A One-Health Platform for Antimicrobial Resistance Data Analytics

Benoit Lange, Reza Akbarinia, Florent Masseglia

A One-Health Platform for Antimicrobial Resistance Data Analytics

Benoit Lange
benoit.lange@inria.fr
Inria, University of Montpellier, CNRS, LIRMM
Montpellier, Occitanie, France

Reza Akbarinia
reza.akbarinia@inria.fr
Inria, University of Montpellier, CNRS, LIRMM
Montpellier, Occitanie, France

Florent Masseglia
florent.masseglia@inria.fr
Inria, University of Montpellier, CNRS, LIRMM
Montpellier, Occitanie, France

ABSTRACT

Antimicrobial resistance (AMR) poses potentially critical health issues for human and animal populations in the near future. To meet this challenge, we need to adopt a "One Health" strategy, which involves studying and linking information from human and animal populations, as well as from the environment.

In this demonstration, we present an early prototype of Promise platform, which we are developing for One Health data management and analytics, to enable experts from different fields to gain insights into AMR. It is designed to handle data from 25 academic networks and 42 partners. Our demonstration illustrate the capabilities of our methodology for analyzing these data. The user is freed from considerations related to data heterogeneity, as interoperability issues are managed by the platform. Additionally, each data provider will be able to stay within his/her own vocabulary, whatever the taxonomy used by other data providers.

CCS CONCEPTS

• **Applied computing** → **Health informatics**; • **Information systems** → **Data management systems**; **Web interfaces**.

KEYWORDS

one health, data management, data visualisation

ACM Reference Format:
Benoit Lange, Reza Akbarinia, and Florent Masseglia. 2024. A One-Health Platform for Antimicrobial Resistance Data Analytics. In *Proceedings of CIKM '24*. ACM, New York, NY, USA, 4 pages. https://doi.org/3627673.3679237

1 INTRODUCTION

Antibiotics have revolutionized the management and treatment of bacterial infections for humans and animals. This solution provides to healthcare professionals a powerful and efficient tool to combat the harmful microorganisms. However, the overuse and misuse of antibiotics have led to the emergence of antibiotic-resistant bacteria, which are becoming increasingly difficult to treat. This is a major public health concern, know as Antimicrobial resistance (AMR), affecting both humans and animals. To address this challenge, researchers are constantly working to develop new antibiotics, since bacteria keep evolving. Therefore, it is crucial to understand the mechanisms behind AMR.

In this context, many projects have been initiated to study AMR for a single population (human or animal). However, researchers have recently started analyzing data from multiple populations to understand the evolution and source of AMR. Various projects have been proposed in Europe and worldwide, and Promise is one of them.

Promise is a French meta-network that aims to understand the evolution and causes of AMR in France. In the context of Promise a dedicated One Health platform that aggregates data from various providers and sources is being developed. The collected datasets are primarily focused on humans, animals, and also the environment. At the moment, the data on the environment is mainly centered on surface water.

This paper presents the early prototype of a visual analytics [2, 6] platform designed to manage datasets from different providers with varying formats. By utilizing the ELT (Extract, Load, Transform) methodology, we simplify the interaction between the different datasets to make them interoperable, allowing health researchers to combine them using their own expertise.

The rest of the paper is organized as follow. Section 2 introduces the Promise data. Section 3 presents the architecture of the platform, and Section 4 provides some demonstration scenarios of the platform's usage. Section 5 presents some results obtained with the Promise data platform.

2 PROMISE DATA

The Promise platform is the result of the collaboration between 25 AMR monitoring networks and 42 academic partners with a shared goal of developing a One Health analytics system. The platform aims to bridge the gap between different health sectors that often operate independently and facilitate cross-disciplinary collaboration. It provides a tool that can aggregate data from various sources and enables project members to create dashboards for visualizing, understanding, and analyzing AMR data among different populations, including humans and animals. The platform is designed using the visual analytics paradigm.

The platform contains data from human, animal, and environmental sources. For the human population, the data comes from hospitals, long-term care facilities, laboratories, etc. For the animal population, the data is about companion and livestock animals of different species. Additionally, the platform collects open data from

CIKM '24, October 21-25, 2024, Boise, ID, USA.

