



### MOOD

### Monitoring Outbreak events or Disease surveillance in a data science context

Project number: 874850

Horizon 2020

SC1-BHC-13-2019

Type of action: RIA

#### Deliverable 3.4

Title: Standardised datasets (covariate, disease datasets, and pattern linking)

Due date of deliverable: 30/06/2023

Revised due date: 31/12/2023

Actual submission date: 19/12/2023

Start date of the project: January 1st, 2020

Leader of the Deliverable: AVIA-GIS

**Duration: 60 Months** 

Dissemination Level			
PU Public	Х		
PP Restricted to other programme participants (including the Commission Services)			
<b>RE</b> Restricted to a group specified by the consortium (including the Commission Services)			
<b>CO</b> Confidential, only for members of the consortium (including the Commission Services)			





Project Information			
Project Acronym	MOOD		
Project Full Title	MOnitoring Outbreak events for Disease surveillance in a data science context		
Name of the funder	European Commission – H2020		
Call Identifier	H2020-SC1-2019-Single-Stage-RTD		
Topic of the call	SC1-BHC-13-2019 Mining big data for early detection of infectious disease threats driven by climate change and other factors		
Grant Agreement Number	874850		
Project Duration (YYYY/MM - YYYY/MM)	2020/01/01 – 2024/12/31		
Project coordinator (Name of institution)	CIRAD		
Cirad Project scientific leader (name, RU, e-mail)	Elena ARSEVSKA, ASTRE, <u>elena.arsevska@cirad.fr</u>		
Project goals	<ul> <li>MOOD aims at using state of the art data mining and data analytical techniques of disease data, Big data, and contextual data originating from multiple sources to improve detection, monitoring, and assessment of emerging infectious diseases (EID) in Europe. MOOD will establish a platform for mapping and assessment of epidemiological and genetic data in combination with environmental and socio-economic covariates in an integrated inter-sectorial, interdisciplinary, One health approach. More precisely, MOOD will develop:</li> <li>1. The epidemic Intelligence community of practice to identify user needs of end-users i.e. national and international human and veterinary public health organizations;</li> <li>2. Data mining methods for collecting and combining heterogeneous Big data;</li> <li>3. A network of disease experts to define drivers of disease emergence;</li> <li>4. Data analysis methods applied to the Big data to model disease emergence and spread;</li> <li>5. Ready-to-use online platform tailored to the needs of the-end users and complimented with capacity building and network of disease experts to facilitate risk assessment of detected signals.</li> <li>MOOD outputs will be co-constructed with end-users at public health agencies to assure their routine use during and beyond project duration. They will be tested and fine-tuned on a set of air-borne, vector-borne, multiple-transmission route diseases, including anti-microbial resistance and disease X. Extensive interactions with end-users, studies into the barriers to data sharing, dissemination and training activities and monitoring of the impacts and innovations of MOOD outputs will support future sustainable use.</li> </ul>		
Key words	Infectious diseases, big data, epidemic intelligence, one health, impact, environmental changes, climate changes, user needs, socio-technical innovation		
Project partners (Name of institutions)	CIRAD, ITM, FEM, ETH, INESC ID, ERGO, SIB, INSERM, ULB, KU LEUVEN, UM, SOTON, AVIA-GIS, MUNDIALIS, OPENGEOHUB, UOXF, ISS, THL, GERDAL, IPHS, ISCIII, ANSES, INRAE, ISID		





#### **Executive Summary**

This report summarised the different normalised datasets (Covariates and diseases data) and their linking patterns. This work is based on the output of WP2, WP3, WP4 and WP5.

#### Keywords

Standardisation, data sets, covariates, pattern linking

Date	Revision	Comment	Author/Editor	Affiliation
06/12/2023	V0.1	Initial version	Sarah Houben, Maguelonne Teisseire, Tom Matheussen, Mathieu Roche, Roberto Intederdonato, Rémy Decoupe, Lucille Sautot, William Wint	Avia-GIS, INRAE, CIRAD, Ergo
14/12/2023	V0.2	Revised version	Sarah Houben, Maguelonne Teisseire, Tom Matheussen, Mathieu Roche, Roberto Intederdonato, Rémy Decoupe, Lucille Sautot, William Wint	Avia-GIS, INRAE, CIRAD, Ergo
19/12/2023	V0.3	Revised version	Julia Haas, Markus Neteler	Mundialis

#### **Document History**





### 1. Introduction

This document reports all standardised datasets and tools available on the MOOD's platform, developed or used in the context of the MOOD project.

