01 04 2021

M. Tobin

VALUE-Dx WP3 – Task 3.4

Interoperable Networks for AMR Laboratory Data

An overarching review of Task 3.4 and federated networking

VALUE-Dx

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 820755. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and bioMérieux SA, Janssen Pharmaceutica NV, Accelerate Diagnostics S.L., Abbott, Bio-Rad Laboratories, BD Switzerland Sàrl, and The Wellcome Trust Limited.



University
of Antwerp

BIOMÉRIEUX

wellcome

UNIVERSITY OF
OXFORD

UMC Utrecht

Penta
Child Health Research

UNIVERSITÀ
di VERONA

THE UNIVERSITY
of EDINBURGH

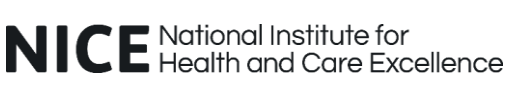
University Medical Center Groningen

RAMBAM
Health Care Campus

UNIVERSIDAD
DE LA RIOJA

BOSTON
UNIVERSITY

BIOASTER
MICROBIOLOGY TECHNOLOGY INSTITUTE

FIND

NICE National Institute for
Health and Care Excellence

Gesundheit Österreich
GmbH

ESCMID
EUROPEAN SOCIETY
OF CLINICAL MICROBIOLOGY
AND INFECTIOUS DISEASES

ERS
EUROPEAN
RESPIRATORY
SOCIETY
every breath counts

Berry Consultants
Statistical Innovation

IBBL
INTEGRATED BIOBANK OF LUXEMBOURG
FOR NEXT GENERATION HEALTHCARE

ZonMw

janssen

ACCELERATE
DIAGNOSTICS

Abbott

BIO-RAD

BD

Table of contents

1. Background to Task 3.4

1.1. Purpose

The proliferation of antimicrobial resistance (AMR) poses a serious threat to global disease control efforts. "The Review on Antimicrobial Resistance" estimates that, left unchecked, AMR will cause ~10 million deaths per year and a loss of over 100 trillion USD in economic output by 2050; most of the direct and indirect impact of AMR will fall on low and middle-income countries (LMICs).

To effectively combat AMR, it is critical to have clear visibility into the scale and epidemiology of the problem at global, regional, country and local levels, to inform policy and implementation decisions.

Building a comprehensive picture of AMR, particularly within the context of the WHO One Health approach, is a daunting task and requires data from multiple sources – across different sectors (e.g. human health, agriculture, animal health), from different service delivery points (e.g. field level, hospitals, laboratories, farms and pharmacies), and across different pathogens and sample types. There is enormous value within the wealth of AMR data collected, not just in supporting surveillance and national-level decision making but also in use cases as diverse as patient level treatment decision making (to support good stewardship), to characterisation of laboratory testing to select laboratories for participation in studies (such as Value-Dx).

Data management systems and networks, looking to mobilise AMR data, are challenged not only by the complexity of the data relationships that describe the microbiology of AMR, but also the practicalities of handling patient identifiable health information. This challenge is compounded by the need to handle a diversity of data sources, and the data sovereignty and security issues that accompany this. The challenge is daunting, yet the simple fact is that without digital tools and automation, the manual effort required to capture, consolidate and analyse the necessary data, will far outstrip the available human resources. Particularly in LMIC's such manual efforts would add burden to already resource constrained health systems.

In the face of these challenges Task 3.4 was directed to look at how digital tools, in the form of connectivity and interoperability (middleware) can be exploited to mobilise AMR data.

1.2. A Use Case for AMR Data Sharing – Lab-Select

To bring focus to the treatment of 'interoperability for AMR data' a use case, given the name Lab-Select, was conceived. In the Lab-Select use case, laboratories, conducting AMR testing, make their Antimicrobial or Drug Susceptibility testing (AST/DST) data along with patient case level data (known in this document collectively as AMR data) available alongside information characterising their laboratory's capabilities (size, equipment, standards etc…) referred to as inventory data. It is envisaged that information gathered from numerous laboratories can be aggregated, on a pan European basis. Such a network would allow a researcher to query this data in order to firstly, select laboratories to collaborate with in AMR

related clinical studies and secondly work with the individual laboratories case-level data, as part of the study activities.

In order for the Lab-Select concept to work, participating laboratories must present their data in such a way that the researcher can access and query the information according to the researchers needs. This simple requirement presents a number of non-trivial data management challenges including, but not limited to:

How to…

- Protect the sovereignty of the data owner (the health care provider)?
- Manage and control processing of the data (who, when and why)?
- Secure sensitive case level data and respect the privacy of subjects and patients?

… whilst allowing the research community to work at all levels across this extraordinarily powerful and critical data, in order to confront and resolve one of the biggest global public health challenges of our day.

Local datasets on their own can be very powerful for specific use cases; however 'analysis of global issues requires globally shared data'. The challenge for technologists is therefore:

- What architectural solutions and technologies are available to allow multi-national sharing of data whilst solving these data management challenges?

The subject area of health data exchange is vast and concerns itself with a full spectrum of considerations ranging from commercial, regulatory, operational and technical. The resources available to Task 3.4 did not allow for a comprehensive and in-depth treatment of the health data exchange landscape! In order to provide tangible benefits and understandings to the Value-Dx project, and the wider AMR community, an 'action research' approach was adopted focussed on construction of a working 'proof of concept' (POC), to propose a credible technical solution to AMR data sharing, as framed by the Lab-Select requirement.

1.3. Centralised Data Warehousing verses Federated Networks

As a result of the action research approach Task 3.4 considered two architectural approaches to 'sharing' AMR data from a network of participating laboratories, firstly a traditional 'Centralised' approach, based on data warehousing (the bringing together of copies of data, from multiple sources, into one place) and secondly a more novel 'Federated Network' approach that does not physically copy and move data, but instead distributes the question (data query) to multiple sources (Data Nodes) who share back only the 'aggregated' results. Task 3.4 chose OHDSI-OMOP (V5.3.1), as the federated network solution, to investigate in detail.

The remainder of this document provides:

- A generalised description of the data management concepts relevant to this domain
- A description of the Lab-Select use cases
- An overarching comparison of centralised data warehousing and federated network technologies
- An overview of the Proof of Concept (POC1) investigation of OHDSI-OMOP

- Conclusions around acceptability of OHDSI-OMOP and next steps

1.4. Conclusions

Our investigations showed that OHDSI-OMOP is capable (with reasonable modifications and enhancements) of supporting a pan European laboratory network, servicing laboratory inventory data and providing access to AMR case level data analytics.

1.4.1. The AMR Technology Gap

Our work has highlighted a well observed fact that the complex relationships that characterise microbiology and AST/DST testing, is often neglected or poorly handled by many health systems(IT), resulting in a technology gap with respect to IT systems support for AMR data.

In-line with this trend it was observed that the OHDSI-OMOP's data model does NOT natively (out-of-the-box) support AMR data without modification. We were able to show a number of OHDSI-OMOP customisation options that could be used to bridge this gap: although we have shown our chosen modifications work, they are sub-optimal and unlikely to scale. Either a more elegant customisation is required or better still AMR support should be built into OHDSI-OMOP.

It is clear from early efforts to influence OHDSI-OMOP into adopting support for AMR data, that strong advocacy is required to justify the effort and added complexity that this requirement brings.

The threat posed by AMR, if unchecked, will come to dominate the future health agenda, challenging public health in a profound way. Mobilising health data to drive decision making is an indispensable element at the heart of containing this emerging threat: closing the AMR technology gap (through awareness and technical education, creation and access to tools, code libraries, and automation) is therefore essential, and needs to be established as a global initiative. It is hoped that task3.4 and this conclusion, can be used as a springboard to help establish such an initiative.

