DE → EN | Show source language | CL

ID: 8QTA0BGS1Q

Jun 4, 2023 · Jun 5, 2023 · Visit page — Source: latina-press.com

KEYWORDS >

CLASS LABELS

Relevance	RELEVANT
Fine-grained classificati...	OUTBREAK DECLARATION
Topic (BERT approach)	OUTBREAK DECLARATION

Class (topic)

see less

Sentences

Chile: Mysterious mass extinction of guano cormorants

In the South American country of Chile, environmental authorities are investigating the deaths of thousands of seabirds and other marine life whose carcasses were found in the Coquimbo region. According to estimates by the Agriculture and Livestock Administration (SAG), at least 3,500 guano cormorants, about ten percent of the 10,000 Humboldt penguins, an endangered species, and 9,000 sea lions, twice as many as in the last 14 years, have been found dead since March 26 this year. The death of the penguins and sea lions is attributed to the outbreak of bird flu that affected the entire Chilean coast, but the Comorians had not contracted the virus. The director of SAG in Coquimbo, ----------, reiterated that it is safe to say that these deaths are not due to avian influenza.
Authorities are now investigating what may have caused the birds' deaths, a phenomenon that has frightened local residents. According to the regional director of SAG, "something is happening at sea" that causes the death of these animals that plunge into the oceans in search of food. The guano cormorant is one of the most important guano producers in South America and is found all year round between Isla Lobos de Tierra in northern Peru and ---------- in central Chile.

Thematic Dimension

Chile: Mysterious mass extinction of guano cormorants

DE → EN | Show source language | CL

ID: 8QTA0BGS1Q

Jun 4, 2023 · Jun 5, 2023 · Visit page — Source: latina-press.com

KEYWORDS

USER KEYWORDS

User dictionaries

disease	AVIAN INFLUENZA, INFLUENZA
host	AVIAN, MAMMAL
symptom	MORTALITY
various	OUTBREAK, VIRUS
location	AMERICAS, REPUBLIC OF CHILE, REPUBLIC OF PERU, SOUTH AMERICA

Keywords (epidemiological information)

AUTOMATIC KEYWORDS

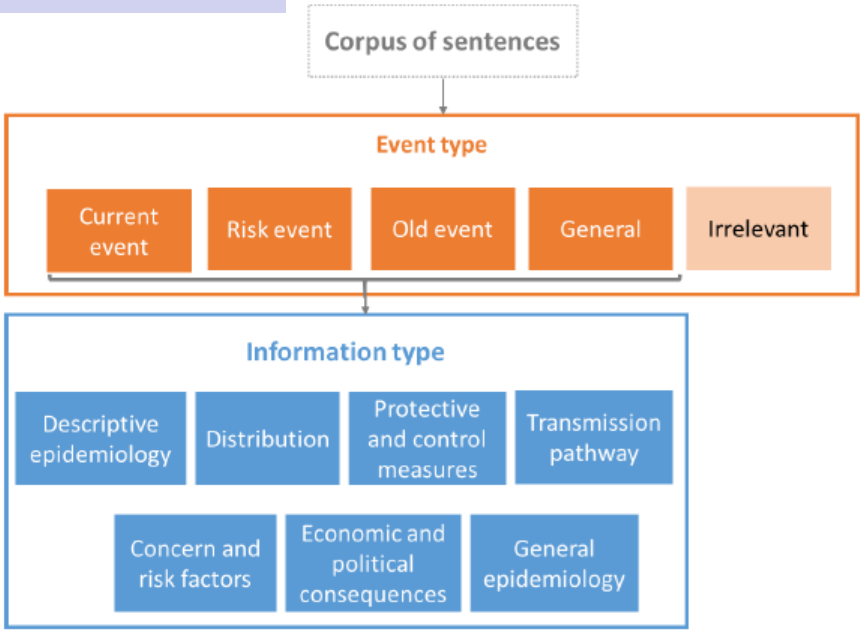
Named Entity Recognition

Extraction Institution	AGÊNCIA LATINAPRESS NEWS & MEDIA, COQUIMBO, HUMBOLDT, IAP, ISLA LOBOS DE TIERRA, SAG, THE AGRICULTURE AND LIVESTOCK ADMINISTRATION
Extraction People group	CHILEAN, COMORIANS, SOUTH AMERICAN
Extraction Location	COQUIMBO