ACM ISBN 979-8-4007-0436-9/24/10
https://doi.org/3627673.3679237

two environmental sources: INSEE and NAIADES. INSEE data provides information related to the population, such as the number of inhabitants per city and the age distribution of the population. NAIADES, on the other hand, contains water surface analysis done on demand when pollution has been detected. The platform is GDPR-compliant, as we exclusively store and analyze data anonymized by partners, without any possibility of linking the data to individuals.

Cross-data analysis between different populations and sources provides a more comprehensive and integrated approach for understanding AMR. By combining data from different populations, researchers can identify potential causes and patterns of AMR evolution, leading to more effective prevention and treatment strategies.

The geographic scale of the data varies depending on the provider, with some providers managing their datasets at the departmental level, while others operate at the regional or national level. Each provider has its own unique data format and thesaurus. Therefore, the data platform needs to be compatible with all these data structures and thesauri. To achieve this, the platform employs an ELT[3] (Extract, Load, Transform) methodology, where the data is stored in its original format and transformed on-the-fly based on the specific queries and requests of users. This approach enables the platform to handle a wide range of data sources and formats while providing users with flexible and customizable access to the data. The platform solves this interoperability issue by creating a software transformation layer that allows querying of all the datasets using a translator module.

3 PLATFORM ARCHITECTURE

The Promise platform is composed of multiple components (blocks) which can be scaled in order to meet the on-demand workload of the system. The platform is hosted in a Kubernetes environment with different Namespaces. Each data provider owns a dedicated Namespace to isolate data from other users. Namespaces can be created, updated, or deleted by admin users or owner of the Namespace. The platform is designed with a block architecture, consisting of four distinct blocks: Administration, Storage, Analytics, and Input (see Figure 1). Each block has a specific function. In addition to the four blocks, the platform includes a web UI, which is not central to the system, and can be considered an extension.

3.1 Input block

This block is responsible for managing secure access to resources through a REST API, which is closely linked to the administration block. All network traffic is routed through the gateway of the Input block to validate whether queries or data are permitted. When a user needs to retrieve data, a query is sent to the gateway and the security engine performs user validation. If the user is authorized to access the requested dataset, the query is forwarded to the next block.

3.2 Administration block

The Administration block is responsible for the management and security of the platform. For this specific part, the platform uses an OAuth service based on Keycloak, which is in charge of the platform's overall security. Every part of the system is monitored to check if the system is overloaded and if optimization can be applied. This strategy uses an OpenTelemetry framework to drive the Kubernetes platform.

3.3 Storage block

The storage block is responsible of managing data in the platform. Our platform can be extended to add new storage components dynamically for different data providers. Each storage module is isolated in a specific Namespace, and data is only available through the dedicated service directly connected to the database.

The storage system is designed to be flexible and can support any type of database, allowing it to be flexible and to utilize the most efficient database for each specific use case [1]. The platform can deploy a dedicated relational database or a time-series database to manage data from any provider. By using the source data structure, the platform can perform the conversion on the fly by the query engine, providing a highly adaptable and scalable storage solution.

To retrieve data, the platform incorporates a query system that translates a high level language to database queries, enabling the abstraction of user queries and database queries. The query translator module is responsible for converting the user's query into a database query, and the results are then transmitted to the web UI through this block. Moreover, the query module interacts with the normalization component to transform the datasets on the fly. This component provides the final user with the ability to manipulate queries and extracted data using their own taxonomy. This component provides a solution for simplifying the understanding and translation of codes used in data analysis. As an example, NAIDES uses codes from Sandre that represent measurement types (e.g., a drug, temperature, etc.). For example, code 6540 represents Ciprofloxacin, which is a Fluoroquinolone drug. The Normalization component manages the mapping to simplify query and user understanding. For instance, the user can get data with: 6540 or Ciprofloxacin. Additionally, the storage block creates aggregations, so when a user searches for Fluoroquinolone, different Sandre codes will be attached to the query: 6540, 6533, 6761, 6533 (Ciprofloxacin, Levofloxacin, Norfloxacin, or Ofloxacin). As a result, the data provided to the final user will be displayed using the desired taxonomy, e.g., Sandre code, Fluoroquinolone, or J01MA02 (ATC code).