1.5. Related Documents

This document provides an overarching summary of the details to be found in the following artifacts delivered by Task 3.4 in Jan 2021.

- Report #1 "VALUE-DX-WP3-4-POC1-Modeling_Feasibility": A detailed description of the application of federated networking to AMR testing data, using the OMOP tool set.
- Report #2 "POC1 Documentation"
- POC1 - Federated Network Proof of Concept (POC1): A working example of the OMOP tool set configured to illustrate the Lab-select process using data from bioMérieux's MYLA® data management platform and test data from WHONET a client software from WHO Collaborating Centre for Surveillance of Antimicrobial Resistance
- Lab-Select Project design requirements and specification VALUE-Dx Lab Select PDR and PDS v1.0.XLSX

These artefacts are stored and referenced in the Value-Dx SharePoint project site.

Acknowledgements

Acknowledgements are given to the following consortium members:

Jean-François GORSE - bioMérieux

Xavier GANSEL – bioMérieux

Emmanuelle Theron – bioMérieux

Frank Leus - University Medical Center Utrecht

Marc Padros Goossens – University Medical Center Utrecht

Tuba Vilken - University of Antwerp

Mauro Tobin - FIND

2. Principles of Data Sharing

This section considers the data management components required to support AMR data management by defining some of the key entities involved in moving data and highlighting the common architectural approaches to show how these entities typically combine in health systems(IT).

Data Definitions

- Data management – The storage, management (i.e. creation, update, removal, back-up, quality control …) access control, usage and sharing, of data
- AMR Data – In the context of this document AMR data is defined as the array of Antimicrobial or Drug Sensitivity (AST/DST) testing results, (quantitative and qualitative) generated by the relationship of Patient and sample data, with a pathogen and many antibiotics
- Case-level - AMR Data at the individual patient level (Patient x tested for drug resistance)
- Aggregate Level AMR Data – Data calculated using case-level data to provide an indicator (for example % drug resistance detected) and therefore removing case level identifiers such as patient name or age

2.1. Data Sharing – Components

The terms connectivity interoperability and middleware are often used together and occasionally interchangeably. Task 3.4 defines these terms as follows:

> **Connectivity**: The ability for physical objects, particularly diagnostic instruments or mobile devices (i.e. phone-based applications or a network of desktops) to electronically connect to neighbouring devices or centralised systems. Example AMR testing instrument communicating over a Local Areas Network (LAN) with a Laboratory Information Management System LIMS. Another example might be a fleet of 'Point of Care' diagnostic instruments, in health centre settings, using mobile internet to communicate with a centralised management system.
>
> **Interoperability**: The ability for centralised systems to communicate electronically with each other using standard automated methods. An example would be an LIMS providing laboratory results to a hospital Electronic Health Record EHR system. Interoperability concerns itself with more than just the exchange of data; a shared vocabulary, workflow and ultimately the 'meaning' conveyed in the data, must be unified and understood, if the full interoperability of systems is to be achieved.
>
> **Middleware**: A software application that Extracts, Transforms and Loads (ETL) data from one electronic system to another (data source to data endpoint), in order to establish interoperability; also referred to as ETL tools or ETL systems. Middleware solutions provide a powerful and flexible way of connecting multiple systems together. They can create a single standardised data output from a diversity of data sources,

equally they can take a single input source and split it into numerous differently formatted outputs containing a different selection of data elements.

Middleware Orchestration: Orchestration is a function of middleware which supports the management of workflow in handling data and interoperability, allowing decisions to be made about what to do with incoming data according to conditions. It allows delivery of data to be sequenced across different systems as well as allowing for the creation of new data packages based on combining data fetched from multiple data sources.

Application Programming Interface API: An API is a standardised method for communication between electronic systems which allows an application to either 'Get' data (request data packages) or 'Put' data (add or update records in another system). There are a number of standards in this area (e.g. Web Services, Restful API) which are well understood and common place. An API will support one or more 'authentication' methods that ensures only systems that have been granted access to an API can issue API calls (i.e. Get or Put data).

These building blocks allow a diversity of configurations and operational models to exist. The combination of these capabilities allows interoperability to support a wide spectrum of use cases and data journeys.

The following examples use diagnostic connectivity to illustrate typical ways in which data flows from data source to data end points using these entities.

2.1.1. Connectivity Example – Connected POC Diagnostics

Point of Care (POC) diagnostics form a distributed decentralised network of testing capability. Electronically collecting testing, patient and machine health data from these devices is essential in efficiently managing the network, amplifying the benefits of rapid diagnostics and improving health outcomes.

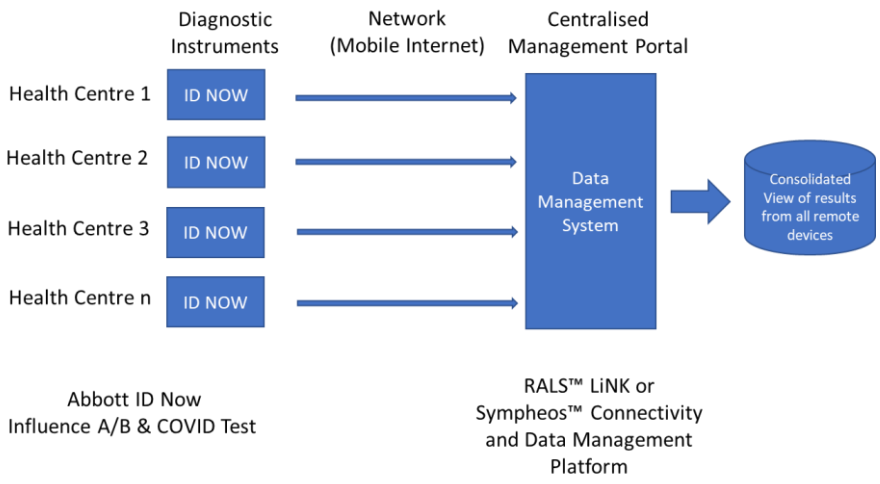


Fig1: Example of connectivity – Abbot ID NOW™

The Abbott ID NOW™ point of care molecular diagnostic device can connect via mobile internet (using an external modem 'CONNECT Universal Gateway'), in order to transmit result data. Abbott's RALS LiNK (or alternatively Abbott's Sympheos™) data management backend systems can collect, store and report on the connected data. This is an example of how a fleet of remote diagnostic devices can use 'connectivity' for case-level reporting, surveillance and inventory management. Note that these platforms can also share data onwards with other health systems(IT).

2.1.2. Middleware Example - Laboratory Diagnostics

In the laboratory setting, middleware (also known as Domain Middleware) is often used to communicate and aggregate data coming from a manufacturer's portfolio of devices, in order to provide a single source of data to participating systems(IT), typically the lab's Laboratory Information Management System (LIMS). Detailed results and instrument level data are brought together providing instrument controls and specialised reporting of results and machine activity, whilst presenting a consolidated and standardised view to the LIMS.

An example of this is the MYLA® middleware from bioMérieux. This software uses connectivity to communicate with a range of bioMérieux devices providing a centralised management portal. This can be described as a many to one integration. MYLA can itself use standard interface connections to interoperate with a LIMS.