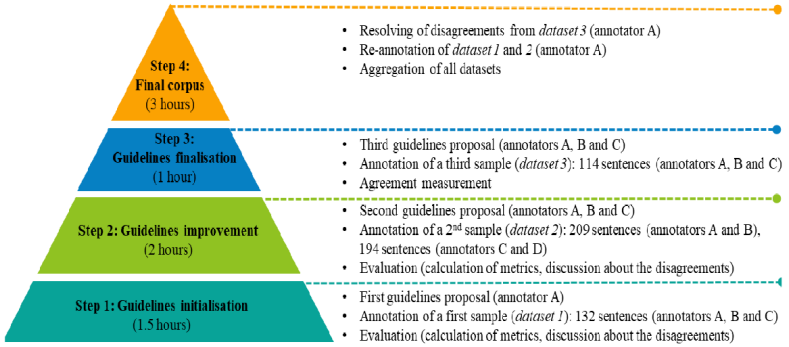
PADI-web process

Fine-grained classification

Data annotation



Thematic Dimension



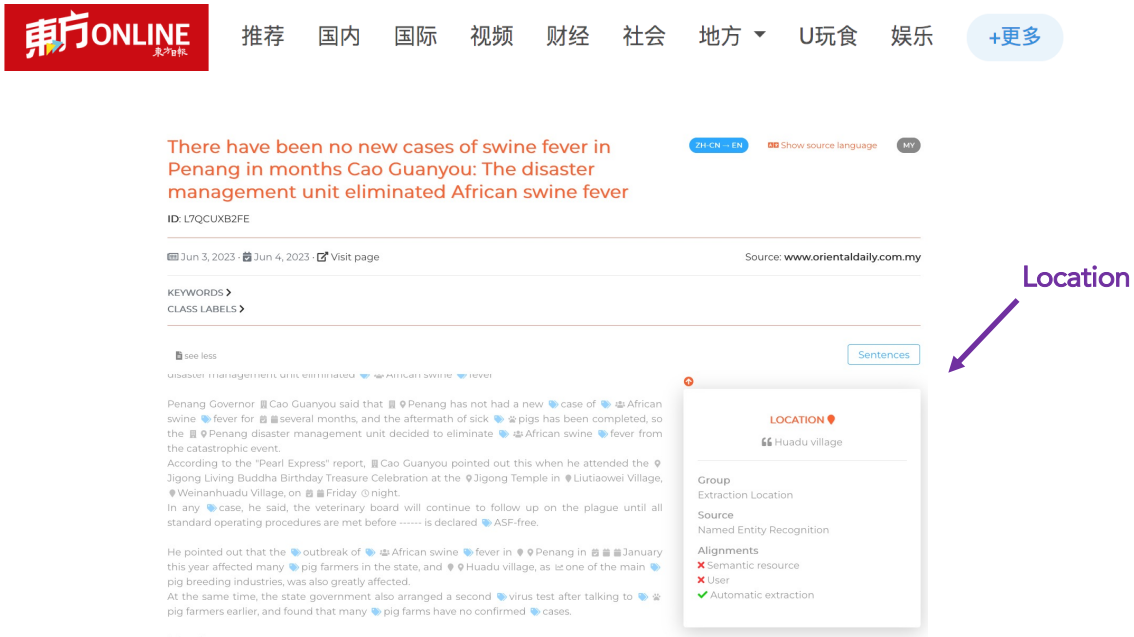
Inter-annotator agreement

		Total agreements	Partial agreements	Disagreements	κ
Step 1	Event type	29%	48%	23%	0.30
	Information type	49%	43%	8%	0.53
Step 3	Event type	87%	19%	4%	0.71
	Information type	75%	22%	3%	0.78

[Valentin *et al.*, Scientific Data'2022]

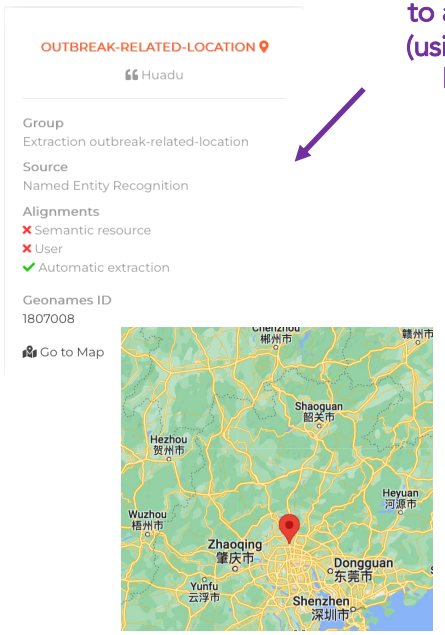
PADI-web process

Spatial Dimension

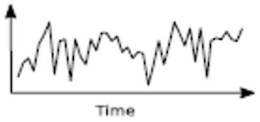


Location

Location related to an outbreak (using Machine Learning)



PADI-web process



Temporal Dimension

東方ONLINE 推荐 国内 国际 视频 财经 社会 地方 U玩食 娱乐 +更多

There have been no new cases of swine fever in Penang in months Cao Guanyou: The disaster management unit eliminated African swine fever

ZH-CN → EN | Show source language | MY

ID: L7QCUXB2FE

Jun 3, 2023 · Jun 4, 2023 · Visit page

Source: www.orientaldaily.com.my

KEYWORDS >

CLASS LABELS >

see less | Sentences

Penang Governor Cao Guanyou said that Penang has not had a new case of African swine fever for several months, and the aftermath of sick pigs has been completed, so the Penang disaster management unit decided to eliminate African swine fever from the catastrophic event.
According to the "Pearl Express" report, Cao Guanyou pointed out this when he attended the Jigong Living Buddha Birthday Treasure Celebration at the Jigong Temple in Liutiaowei Village, Weinanhuadu Village, on Friday night.
In any case, he said, the veterinary board will continue to follow up on the plague until all standard operating procedures are met before ------ is declared ASF-free.