For some, the data (covariates, disease data) or code (software packages and tools) are stored directly, while for others, only metadata and links to the actual data or code are provided through the MOOD's geonetwork (See red arrows in Figure 1).

Also in this document, the implementation of the data linking and the metadata linking patterns are described.

This report is a living document and will be updated until October 2024.



Figure 1: MOOD platform and GeoNetwork data architecture





# 2. Core MOOD Platform

### **Covariates data**

Covariates were first identified by WP2 (Milestone 16) and further expanded by WP3 during the course of the project (e.g. Milestone 29). All covariates produced for the MOOD project are summarised in the following table (Table 1).





Covariate group	Parameter	Disease	Source	Archive
Climate	Temperature	Tularaemia, Lyme, TBE, WNV, USUTU, Mosquito borne Flaviviruses	MODIS, VIIRS, Worldclim, ECMWF, (Sentinel-3)	mundialis, ERGO
	Humidity	Tularaemia, Lyme, TBE, WNV, USUTU, Mosquito borne Flaviviruses	Worldclim, ECMWF	ERGO, mundialis
	Precipitation	Tularaemia, Leptospirosis, Mosquito borne Flaviviruses	ECMWF, CMORPH	ERGO, mundialis
	Frost	Lyme, TBE, Mosquito borne Flaviviruses	ECMWF	TBA
	Snow	Lyme, TBE	MODIS, NCIS	ERGO
Matar	Divers			5000
water	Rivers	i ularaemia, Mosquito borne Flaviviruses	HYDROIK, OSM	ERGO
	Surface water	Avian Flu, Tularaemia, Leptospirosis, Lyme, TBE, WNV, USUTU, Mosquito borne Flaviviruses	VITO	ERGO
	Flooding	Leptospirosis, Mosquito borne Flaviviruses	JRC	ERGO
	MNDWI	Avian Flu, Tularaemia, Leptospirosis, Lyme, TBE, WNV, USUTU, Mosquito borne Flaviviruses	MODIS	mundialis
Topography	Altitude	Tularaemia, Lyme, TBE, WNV, USUTU, Mosquito borne Flaviviruses	GMTED	ERGO
	Slope	ТВЕ	GMTED	ERGO
		Mosquito borne Flaviviruses		
	Digital Elevation Model	Tularaemia, Lyme, TBE, WNV, USUTU, Mosquito borne Flaviviruses	Copernicus	mundialis
Infrastructure	Roads	USUTU	ESRI, OpenStreetMap	ERGO
	Accessibility	Avian Flu, Tularaemia, Lyme, TBE, Covid	JRC, MAP	ERGO
	Railways	Mosquito borne Flaviviruses	ESRI	ERGO
Demography	Population	Avian Flu, Tularaemia, Lyme, TBE, WNV, USUTU, Mosquito borne Flaviviruses, Covid	Worldpop	Worldpop
	Urbanisation	Tularaemia, slums, Lyme, WNV, USUTU, Mosquito borne Flaviviruses, Covid	Worldpop, GPW	Worldpop