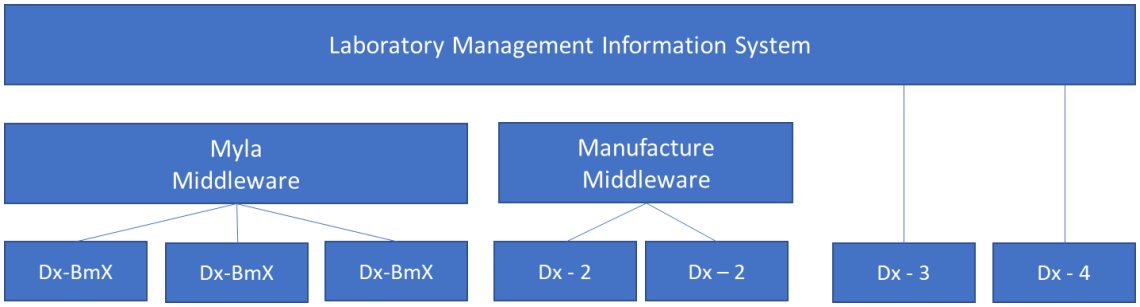


Fig2: Example of interoperability in the laboratory

Note: A laboratory may contain other manufacture's devices, which may either use that manufacture's middleware, or connect directly to the LIMS.

2.1.3. Example of Middleware - System(IT) Interoperability

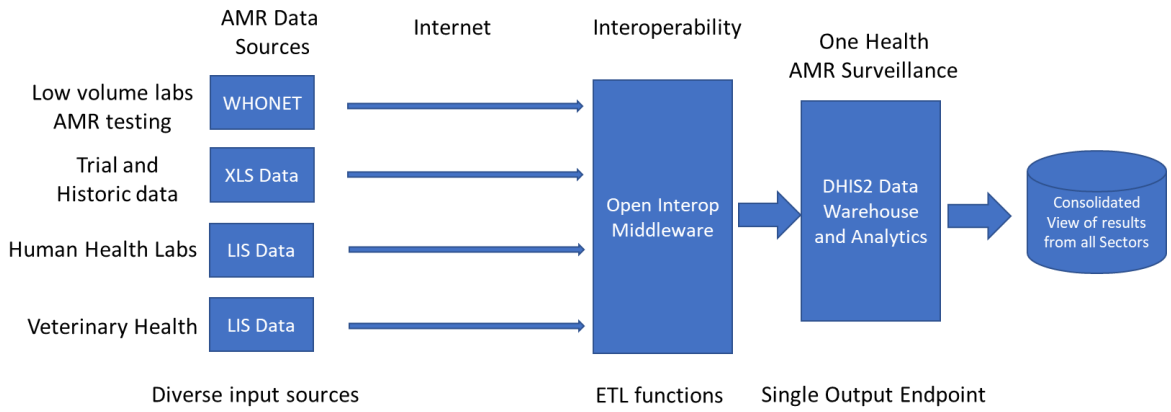


Fig3: Example of interoperability in One Health Surveillance

This example, taken from the Zambian National One Health AMR surveillance system, provides an illustration of middleware (Open Interop) being used to collect data from multiple input sources, in this case Microsoft Excel, WHONET (microbiology laboratory database software), OpenLDR (a data warehouse containing human AMR testing), and SILAB (LIS used in veterinarian labs). Collected data is sent to DHIS2 (a widely adopted Open Source project) which acts as a data warehouse and analytics tool.

The role of middleware in this architecture is to extract and transform the AMR data from the various data input sources and update the One Health data warehouse via a single standardised data DHIS2 interface.

The benefit of middleware in this case, is to allow any number of additional data sources to be added to the configuration, without the need to change the One Health system itself. Equally the numerous data sources are protected from the need to implement changes whenever a change in the endpoint (DHIS2) is made. Additional systems can be added to the architecture with minimum disruption including the retirement and/or replacement of key components.

3. The Lab-Select Use Case

Lab-Select is a visionary service defined as part of Task 3.4, and describes a pan European service allowing 'research organisations' to identify and select laboratories to collaborate with in AMR studies. The Lab-Select concept serves two purposes, firstly it provides a use case for the Task 3.4 investigation into data sharing architectures, and secondly it provides information and thinking to input into ECRIN/ECRAID (initiatives to establish a sustainable Clinical Trial Network).

To fully describe the Lab-Select proposition and resulting requirements, processes and commercial considerations, is a vast undertaking. For the purposes of this document Task 3.4 have focussed on the following characteristics and has identified a number of so called 'epic' requirements, that frame the proposition. These epic characteristics are listed in the sections below.

An initial set of use cases are described in the product design specification and requirements document on the Value-Dx project site (VALUE-Dx Lab Select PDR and PDS v1.0.XLSX)

3.1. Lab-Select Characteristics

This section lists the major epic Lab-Select characteristics. Many of these characteristics mirror those expected to be a part of ECRAID. In order to manage the scope of our activities, Task 3.4 has not attempted to perfectly align with the emerging ECRAID requirements. However, since the fundamental requirements are the same there is strong alignment between the two.

3.1.1. Target Setting

Lab-Select is a pan-European service allowing individual laboratories conducting AMR testing to enrol as part of the network and European researchers and study groups, to apply for access.

3.1.2. Target Users

Lab-Select is targeted primarily at research organisations such as pharmaceutical companies who are planning to conduct drug discover investigations or clinical trials for new drug treatments. It can also be used by other academic research institutions or research programmes such as Value-Dx. These stakeholder groups mirror those expected to be a part of ECRAID

This simple epic requirement demands that a number of fundamental entities and components be in place, these items are as follows:

1 – A centralised provider for delivery of the service, known here as the Lab-select Observatory

2 – A network of participating laboratories – known as the Lab-Select network

3 – Individual laboratories generating fully characterised AMR case level data – known later in this document as Data Nodes

4 – Supporting 'inventory' data describing the laboratory itself

3.1.3. Service Proposition

The Lab-Select service proposition is to provide the researcher with the ability to;

- Select laboratories from the network
- Contact the laboratories and agree collaborations
- Gain access to the collective case-level data from the collaborating laboratories
- Provide the ability to conduct research activities (data analytics) on real world data

This service proposition is underpinned by a number of key user journeys constructed from the following additional features:

- Inventory search – Laboratory search and select, based on matching items in the inventory e.g. location, Quality Assurance standards used, available diagnostics
- Surveillance and testing profile search – search and select based on trending and profiles data i.e. prevalence of pathogens regularly tested for, samples handled, antibiogram. This is akin to maintaining surveillance type data for each lab
- Bespoke analytics – bespoke, researcher defined case-level analysis and data mining, across all selected laboratories, in order to either further refine selection or as part of conducting real world data research. The selection and any subsequent research analysis, can span years and by its nature be multi-site (i.e. real-world comparison of results with trial results, drug safety analysis)

3.1.4. Laboratory Network Management

In order to effectively run the service the following 'epic' requirements must also be in place:

Data capture: Data must be continuously gathered from the individual laboratories resulting in the following data management features:

- Inventory data – is initially collected when a laboratory enrols in the Lab-Select network and is maintained manually by the laboratory, on an annual basis
- The case-level data - is collected in near real-time from each participating laboratory
- The surveillance / testing profile data – aggregated (that characterises each lab) which can be updated on a monthly basis, using routine analytics, run against the data owners case level data
- Bespoke queries – queries commissioned by the researcher which can be run on demand

Access Control: Eligibility rules must exist to allow Labs to restrict how their data is used in research, and what types of collaboration are acceptable

Security: The data must be secure and patient privacy respected

Regulatory compliance: The system must comply with both local, nation and European regulatory requirements and guidelines, and must respect data sovereignty laws

Sustainable operations: – The service must be able to scale and be maintainable

3.1.5. Establishing a Collaborations in Lab-Select

It is envisaged that establishing a collaboration and data sharing agreement will be a manual offline process involving negotiation and development of an understanding between the participating organisations. At this point in time it seems unlikely that this manual step can reasonably be avoided, however it is possible to see how increasing levels of automation can help to make this process as quick and painless as possible.