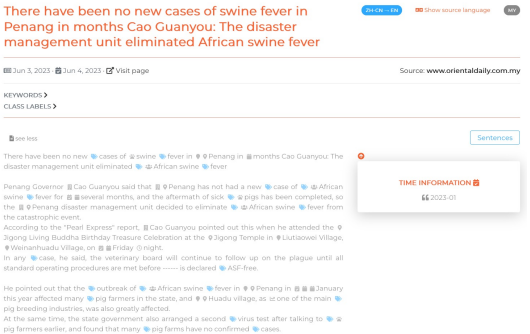
He pointed out that the outbreak of African swine fever in Penang in January this year affected many pig farmers in the state, and Huadu village, as one of the main pig breeding industries, was also greatly affected.
At the same time, the state government also arranged a second virus test after talking to pig farmers earlier, and found that many pig farms have no confirmed cases.

LOCATION
Huadu village

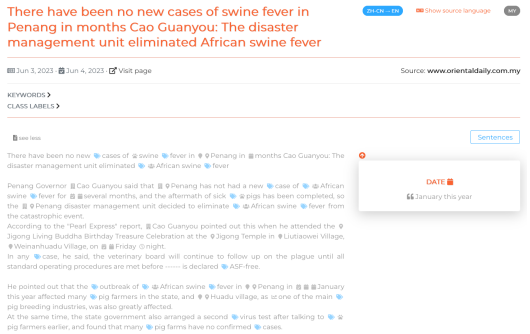
Group
Extraction Location

Source
Named Entity Recognition

Alignments
✗ Semantic resource
✗ User
✓ Automatic extraction



Temporal information (*normalised*)



Temporal information (*in text*)

PADI-web output

Outputs

Email Notifications

Website consultation

CSV

14 new texts have been collected between **04 November, 2024 (00:00)** and **05 November, 2024 (07:00)**.

> Go to search page

If this email is not fully working in your mail client, please click here to read it in PADI-web.

CONTENTS

- Bluetongue (1)
  - Africa (1)
- African Swine Fever (6)
  - Europe (5)
  - Unknown (1)
- Avian Influenza (5)
  - Asia (1)
  - Europe (3)
  - North America (1)
- West Nile Virus (2)
  - Europe (1)
  - Unknown (1)
- Classification Information
- Keywords and Extracted Information

Bluetongue (1) ↑

Africa (1) ↑

Libyan Arab Jamahiriya (1) ↑

1 **Bluetongue disease affects sheep populations in Libya**
04 November, 2024 LIBYAN ARAB JAMAHIRIYA EN

Bluetongue disease affects sheep populations in Libya A veterinary official inspects sheep for signs of bluetongue disease at a local farm.The head of the National Center for Animal Health, Abdulrahman Ajbil, has confirmed an outbreak of blue-tongue disease affecting sheep across eastern and western regions of the country. Speaking to Libyan Express, Ajbil reassured that despite the spread, sheep mortality rates remain low.Bluetongue is a viral disease affecting ruminants, particularly sheep, and is characterised by inflammation of the mouth, nose, and tongue. While not transmissible to human...

www.libyanexpress.com
BLUETONGUE3_EN

RELEVANT | ALERT, PREPARDNESS | OUTBREAK DECLARATION

African Swine Fever (6) ↑

Europe (5) ↑

Italy (4) ↑

1 **600,000 euros for the fight against African swine fever in Lombardy**
04 November, 2024 ITALY IT

Different view points to define an event

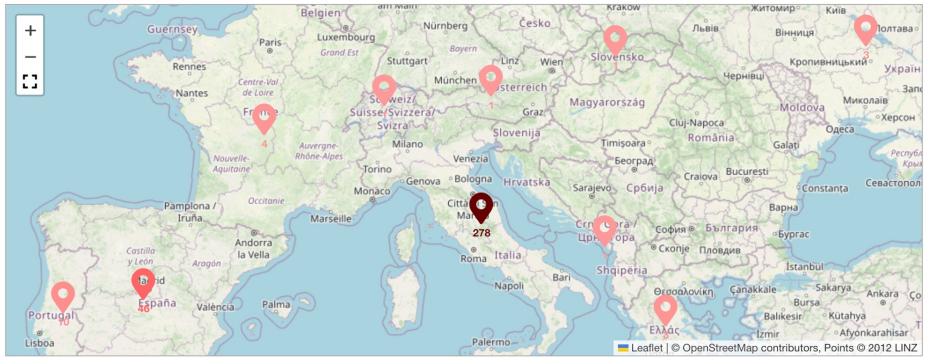
MUST
Multi-source Surveillance Tool

Generate Map

Only show article countries

Clusters

Maximum article number 500



PADI - WEB

Alarm as HPAI kills almost 9,000 sea creatures in Chile

May 30, 2023 · Jun 23, 2023 · Visit page — Source: www.wattagnet.com

KEYWORDS
CLASS LABELS

<p><p>Alarm as HPAI kills almost 9,000 sea creatures in Chile</p>
<p>Over 8,887 sea creatures have been recorded dead on the Chilean coast so far this year due to infection with...