3.4 Analytics block

The final block of the platform is dedicated to analyzing the relationships between the different datasets. This block is intentionally designed to be flexible, enabling project researchers to incorporate their own analytical techniques for extracting features from the data. The block is built on a Function-as-a-Service (FaaS) architecture[4], which allows the extension of analytical modules in any programming languages to detect correlations. The results of the analysis are initially stored in a temporary database. If the results are significant, they can be serialized and transferred to a permanent database.

3.5 WebUI

The Promise data analytics platform enables users to create and customize their analytical dashboards by means of a web-based user interface (webUI). The data visualization tool is composed of various components such as maps, scatter plots, and pie charts to present data in a clear and comprehensive manner, while respecting visual

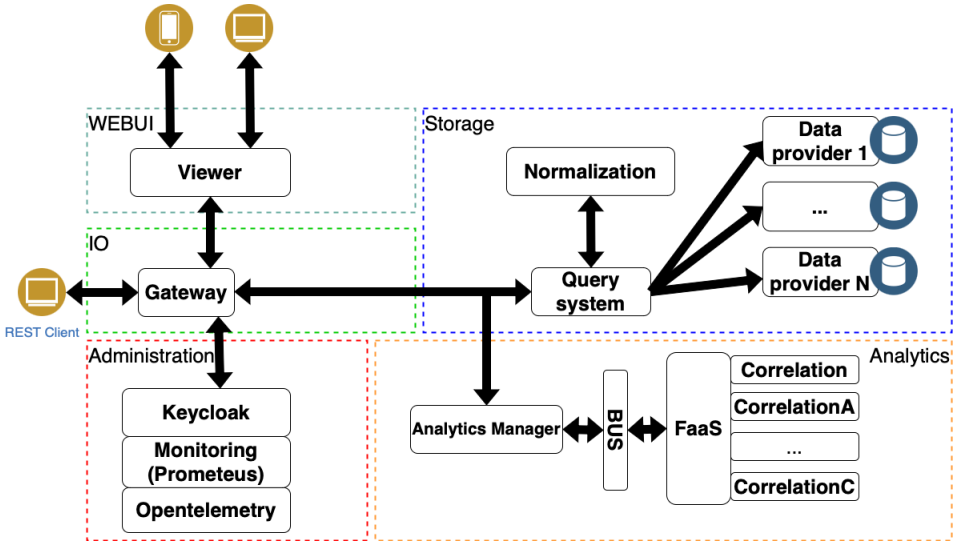

Figure 1: The architecture of Promise AMR analytics platform and its main blocks

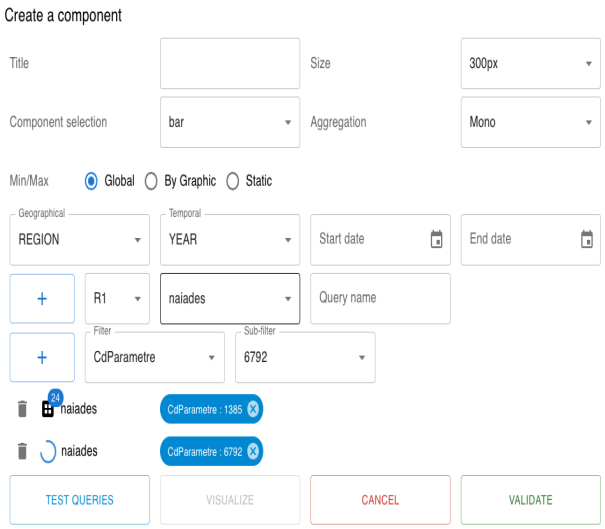

Figure 2: The user interface for building queries allows users to create and customize queries for data retrieval. It provides various options to select data sources, filter data based on specific criteria, and sort the results