3.2. Benefits of Lab-Select

It is believed that vast efficiencies can be gained from making laboratory inventory and surveillance data available, as part of identifying partners and establishing data sharing agreements for AMR related studies. Additionally a Lab-Select process can reduce the overhead of conducting bespoke data analysis and ensure that such activities are focussed and more effectively executed. It should also be noted that the case-level data required to support a meaningful implementation of Lab-Select would represent an enormously valuable pan-European wide asset, that could have numerous and potentially unforeseen benefits, especially when merged with non-laboratory data sources (e.g. One Health Surveillance)

Thus the fundamental benefits of Lab-Select can be summarised as the ability to:

- Speed up identification of appropriate partners
- Speed up initiation of suitable collaborations
- Allow multi-site, multi-year real world data to compliment trial data
- Bring efficiencies and accuracy to collection and analysis of existing data
- Bring efficiencies to the collection and usage of key AMR data
- Provide a platform for the re-use and wider exploitation of this critical data

In short – '*To better mobilise laboratory diagnostics to fight AMR*'

3.3. Lab-Select Data and User Journey

The Lab-Select data journey describes how data and is managed and flows through the Lab-Select system. This is envisaged to happen at three levels. Firstly inventory data is entered and maintained by the laboratory. Secondly routine data collection scripts assemble aggregate testing profile data, and finally case level data is made available for bespoke analytics.

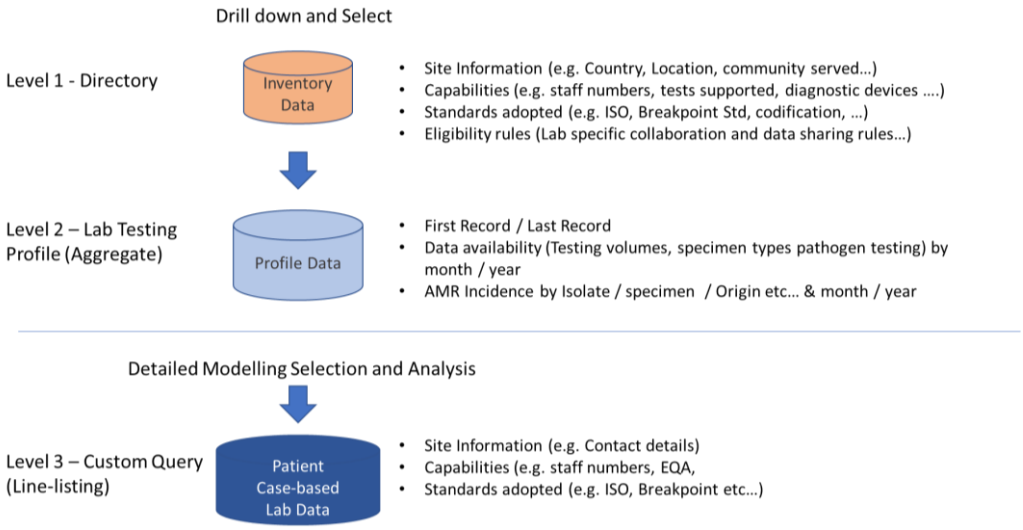


Fig4: Three level data selection process

A researcher using the system to select laboratories, is an example of a 'user journey'. The lab selection user journey represents the process of a researcher viewing the inventory, viewing the structured profile data, and then optionally issuing bespoke queries, in order to make choices. The metaphor for this selection process is one of progressively drilling down and adding labs to a shopping basket. As with filling an eCommerce shopping basket with shopping items, the decision-making process can be varied (comparing, adding and removing items in an unprescribed way) and therefore Lab-Select must allow the user journey to be fully flexible, allowing the user to add, remove, query, re-query and adjust, their selections.

3.3.1. Example User Journey

Below is an example of a researcher selecting laboratories based on the presence of drug resistant K. pneumoniae in tested blood samples.

Step 1 – First search for appropriate laboratories, using inventory data, and add to basket.

Example Inventory data below

- Laboratory name, location, hospitals or community served
- Laboratory size and testing capacity

10 employees or less	*Yes / No*
11 to 25 employees	*Yes / No*
26 to 50 employees	*Yes / No*
More than 50 employees	*Yes / No*

- Full-time equivalents (FTE)* occupying the following positions?

	Number
Clinical/medical microbiologist, clinical/medical biologist	
Biologist (MSc)	
Laboratory technician	
Others	

- Automated systems used to identify microorganism?

VITEK (bioMérieux)
Phoenix (Becton Dickinson)
MALDI-TOF Mass Spectrometry (Brucker Daltonics)
BD BBL™ Crystal™ identification system
MicroScan WalkAway plus System (Beckman Coulter)
Others (specify instrument, manufacture, version)

- Universal Standards used to identify medical laboratory observations and organisms?

LOINC (**L**ogical **O**bservation **I**dentifiers **N**ames and **C**ode**s**)
SNOMED-CT (**S**ystematized **No**menclature of **Med**icine – **C**linical **T**erms)
ICD-9 (International Classification of Diseases)
ICD-10
ICD-11
Others (please specify name and version)

- Breakpoints used to interpret antimicrobial susceptibility testing results?

EUCAST	The European Committee on Antimicrobial Susceptibility Testing (Europe)
CLSI	Clinical Laboratory Standards Institute (USA)
BSAC	British Society for Antimicrobial Chemotherapy (UK)
CA-SFM	Comité de l'Antibiogramme de la Société Francaise de Microbiologie (France)
DIN	Deutches Institute für Normung (Germany)
CRG	Commissie Richtlijnen Gevoeligheidsbepalingen (The Netherlands)
NWGA	Norwegian Working Group on Antibiotics (Norway)
SRGA	Swedish Reference Group of Antibiotics (Sweden)
	According to the manufacturer of the disks or commercial MIC tests
	Other

Step 2 – Use profile data to select from laboratories in the basket, by looking for test results for K. pneumoniae, reject those that don't.

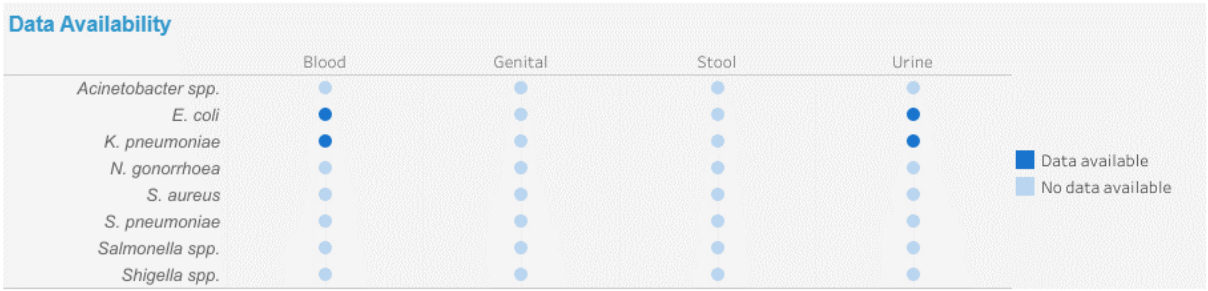


Fig5: Example profile view for a laboratory

Step 3 – Run bespoke query against remaining selected labs, to highlight best candidates