Sentences

see more

Source feed Avian Influenza Mammals (EN) https://news.google.com/rss/search?q=avian+influen... · (RSS Feed)

Sernapesca confirms first case of Avian Influenza in marine mammal in Aysén region

ES → EN · Show source language

May 29, 2023 · Jul 27, 2023 · Visit page — Source: www.sernapesca.cl

KEYWORDS
CLASS LABELS

<p><p>Sernapesca confirms first case of Avian Influenza in marine mammal in Aysén region</p>
<p>With this finding, there are 13 regions with confirmed cases of Avian Influenza i...

Sentences

see more

Source feed Avian Influenza Mammals (ES) https://news.google.com/rss/search?q=gripe+aviar+m... · (RSS Feed)

Chile: Nearly 9,000 marine species hit by avian flu

FR → EN · Show source language

May 25, 2023 · Dec 15, 2023 · Visit page — Source: www.sciencesetavenir.fr

KEYWORDS
CLASS LABELS

<p><p>Chile: Nearly 9,000 marine species hit by avian flu</p>
<p>Nearly 9,000 sea lions, penguins, otters and small cetaceans have died since the beginning of the year in Chile ...

Sentences

see more

MUST

MUST
Multi-source Surveillance Tool

[Trevennec *et al.*, KES 2024]

MOOD

ASTRE
Animal health
Territories Risks Ecosystems

Carlène Trevennec
French Animal Health
Epidemic Intelligence

Provide a comprehensive view of emerging disease events

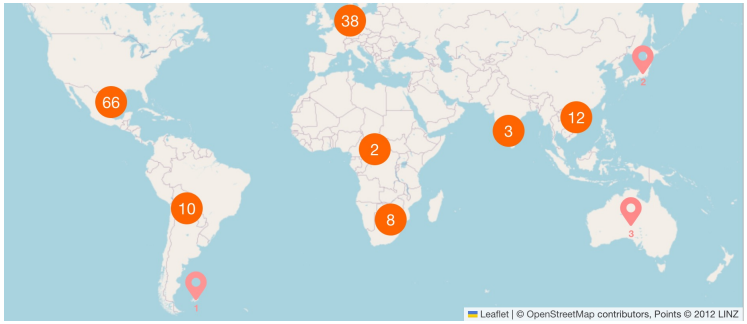
MUltisource Surveillance Tool (MUST) for Event-Based Surveillance (in the public health field)

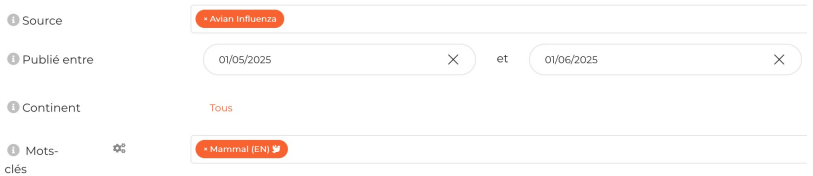
Public interface

- Extract disease event from various sources: official (IBS) and unofficial (EBS)
- Space and time visualisation
- Add complementary context information related to the outbreaks

Advanced interface

- EBS optimization (validation of events, model training)
- Epidemiological analyses





Avian influenza:

Recent increase in reported infections in **mammals** and sporadic cases in humans
Pandemic potential

Important for the French epidemic intelligence team – international monitoring of animal health, risk of introduction in France

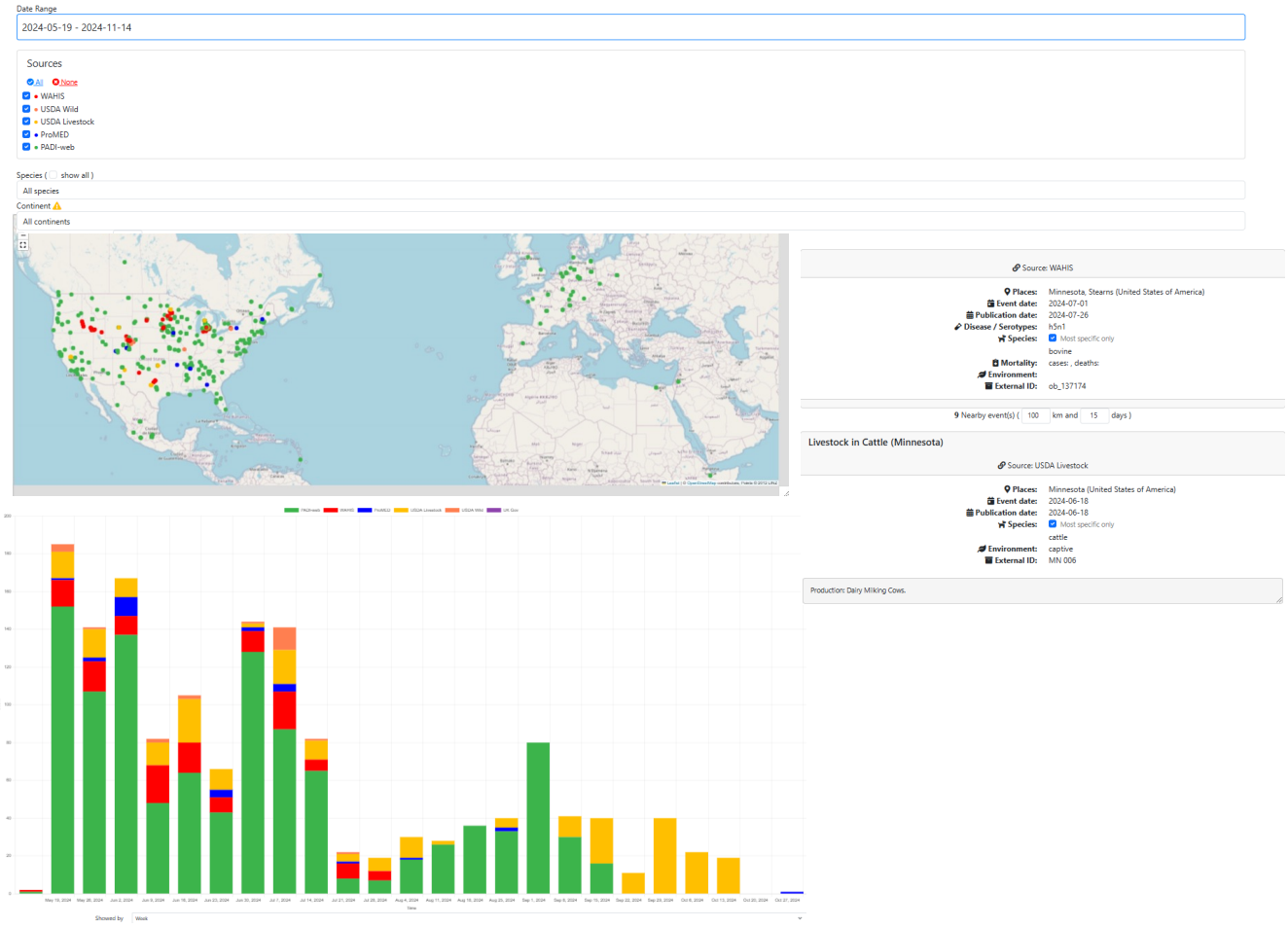
MUST-AI integrates 3 different sources:

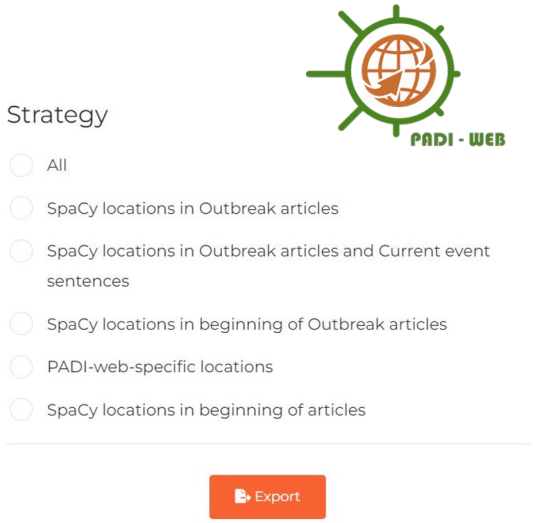
- **WAHIS**: official source of public health notifications (database)
- **PADI-web**: unofficial sources that collect events from online media (Google news…)
- **ProMED**: expert networks in infectious diseases (mails and gradually a database)

=> collecting unstructured data from ProMED and PADI-web and extracting valuable information to make comparisons and complete official data

MUST-AI INTERFACE

Open access tool:
https://must-surveillance.com/





Various types of extractions are proposed according on a number of criteria:

- Locations extracted at the beginning of the outbreak articles or not
- Locations extracted with SpaCy learned with labeled data or not

Relevant articles or not: an article is considered very relevant if its publication date is before the date of the report or official notification

[Trevennec *et al.*, KES 2024]

Fusion

There are three main criteria to determine if two events are similar:

- **Geographical proximity:** if the distance is less than a reference distance
- **Temporal proximity:** if the difference in dates is less than a specified maximum of days
- **Species proximity:** if the Levenshtein distance of the species names is less than a specified maximum value

Work in progress

- EpidGPT: A Combined Strategy to Discriminate Between Redundant and New Information for Epidemiological Surveillance Systems

Doc 1: African swine fever alert in Arad. Hundreds of pigs will be slaughtered The Local Center for Disease Control (CLCB) Arad approved a plan of measures to prevent the spread of the disease ... The Romanian authorities decided that Pig housing buildings and equipment will be disinfected...

Doc 2: An outbreak of African swine fever has been confirmed in a farm with over 300 pigs in Frumusani commune, Aluniş village, Arad county, county authorities announced today, Friday – August 18, which specified that all animals will be euthanized.