	Age	Tularaemia, TBE, WNV, USUTU, Covid	Worldpop	Worldpop
	Gender	Tularaemia, Lyme, TBE	Worldpop	Worldpop
Health capacity	Veterinary schools	Health capacity, AMR	ТВА	ТВА
	Hospitals	Covid	EUROSTAT, OpenStreetMap	ERGO
Movement	Road	Lyme, Mosquito borne Flaviviruses, Covid	EUROSTAT	EEGO
	Rail	Lyme, Mosquito borne Flaviviruses, Covid	EUROSTAT	ERGO
	Air	Avian Flu, Lyme, Mosquito borne Flaviviruses, Covid	Details to be provided	Worldpop
	Overall	Lyme, TBE, Mosquito borne Flaviviruses, Covid	Google	UOXF, ERGO
	Tourism	Tularaemia, Lyme, TBE, Covid	EUROSTAT	ERGO
	Commuting	Covid	EUROSTAT	ERGO
Agriculture	Agricultural Population	Avian Flu, Leptospirosis, Lyme, TBE, AMR	FAO, GPW	ERGO
	Farm densities	Leptospirosis, TBE, AMR	EUROSTAT	ТВА
	Livestock	Avian Flu, Tularaemia, Leptospirosis, Lyme, TBE, AMR	FAO GLW	ERGO
	Cropping	TBE	SAGE	ERGO
	Trade	Avian Flu	ТВА	ТВА
	Markets	Avian Flu	ТВА	ТВА
	Wildlife Trade & Markets		ТВА	ТВА
Hosts	Bird Abundance	Avian Flu, Lyme, TBE, WNV, USUTU	ТВА	
	Birds range	Avian Flu, Tularaemia, Lyme, TBE, WNV, USUTU	Birdlife	ERGO
	Bird Migration	Avian Flu, TBE, WNV, USUTU	ТВА	
	Deer	Lyme, TBE	ERGO	ERGO
	Small Mammals	Tularaemia, Lyme, TBE	IUCN	ERGO





	Hares			
Vectors	Ticks	Tularaemia, Lyme, TBE	VECTORNET, FEM	ERGO
	Mosquitoes	Tularaemia, WNV, USUTU, Mosquito borne Flaviviruses	VECTORNET, UOXF	ERGO
	Sandflies		VECTORNET	ERGO
	Midges		VECTORNET	ERGO
Vegetation	NDVI/EVI	Lyme, TBE, WNV, USUTU	MODIS, VIIRS, PROBA-V, (Sentinel- 2)	ERGO, mundialis
	Land Use	Avian Flu, Tularaemia, swamps, Lyme, TBE, WNV, USUTU	ESA, Corine	ERGO
	Forest masting	Tularaemia, Lyme, TBE	ТВА	ТВА
Miscellaneous	Air Pollution	Seasonal Flu, Covid	Copernicus	CAMS
Proxies	Rats	Leptospirosis	ТВА	TBA
	Hunting	Tularaemia, mammals, mammals, WNV, USUTU	ТВА	TBA
	Fishing	Tularaemia	ТВА	ТВА
	Dogs	Tularaemia, Leptospirosis, TBE	ТВА	ТВА
	Lockdown measures	Tularaemia, Lyme, TBE, Covid	OWID, ECDC	CIRAD
	Outdoor Activities		ТВА	ТВА
	Global Daylength Map	Lyme, TBE	Computation of solar position and intensity from time and place (NREL SOLPOS 2.0 sun position algorithm)	mundialis

Table 1: Covariates data, associated disease and data provider





Data layers are directly stored and downloadable from the MOOD platform.

All the covariates metadata are available on the MOOD GeoNetwork (<u>https://geonetwork.mood-h2020.eu/geonetwork/srv/eng/catalog.search#/home</u>) and the links to download the data layers are also directly available on GeoNetwork. See Deliverable 3.3, point 2: "**The MOOD GeoNetwork**" for the description of the GeoNetwork populating process.

#### Download and links



Figure 2: Link to download the covariates data on the MOOD GeoNetwork.

### Disease data

#### Disease data referenced in the GeoNetwork

See Deliverable 2.2, point 2 "**MOOD disease datasets referenced into the MOOD GeoNetwork**", for the list of datasets produced within the MOOD project and already integrated into the MOOD GeoNetwork.

### Disease data integrated in the MOOD platform

Some disease data are available on the MOOD platform through softwares such as PADI-web.

The publicly available disease data generated by this software will be stored in the dedicated NoSQL database inside the MOOD platform, according to the following conceptual (See Figure 3) and logical models (See Figure 4).