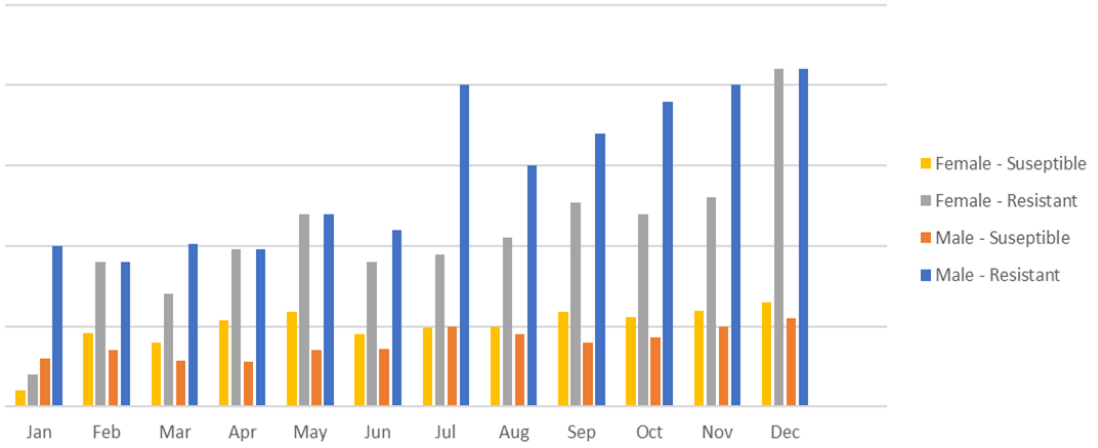


Fig6: Example results for a bespoke query to show detailed stratified case data

Graph to show trend over time, by month, of incidences of K. pneumoniae, resistance and susceptibility, in blood samples, from tested individuals within an age range of 23<34, by gender, for individual labs or group of selected labs that form part of the study.

4. Candidate Solutions

The challenge faced by Task 3.4 was to find potential technical solutions to the Lab-Select service proposition and epic requirements, and to prove its feasibility in the form of a 'proof of concept' POC. The investigation started at the highest level by considering two opposing data aggregation architectures for data management:

1 – A Centralised Data Warehouse Approach

2 – A Federated Network Approach

In the centralised model data is copied from the participating data providers and brought together in a single place where the Observatory is free to process it. In the federated model data queries are sent from the Observatory to the participating data providers who return the answer which is then aggregated by the Observatory.

The sections below describe both these approaches. Section 4.2 'Federated Networking – The Promises' considers the reason why the Federated approach was selected and lists the 'Promises' that federated architectures offer.

4.1. Data Aggregation Models

4.1.1. The Centralized Data Warehouse Approach

The centralise data warehouse approach is a common and widespread data management solution, in use across all industries, to collect store and analyse data. In this model data is copied from participating data providers (also known as Data Nodes).

A typical data warehouse architecture takes a layered approach to managing data, as follows:

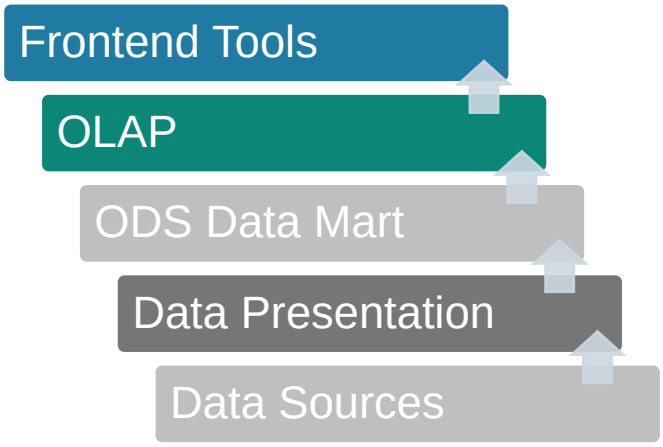


Fig7: Layered data warehouse model

Data Presentation Layer: Structured line listing (or case-level) data is collected from participating data sources or Data Nodes, and copied into the data presentation layer, where the data is stored. This detailed data can then be cleansed, transformed and mapped (against a common set of reference data using ETL functions), as part of being promoted into the next

layer, which is the Operational Data Stores (ODS) or Data Mart layer. This represents the raw input data.

Operational Data Stores (ODS): The ODS consists of the cleansed and mapped data, which can be either directly queried, or further processed and typically aggregated into the OLAP layer, according to the analytical needs and requirements of the data warehouse users.

Data Marts are 'functional' sub-sets of data that have been established for a specific domain purposes within this layer.

OLAP: OLAP (Online Analytical Processing) databases store the aggregated view of the data using multi-dimensional schemas (sometimes called an OLAP cube) that allows users to drill up, drill-down and slice-and-dice the data. The OLAP data is normally generated on a scheduled basis (hourly, daily etc.) from the ODS using a set of transformations that aggregate the data according to the 'needs and requirements of the users and the domain they are working in.

Frontend Tools: Finally the front-end tools layer provides the user with a set of analytics and reporting tools to allow the data to be queried, analysed and visualised in dashboards graphs and maps.

4.1.2. Centralised Architecture for Lab-Select

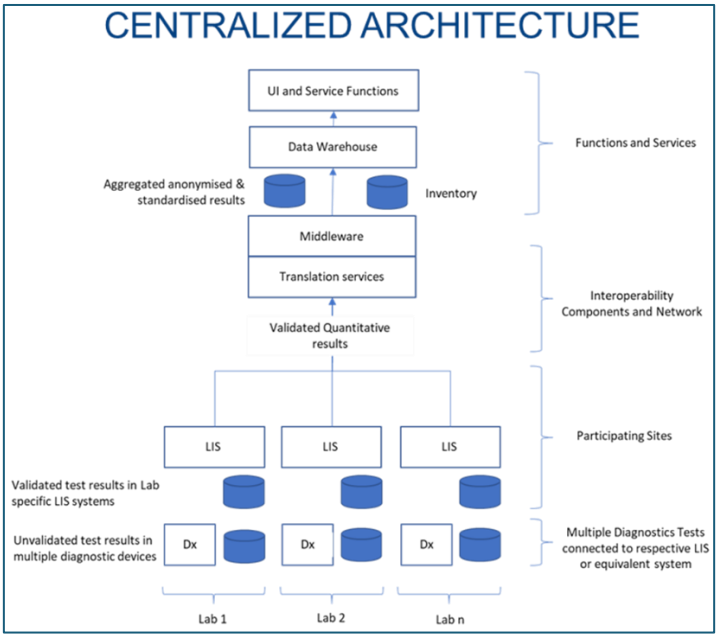


Fig8: Centralised data warehouse model

The diagram above shows one of many possible implementations of a Centralised approach for Lab-Select. This particular option starts the data journey with diagnostic devices in the laboratories feeding data to their respective LIS's, and each LIS sharing its data with a centralised data warehouse. ETL functions are provided by middleware and translation services in order to create the required OLAP datasets in the data warehouse.

To use such a data warehouse approach, to support the Lab-Select use case, requires creating a local copy of the case-level data, from all the Labs in the Network, by transmitting the data into a data presentation layer across a network. Once the data arrives at the data warehouse individual datasets are 'transformed' into a standard set of Data Marts and made available to the Observatory to allow researchers to analysis and select labs.