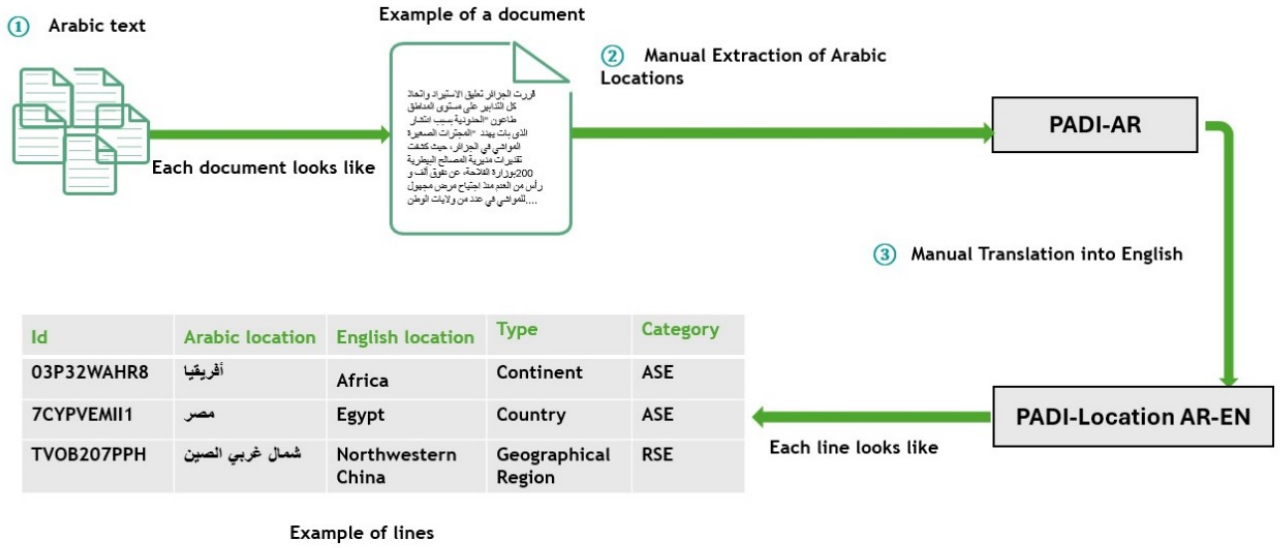
Model	Prec%	Rec%	F_1%
$BERT(PADIWeb_{novel} + BookCorpus)$	81.49	79.63	80.55
$RoBERTa(PADIWeb_{novel} + BookCorpus)$	80.94	79.53	80.23
$BioBERT(PADIWeb_{novel} + PubMedAbstracts)$	85.67	87.34	86.50
$BioELECTRA(PubMed + PADIWeb_{novel})$	90.67	88.34	89.49
$NovelEpidGPT(WebCorpus + PADIWeb_{novel})$	**92.89**	**91.23**	**92.05**

New information extraction with AI methods

[Menya *et al.*, NLDB'2024]

Work in progress

- Evaluation of the pipeline with Arab texts in Arabic language



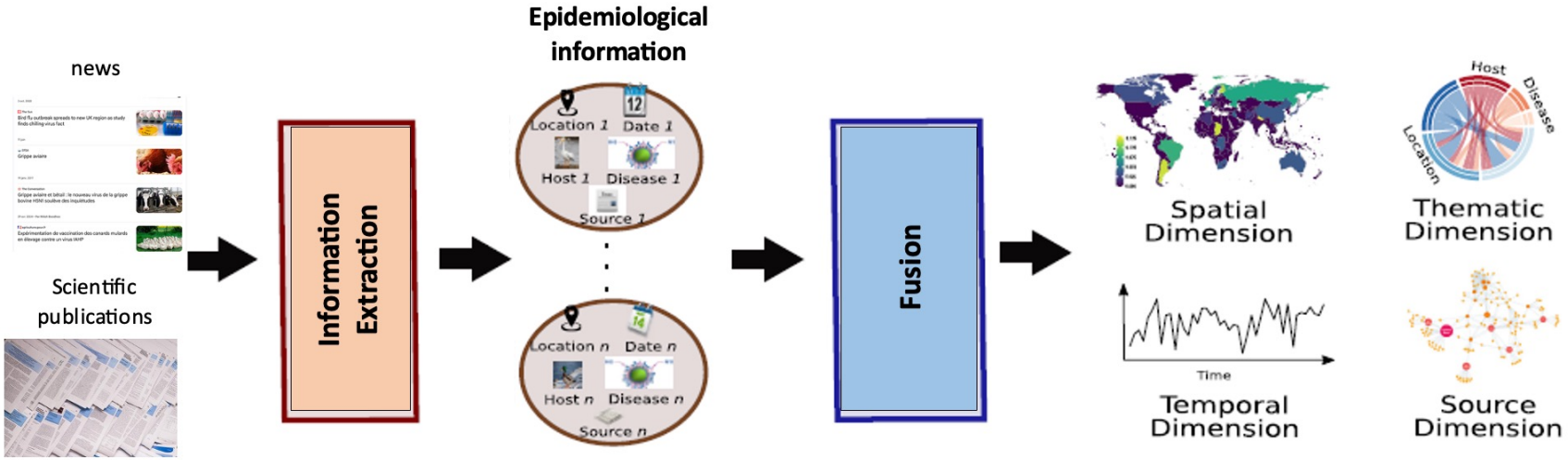
Id	Arabic location	English location	Type	Category
03P32WAHR8	أفريقيا	Africa	Continent	ASE
7CYPVEMII1	مصر	Egypt	Country	ASE
TVOB207PPH	شمال غربي الصين	Northwestern China	Geographical Region	RSE