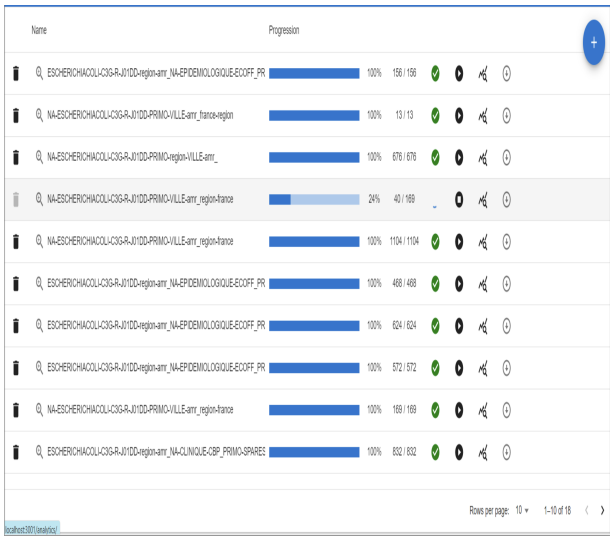

Figure 3: The Analytics manager of the Promise platform is a module that manages the execution of analytical tasks on the platform. It is responsible for scheduling, distributing, and monitoring the analytics tasks across the available computing resources.

analytics rules [5]. Additionally, users can run automatic analytics on the dataset by piloting the Function-as-a-Service (FaaS) platform.

4 DEMONSTRATION OVERVIEW

The web user interface (UI) of our platform enables the visitors to create customized dashboards by selecting and filtering data from various datasets [1]. It also allows users to create and customize queries for data retrieval, offering various options to select data sources, filter data based on specific criteria, and sort the results. The demonstration will allow visitors to run queries and visual analytics on the complete datasets gathered by the Promise project. If multiple values are obtained for a single data point, an average value is calculated, but users can also specify an aggregator method such as minimum, maximum, or median.

It is possible to identify correlations by selecting multiple datasets from different provider with the user thesaurus. Figure 2 shows the query creation interface. When the user selects a parameter, a

[1]The demonstration video can be downloaded here: https://seafile.lirmm.fr/f/15977b76bc3d4eddb2fd/?dl=1