This approach presents a considerable number of practical issues including (but not limited to):

- Scale – As a result of taking a copy of each laboratories data, the Observatory will be required to support a very large and ever-growing volume of raw data which is further duplicated to create operational datastores
- Security – Highly secure network connections and data storage facilities are required in order to secure this sensitive data
- Data Sovereignty and Patient Privacy – Sensitive patient data will be centralised outside the data sovereignty of the nation states in which the laboratories are located, presenting considerable issues regarding regulations and attitudes towards data sovereignty. A large burden comes from the need comply and stay up to date with local regulations to protect patient privacy for each laboratory in the network.
- Data Standardisation - There is also considerable heavy lifting in terms of ETL as unification of the vocabulary is required to establish a single standardise dataset for Lab-Select users to work with. In this model the burden is with the Observatory which also therefore includes a change management overhead in order to ensure that any changes in the source data is reflected in the ETL and translation services.

4.1.3. The Federated Network Approach

In the federated approach the principle is to avoid the need to copy and transmit the source data, by establishing standardised datasets, at the local Data Node level; analytical queries are distributed over the network, to the participating Data Nodes, which return their individual aggregated results. The returned results can then be used to either directly answer the analytical query, or to build OLAP datasets for subsequent aggregated analysis. In this way it is the question and answer which is shared; the raw data remains in place and is never actually shared.

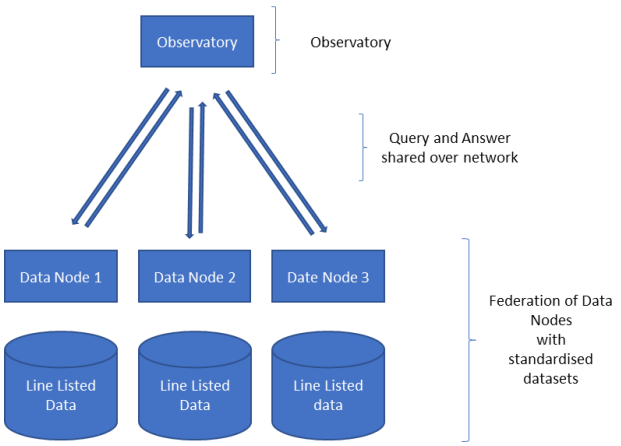


Fig9: Federated network model

In order for this approach to work Data Nodes need to present their data in a consistent, unified and standardised format, such that each Data Nodes structure is almost identical. This allows a query constructed in the Observatory to be executed in an identical way against each Data Node's data. This provides a coherent set of responses which allows the Observatory to combine the responses to create a meaningful and accurate answer to the original query.

4.1.4. The Federated Network Approach for Lab-Select

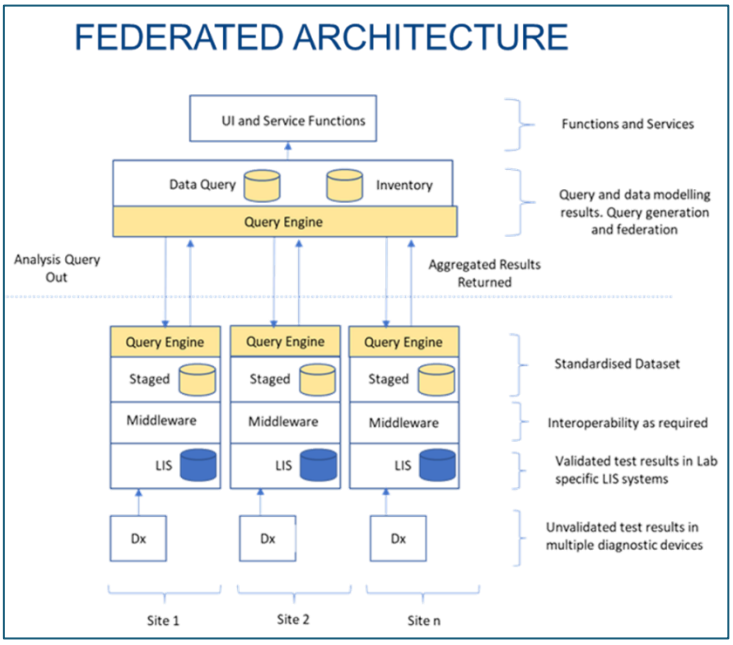


Fig10: Federated network model for Lab-Select

The diagram above shows a possible implementation of Federated networks for Lab-Select. In this option laboratory diagnostics are connected to their respective LIS systems. Middleware is used to perform ETL functions in order to stage the AMR testing data in a standardised format managed and hosted by the local laboratory system(IT). A query engine, at the Observatory communicates with corresponding query engines installed and hosted at each

laboratory site. Queries and answers are passed across the network between the query engines.

This approach promises to solve a number of issues present in the Data Warehouse model whilst presenting a number of unique practical issues as summarised below:

- Scale – By decentralising the storage of detailed data, the burden of scale is distributed across the network. Also the amount of duplication of data is reduced.

- Security – Sensitive data is maintained within the secure network of the Data Node. Secure network access remains a priority, however no sensitive data is transmitted across the network reducing the security risk.
- Data Sovereignty and Patient Privacy – Sensitive patient data remains with the domain of each individual laboratory and associated systems(IT) and is therefore, theoretically under full control and jurisdiction of the laboratory and hospital system(IT). The local hospital system maintains sovereignty over the patient level data and can withhold or withdraw access at any time.

- Ownership of aggregated results passed to the Observatory is the subject of data sharing agreements between the Observatory and Data Nodes which are expected to focus on anonymised aggregate data, rather than patient identifiable case level data. This promises to reduce the burden on complying and staying up-to-date with local regulations to protect patient privacy for each laboratory in the network.

- Data Standardisation - There remains considerable heavy lifting in ETL to support unification of the vocabulary and establishing a single standardise dataset for Lab-Select users to work with. The burden is distributed amongst the laboratories, who are accountable and responsible for creating the standard dataset. Although the overall ETL burden (compared to the centralised approach) is not reduced, it is focussed mostly at the Data Node level.

4.2. Federated Networking – The Promises

Task 3.4 chose to focus on investigating Federated Networks as a solution architecture for Lab-Select in order to investigate the novel potential of this new approach, with a particular focus on the following promises offered by the technology:

Ability to:

- Handle the complexities of AMR diagnostic data
- Accommodate diverse source datasets
- Avoid data sovereignty issues
- Secure sensitive data and avoid un-authorised sharing
- Federate complex custom queries across the network
- Federate queries automatically
- Allow creation of a secure and usable service wrap to support Lab-Select

5. Federated Networking using the OHDSI toolkit

The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") program is a multi-stakeholder, interdisciplinary collaboration focussed on developing open-source solutions to deliver value from health data, through large-scale analytics. The initiative also goes by the name OHDSI-OMOP: Observational Health Data Sciences and Informatics – Observational Medical Outcomes Partnership. OHDSI and OMOP are often used interchangeably when describing tools and facets of the OHDSI-OMOP world.

The OHDSI-OMOP standards and tools are supported and maintained by a dedicated community.

5.1. OHDSI-OMOP Common Data Model

OHDSI-OMOP have defined the Common Data Model (CDM) for storage of observational health data. The model attempts to harmonise data from a number of different observational domains relative to patient health, into a single data model to support analytics. The model is described as 'Patient Centric' in that all clinical events recorded in the model are linked to a PERSON table which describes the person. The model allows a longitudinal record containing multiple clinical events to be captured.

The graphic below provides a view of the tables that make up the CDM.