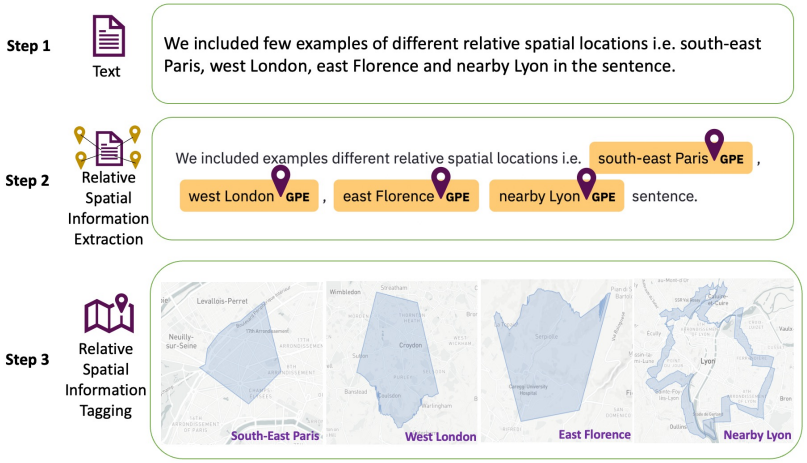
Example of lines

Fatima Ezzahra El Houbri, Najlae Idrissi (Université Sultan Moulay Slimane, Faculté des Sciences et Techniques, Morocco) and Sandra Valentin (Cirad, TETIS)

- Integration of new NER system based on AI method (**GliNER**)

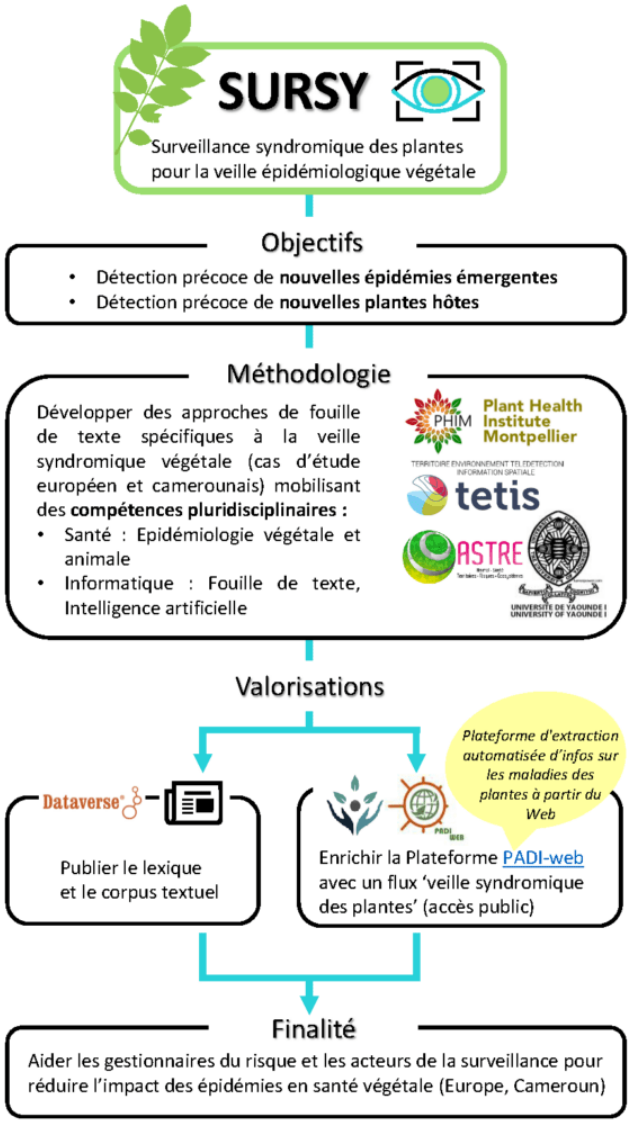
Future work





[Syed *et al.*, EACL'2024]

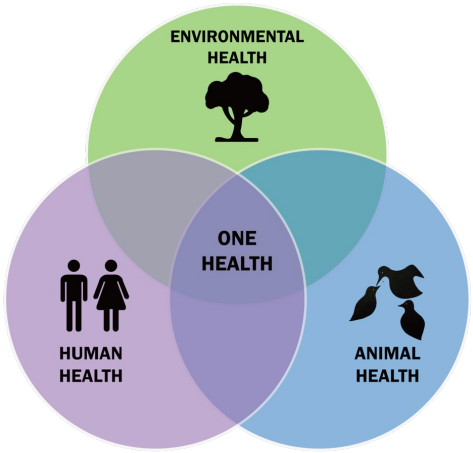
Future work



In progress:

- Constitution of a corpus of syndromic surveillance. 2 types of queries: *vocabulary of « mysterious » events & combination of « hosts » and « symptoms »*
- Automatic augmentation of this dataset
- Proposition of a general One Health pipeline for syndromic surveillance

Future work



Generecity ***Quality criteria*** ***Other media***

PADI-web documentations

Basic user

18/11/2024

Objective: **Topics**

PADI-web

Animal Health: https://padi-web.cirad.fr/

Query 1

I would like to identify articles dealing with alert and preparedness for Avian Influenza in USA between 18/06/2023 and 01/09/2023

How many texts?

a. 0 ☐
b. 8 ☐
c. 56 ☐
d. Other ☐

Basic user

18/11/2024

Objective: **Maps**

PADI-web

Animal Health: https://padi-web.cirad.fr/

Query 2

I would like to visualize countries classified as outbreak declaration of Avian Influenza in Europe between 01/12/2023 and 01/01/2024

How many texts?

a. 34 ☐
b. 76 ☐
c. 143 ☐
d. Other ☐

Basic user

18/11/2024

Objective: **Hosts / Languages**

PADI-web

Animal Health: https://padi-web.cirad.fr/

Query 3

I would like to identify how many articles mentioned Avian Influenza and bovine in March 2024?
What are the original languages of the articles dealing with Avian Influenza collected on the 14th of March 2024 with poultry in the title?