#### Database conceptual model

The conceptual model was developed within the WP2 (i.e. Subtask 2.2.1 - Design of the information system for Data Management from multi-sources)).









The logical model is directly derived from the conceptual model (See Section 4 of the Deliverable 2.2). This logical model will be adapted to fit the NoSQL database of the MOOD platform.







Figure 4: Logical model for disease data

### <u>Metadata</u>

All the metadata associated to the covariates and to the disease data are stored within the MOOD GeoNetwork. For the population of the metadata in GeoNetwork, see "**Populating the MOOD's GeoNetwork**" point **2.1** of the Deliverable 3.3.

The following metadata details are available on the MOOD GeoNetwork:

- Full title
- Overview (text)
- Processing steps
- Description
- File naming
- Pixel value
- Coordinate reference system
- Spatial extent
- Spatial resolution
- Temporal resolution
- Time period
- Format
- Software used





- Lineage
- Original dataset licence
- Processor
- Contact
- Link to the dataset
- Categories of the dataset
- Region of the dataset
- Theme
- Metadata language
- Resource identifier
- Legal constraints
- Contact for the resource
- Status
- Update frequency
- Representation type
- Resolution
- Identifier
- Overview (in PNG format)
- Provider

A selected part of these metadata are also accessible into the MOOD platform directly (Figure 5: Red arrow)

- Short title
- Added on
- Updated on
- Full title
- Description
- Processing steps
- Data timespan
- External references

Inside the MOOD platform, external links will bring the user to the MOOD GeoNetwork page for more detailed metadata (Figure 5: Blue arrow) and also to the dataset provider to directly download the data layer (Figure 5: Green arrow).





Figure 5: Metadata in the MOOD platform (Screenshot dated 15.12.2023)

## 3.Tools

Many different tools were developed within the MOOD project. These tools are listed in the MOOD platform and are accessible with a complete description and a tutorial in the MOOD GeoNetwork (Figure 6). They are organised by category:

- Text mining (tools developed within WP2 and WP3)
- Normalisation (tools developed within WP3)
- Risk Mapping (tools developed within WP4)
- Other tools





	Switch Tool	➡ Tool list	
PADI-web		Text Mining	•
SNEToolkit		Normalisation	×
GeospaCy		Risk Mapping	•
EpidBioELECTRA		Other Tools	•
Spatial Opinion Minin	ng of COVID-19 Tweets		
EpidBioBERT			
GeoNLPlify			
Data quality: Classific	cation of news articles		
MOOD Press Tweets	Collector		
EpiDCA			

Figure 6: Tool list on the MOOD platform (Screenshot dated 15.12.2023)

The tools are integrated into the MOOD platform the same manner as for the metadata (see "**Populating the MOOD's GeoNetwork**" point of the Deliverable 3.3).

The code of these tools are stored on web-based software portals like GitHub and GitLab.

### 4. Pattern linking

As illustrated in Figure 7 with red arrows, the data and metadata linking will be based on three dimensions:

- Spatial
- Temporal
- Thematic







Figure 7: Linking process

### Data linking in the core of the MOOD platform

The data linking is based on 3 different dimensions:

- Spatial
- Temporal
- Thematic

according to the specifications described in the Deliverable 3.3.

The implementation of theses linking is described dimension per dimension:

### Spatial

The spatial data linking is made through the covariate part of the platform by allowing the user to visualise the integrated data directly on a map in the Core of the MOOD platform (Figure 8).







Figure 8: MOOD platform (development mode) - Covariate list (Screenshot dated 15.12.2023)

### Temporal

Temporal data linking is achieved by selecting a specific time period within a time series dataset. This allows users to visualise either a particular time point or a time lapse, facilitating a more focused analysis of the temporal aspects of the data (Figure 9).







Figure 9: MOOD platform (production mode) - Time series (Screenshot dated 15.12.2023)

### Thematic

Thematic data linking will be possible through the selection by the user of a specific disease in the disease list on the MOOD platform (Figure 10). This selection will open a menu containing 3 fields:

- Data sets
- Model outputs
- Suggested covariates

The "**Data sets**" field regroups the data set used to develop models. These publicly available data sets are listed in the MOOD platform and can be downloaded through GeoNetwork or query directly from the MOOD platform database and visualised in the Core MOOD platform.