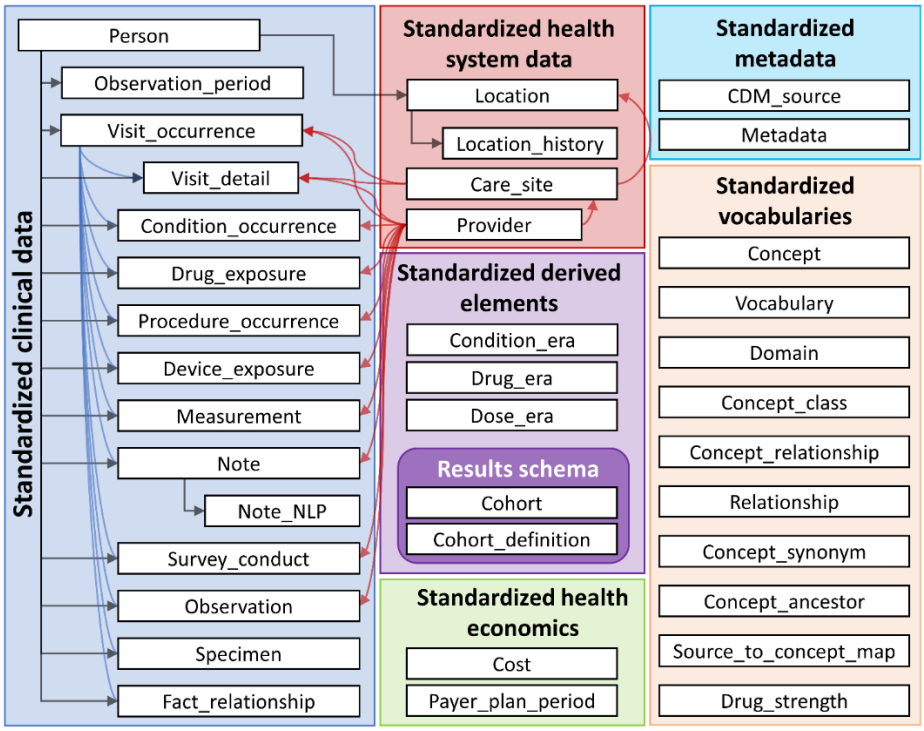


Fig11: Overview of all tables in the CDM version 6.0. Note that not all relationships between tables are shown.

5.2. OMOP-CDM Does not support AMR data without modification!

The OMOP-CDM data model does NOT support the complex data and entity relationships that define the microbiology of antibiotic and drug susceptibility testing, and therefore does NOT natively support AMR.

CDM does however provide methods for extending the data model to meet particular user needs. Task 3.4 has investigated and modelled different ways of accommodating AMR data into the CDM and has broadly shown that AMR data can be incorporated into OMOP-CDM with constraints. Work is ongoing with the OHDSI community to explore how best to include AMR in the model as a native feature of the CDM.

A full discussion of this investigation and how to achieve support for AMR testing data in OMOP-CDM is presented in the Task 3.4 document Report #1 "VALUE-DX-WP3-4-POC1-Modeling_Feasibility".

5.3. OHDSI Toolkit

OHDSI provides a suite of Open Source and paid for tools. For the purposes of this document the tool kit is divided into two categories:

Implementation toolkit: In order to host data in the OMOP-CDM source data must be gathered and translated into OMOP-CDM including a complete mapping of "local" vocabularies" to standardized vocabularies. OHDSI therefore make implementation tools to build the models, prepare the data vocabulary mappings, and migrate data.

Federated Network toolkit: The network toolkit includes components that allow a network of OMOP-CDM Data Nodes to federate with an Observatory. The principle tools for this function are:

- **Arachne** – Components to provide workflow orchestration across federated network. 'Arachne central' provides the 'Observatory' with the ability to establish the network, profile Data Nodes and issue queries and collect results.

- **ATLAS** – A web-based integrated platform for database exploration, standardized vocabulary browsing, cohort definition, and population-level analysis

In short, ATLAS is used to generate data queries, Arachne is used to distribute the queries and receive the results. Note that data queries can also be constructed manually or using the user's own tools.

5.3.1. The EHDEN Project - https://ehden.eu/

The EHDEN project is an IMI funded project, following previous European initiatives in the area of health data infrastructure, and is part of IMI's Big Data for Better Outcomes (BD4BO) programme. EHDEN promotes the adoption of the OHDSI-OMOP Common Data Model through education, certification and financing of Small and Medium Enterprises (SME) to help the Data providers in their efforts to build large data sets according to this model and to undergo data vocabularies mapping.

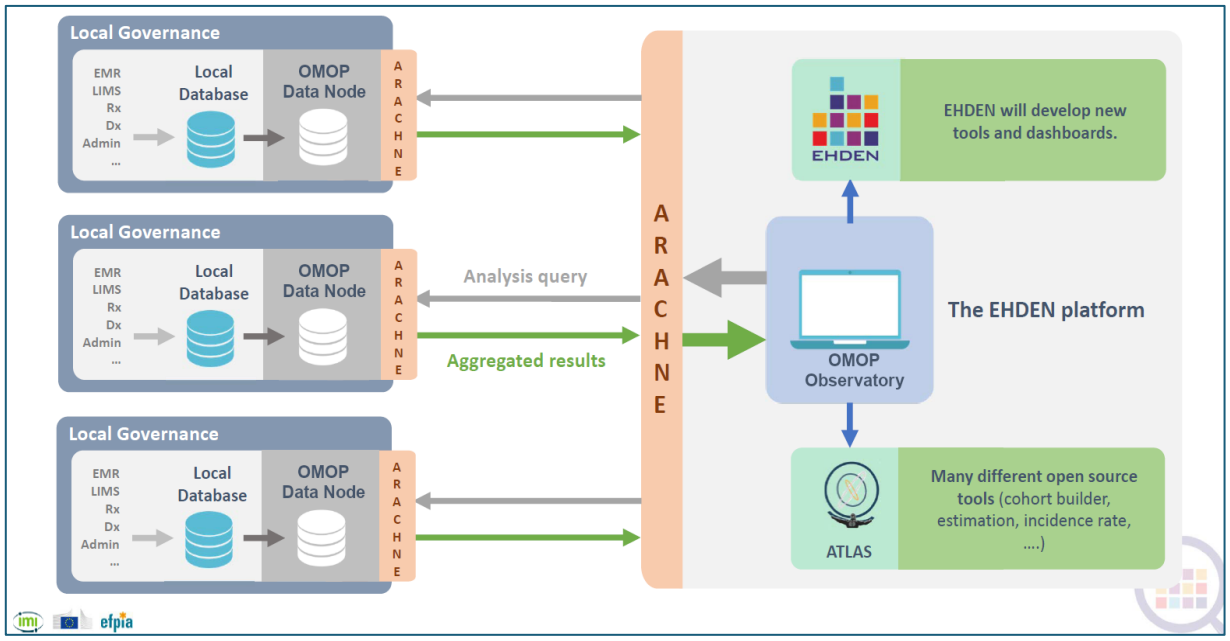


Fig12: The federated network of EHDEN (courtesy of EHDEN project – N. Hughes)

The EHDEN project is creating a federated network with the stated aim of "*allowing access to the data of 100 million EU citizens standardised to a common data model*". The EHDEN network currently contains over 50 European Data Nodes.

Task 3.4 have met with members of the EHDEN project to share experience and knowledge.

6. POC1 Feasibility and POC1

Task 3.4 focussed on examining the feasibility of using OHDSI-OMOP-CDM to support Lab-Select. This examination took two forms, firstly a feasibility study to show that OMOP-CDM could be modified to support AMR data and secondly to build a proof of concept (POC1) in which the technology was used to demonstrate it's capabilities and test its promises.

An 'Action Research' approach was taken to the investigation and an agile iterative approach was used to constructing the POC, which involved:

- Installing and investigating the OHDSI-OMOP software
- Working with sample AMR datasets
- Establishing a federated network
- Iterations of requirements gatherings and POC construction

6.1. Feasibility of using OHDSI to manage AMR data

A database modelling and design exercise was conducted to look in detail at the CDM data structure and to examine how an array of AST/DST results can be stored along with testing resources and protocol (i.e equipment used, specimen processed) for a person. The OHDSI toolkit was used to see if it was possible to map AMR test data from MYLA® middleware and from bioFire FilmArray®, into the CDM.