How many texts?

a. 1 ☐
b. 58 ☐
c. 146 ☐
d. Other ☐

What is the language?

a. English ☐
b. Spanish ☐
c. French ☐
d. Chinese ☐

Basic user

18/11/2024

Objective: **Meta data / content**

PADI-web

Animal Health: https://padi-web.cirad.fr/

Query 4

I would like to explore (1) meta-data and (2) content of articles collected in March 2024 dealing with West Nile virus in Greece

What are other diseases mentioned?

a. Zika ☐
b. Covid-19 ☐
c. Malaria ☐
d. Dengue ☐

What are the locations mentioned?

a. Iraklio ☐
b. Attica ☐
c. Tessaly ☐
d. Samos ☐

Basic user

18/11/2024

Objective: **Syndromic surveillance / Topics**

PADI-web

Animal Health: https://padi-web.cirad.fr/

Query 5

I would like to identify the first sentence dealing with "economic and political consequences" in the article published the 7th of February 2024 and collected with the syndromic feed.

What is the first sentence dealing with *economic and political consequences*?

a. Sentence 1 ☐
b. Sentence 4 ☐
c. Sentence 7 ☐
d. Sentence 11 ☐

Basic user

18/11/2024

Objective: **Information Retrieval**

PADI-web

Animal Health: https://padi-web.cirad.fr/

How to retrieval texts with keywords?

I would like to identify texts dealing with mortality related to Avian Influenza?

Method 1: mortality as an ***indexed*** keyword

Basic user

18/11/2024

Objective: **Statistics**

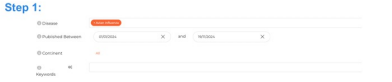
PADI-web

Animal Health: https://padi-web.cirad.fr/

How to retrieval relevant period?

I would like to identify the month where Avian Influenza is the most frequent in 2024?

Step 1:



Step 2:

Advanced user

18/11/2024

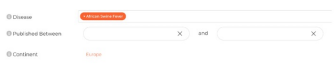
Objective: **Notifications**

PADI-web

Animal Health: https://padi-web.cirad.fr/

How to implement notifications?

Step 1: You propose a query



Step 2: You start the implementation of your notification

PADI-web team

PADI - WEB

4 PhD students
> 8 Msc students
1 post-doc

Computer scientists

Mathieu Roche [2]
Senior Research Scientist

Sarah Valentin [2]
Research Scientist

Julien Rabatel
IT development

Epidemiologists

Renaud Lancelot [1]
Senior Research Scientist

Elena Arsevska [1]
Research Scientist

Carlène Trevennec [1]
French Animal Health Epidemic Intelligence

Isabelle Pieretti [3]
French Plant Health Epidemic Intelligence

1 ASTRE Animal · Santé Territoires · Risques · Écosystèmes

2 tetis TERRITOIRE ENVIRONNEMENT TÉLÉDÉTECTION INFORMATION SPATIALE

3 PHIM

MINISTÈRE DE L'AGRICULTURE ET DE L'ALIMENTATION
Liberté Égalité Fraternité

References

PADI-web v1 (2018)

Web monitoring of emerging animal infectious diseases integrated in the French Animal Health Epidemic Intelligence System. Elena Arsevska, Sarah Valentin, Julien Rabatel, Jocelyn De Goër De Herve, Sylvain Falala, Renaud Lancelot, Mathieu Roche. 2018. *PLoS ONE* 13(8): e0199960. https://doi.org/10.1371/journal.pone.0199960

PADI-web v2 (2020)

PADI-web: A multilingual event-based surveillance system for monitoring animal infectious diseases. Sarah Valentin, Elena Arsevska, Sylvain Falala, Jocelyn de Goër De Herve, Renaud Lancelot, Alizé Mercier, Julien Rabatel, Mathieu Roche. 2020. *Computers and Electronics in Agriculture* 169: 105163. https://doi.org/10.1016/j.compag.2019.105163

PADI-web v3 (2021)

PADI-web 3.0: A new framework for extracting and disseminating fine-grained information from the news for animal disease surveillance. Sarah Valentin, Elena Arsevska, Julien Rabatel, Sylvain Falala, Alizé Mercier, Renaud Lancelot, Mathieu Roche. 2021. *One Health* 100357. https://doi.org/10.1016/j.onehlt.2021.100357

PADI-web Plant Health (2024)

PADI-web for Plant Health Surveillance. Mathieu Roche, Julien Rabatel, Carlène Trevennec, Isabelle Pieretti. 2024. *In Proc. of CAiSE'24 - 36th International Conference on Advanced Information Systems Engineering*, Springer, CAiSE Forum 2024: 148-156.

PADI-web Contact

Further questions and ideas on collaborations

padi-web@cirad.fr

https://www.padi-web-one-health.org