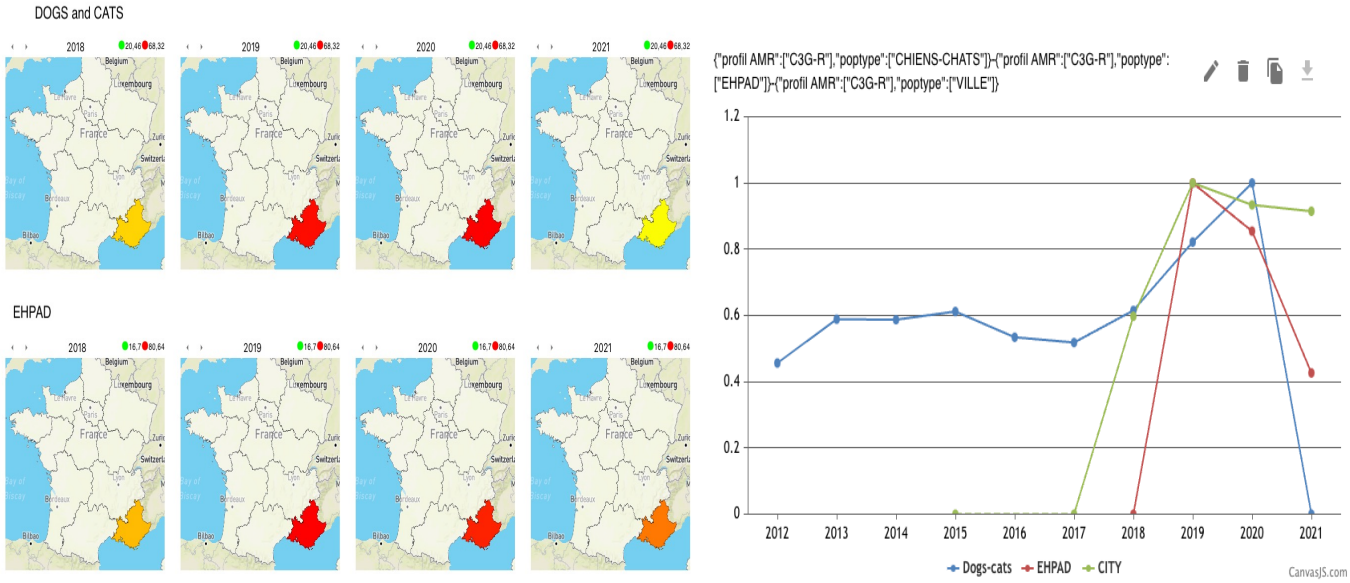

Figure 4: On the left side of the figure, we highlight the amount of AMR for dogs, cats, and senior persons in EHPADs (care homes), in the PACA region. On the right side, we plot AMR for three specific populations in the PACA region, and we observe a similar trend between EHPADs and animals.

query is automatically generated based on the selected criteria. If the user finds a discovery of interest, he/she can send the resulting values to the analytics engine to persist the knowledge.

The analytics manager is presented in Figure 3. This module uses the same query engine as the dashboard tool but the analytics engine decomposes query parameters in order to create sub queries. This step is fully asynchronous, many sub queries can be evaluated, and it can be a long-running job.

5 IS THERE A CORRELATION BETWEEN AMR IN SENIOR CITIZENS AND THEIR PETS?

The aim of this subsection is to determine if there is a link between AMR in senior citizens and their pets. By constructing and executing an analysis on the dataset, this dashboard validates the presence of this link specifically in the PACA region (south of France). Figure 4 illustrates the high correlation between both populations, with green values indicating a strong correlation. This correlation trend is the same for multiple microorganisms and antibiotics.

According to the expert in AMR involved in the Promise project, this result is the first of its kind. The PACA region has a high number of seniors having pets, this is why there is a close correlation between the AMR in senior population and animals in this region.

6 CONCLUSIONS

This paper presents the architecture of a One Health data analytics platform developed for detecting and understanding AMR. The platform aggregates data from various research groups in health management for humans and animals, enabling users from different fields to perform cross-analyses on multiple datasets. The paper demonstrates how the platform can be used to analyze a combination of datasets and generate valuable insights.

Future work on the platform will focus on integrating more data and advanced analytics to keep extracting insightful knowledge with the health researchers involved in the Promise project.

ACKNOWLEDGMENTS

This research was supported by the Promise project. We would like to express our sincere gratitude for their financial support. Promise is a French professional community network on antimicrobial resistance supported by the French Priority Research Programme on Antimicrobial resistance (PPR Antibiorésistance), coordinated by Inserm and funded by the Secrétariat Général pour l'Investissement (SGPI).

REFERENCES

[1] Zhenyu Chen, Dapeng Li, Can Cui, Yunhao Huang, Jiaqi Wang, Fangchun Di, Lixin Li, Jie Zhang, Lingxu Guo, and Xingyu Gao. 2019. A Multi-database Hybrid Storage Method for Big Data of Power Dispatching and Control. In *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation*. 502–507.

[2] Wenqiang Cui. 2019. Visual analytics: A comprehensive overview. *IEEE access* 7 (2019), 81555–81573.

[3] Edward Manopo Haryono, Indra Gunawan, Achmad Nizar Hidayanto, Untung Rahardja, et al. 2020. Comparison of the E-LT vs ETL method in data warehouse implementation: A qualitative study. In *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*. IEEE, 115–120.

[4] Dac-Nhuong Le, Souvik Pal, and Prasant Kumar Pattnaik. 2022. OpenFaaS. *Cloud computing solutions: architecture, data storage, implementation and security* (2022), 287–303.

[5] Vinay Kumar Vennu and Sai Ram Yepuru. 2022. A performance study for autoscaling big data analytics containerized applications: Scalability of Apache Spark on Kubernetes.

[6] Pak Chung Wong and Jim Thomas. 2004. Visual analytics. *IEEE Computer Graphics and Applications* 24, 05 (2004), 20–21.