The "**Model outputs**" field regroups the models output maps that have been developed within the MOOD project.

The "**Suggested covariates**" field will contain the list of the specific covariates used to create a specific model output. The user will be allowed to select a published model (that will be listed in the Core MOOD platform and available through GeoNetwork) and the covariates used to develop this model will appear in the left panel of the Core MOOD platform. The user will be free to select all these covariates or to modify some of them, to select the ones that are of interest for his work.





			Covariates	Diseases
→	~	Avia	n Influenza	
	→	~	Data sets	
			Gridded datasets f Human Development Ir	or Gross Domestic Product and ndex over 1990–2015 1km
			😅 CGLOPS-1 Herbace	eous wetland
			Birdlife distributior potentially important to Request to partners on	n data of species identified as o the spread of Avian influenza. On ly.
			Farm and Agricultu and FAOSTAT xlsx	ural Labour figures from Eurostat
			Annotated dataset surveillance of Avian In West-Nile Virus Disease	s from PADI-web for event-based fluenza, African Swine Fever, and e
			An annotated Avia based surveillance syst	n Influenza dataset from two event- tems
			2021_HPAI_AvianF	lostDistributionIndices
			Dissemination of in surveillance, a case stu	nformation in event-based Idy of Avian Influenza - dataset
			Surveillance tools	ents from differents digital
			Eabeled Entities fro Avian Influenza Disease	om Social Media Data Related to e
	<b>→</b>	>	Model outputs	
-	<b>→</b>	>	Suggested Covariates	

Figure 10: Thematic linking - MOOD platform (development mode) (Screenshot dated 15.12.23)

### Data linking through Epi Data Explorer (EDE)

The query/uploaded data from the core MOOD platform can be visualised directly into it, but this will also be possible in the EDE visualisation tool. Data will be sent to EDE from the Core MOOD platform in a "CSV" format.





The data linking through EDE will be based on 2 different dimensions:

- Spatial
- Temporal

The implementation of these linking is described dimension per dimension:

### Spatial

The spatial aspect of the data linking into EDE is the possibility to visualise the data queried by EDE from the core MOOD platform on two different maps.

### Temporal

The temporal aspect of the data linking is provided by the EDE tool by giving the possibility to the user to compare 2 different maps within a specific time period.

### Metadata linking

The metadata linking follows the same principles as the data linking on the three dimensions. The query is applied on the metadata and not on the data. The result is the list of the datasets that are in the GeoNetwork and not the data themself. That means that the user will have to explore in a deeper way the content of the datasets that are given as a result. To do so, the different tools are provided (standardisation tools described in the Deliverable D3.3 and provided in the software part of the GeoNetwork).

## 5. Conclusion

In conclusion, this report provides a comprehensive overview of the standardised datasets, linking patterns, and tools developed within the MOOD project for monitoring outbreak events and disease surveillance in a data science context. The document highlights the extensive work carried out in various work packages, including WP2, WP3, WP4, and WP5, leading to the co-creation of the MOOD platform.

The Core MOOD platform incorporates a wide range of covariates, disease data and associated metadata, which are essential for understanding and predicting disease outbreaks. The spatial, temporal, and thematic dimensions of data linking enable users to explore and analyse the integrated datasets effectively. The use of GeoNetwork facilitates the dissemination of metadata, making it accessible to a broader audience.

The report emphasises the importance of continuous updates, with the document planned to be a living document until October 2024. This commitment to ongoing refinement ensures that the MOOD platform remains current and relevant in the dynamic field of disease surveillance.





Furthermore, the inclusion of tools, such as those listed in the MOOD platform and accessible through GeoNetwork, enhances the platform's utility for researchers and practitioners involved in disease monitoring and analysis.

The multi-dimensional linking process enables users to navigate seamlessly through the vast dataset, providing a comprehensive understanding of the interconnected factors influencing disease dynamics.