Firstly, it was observed that the CDM could NOT natively support the complex relationships between antibiotics, pathogen and quantitative / qualitative results that characterise AMR testing.

Secondly it was observed that the CDM model does however allow customisations to accommodate user requirements. A modelling exercise was undertaken to establish a method for accommodating AMR testing results. Three models were developed:

> **Model1 – Isolate View**: A root 'measurement record' is used to capture the isolate identification, which is then linked to multiple measures that represent each individual AST tests and additional measures that contain both the code for the antibiotic and the MIC value.
>
> **Model2 - Cultures View**: A root 'measurement record' is created to capture the culture result (positive or negative), which is then linked to an observation thatvreoresent the identification of the isolate which is then linked to multiple measurements that separately contain antibiotic codes and quantitative values
>
> **Model3 - Simplified Isolates View**: A simplified version of model 1 reducing the number of measurement records required to contain the information by containing code for antibiotic and MIC value within the test measurement

All the above models involve the pairing of LOINC codes (coding scheme for diagnostic test requests) and SNOMED codes (coding scheme used to encode laboratory test results as well as observations derived from test results) as captured in the OMOP-CDM.

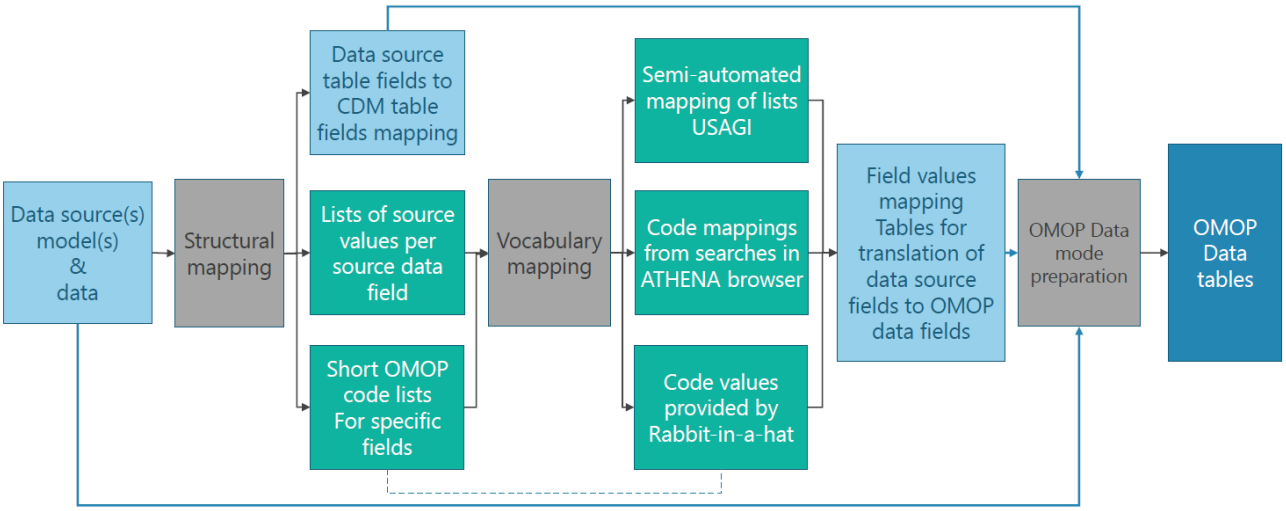


Fig13: Data preparation process using OHDSI tools

The above diagram provides an overview of the steps that must be taken in order to transform an input data source into the OMOP-CDM. In this diagram usage of tools provided by OHDSI are indicated. This gives an indication of the burden each Data Node must accept in both preparing and maintaining the mappings of data from source to CDM via their chosen ETL tools.

Extensive work was performed by bioMérieux in following the process for mapping AMR data into the CDM model. This task was successfully performed.

Although all the proposed models can theoretically be used to support AMR data, the best long-term option is to include AMR support as a native feature of the CDM. Task 3.4 therefore engaged with the OHDSI-OMOP community to validate the correctness of our models and to also propose adoption of AMR support as a standard. These discussions are still ongoing.

For full details of the feasibility work data modelling please refer to "VALUE-DX-WP3-4-POC1-Modeling_Feasibility"

6.2. Feasibility of using OHDSI to deliver Lab-Select

POC1 investigated the feasibility of using OHDSI-OMOP to run a federated network, by implementing a simple federated network based on an Observatory communicating with two test Data Nodes. The Data Nodes use two radically different AMR data sources. Each Data Node owner worked separately from each other and the Observatory. The network implemented 'Model 2 – Cultures View' to support AMR data in the OMOP-CDM.

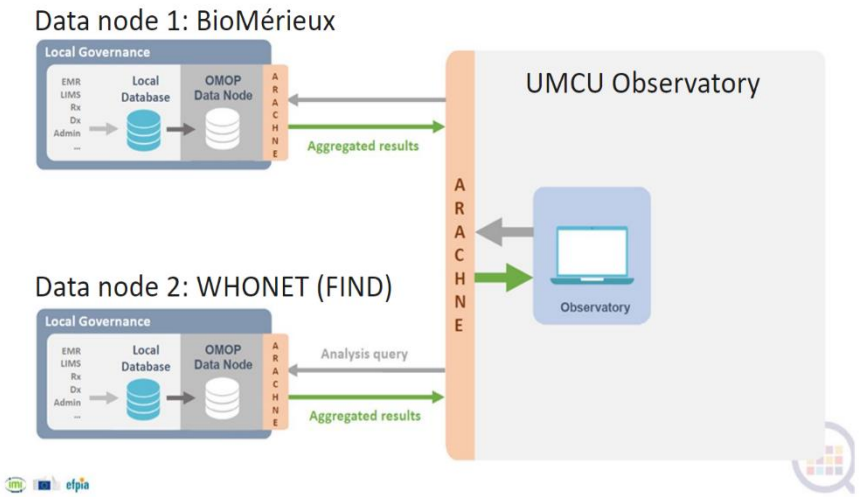


Fig14 POC1 Federated network

The Observatory was established and run by the University Medical Center Utrecht. Data Node1 was established and run by bioMérieux using sample datasets constructed as part of their CDM feasibility work. Data Node 2 was constructed by FIND using test laboratory data from the WHONET microbiology laboratory database software. The WHONET windows client ships with example data, which was manually extended to in order to provide a profile of testing activity overtime, and a one-off load into the CDM was created.

6.2.1. Operating the network.

In order to distribute a query across the participating Data Nodes, several steps need to be taken in the Observatory. Firstly, a new study is created and provided with some metadata on, for example; study type, description, members and participating Data Nodes. The next step is to create and configure a new analysis and load the actual analysis scripts into the Observatory. These analysis scripts will then be distributed to the selected Data Nodes. However, at this stage Data Node owners will only receive a notification that an analysis for this new study is pending execution. Data Node owners will then have to

1) log on to the Observatory to get the analysis files,

2) run the analysis on their local Data Node(s) and

3) manually upload the results into the Observatory and make them visible to the rest of the study team.

Once all invited Data Node owners have run the code and uploaded the results, all aggregated results can be used for further aggregation and/or dissemination.

Having established the network communications and experienced federating queries and aggregating results the final challenge was to exercise the data analytics by creating a series of visualisations with increasing complexity. As an exercise the subject area selected was E. coli resistance, as measures in blood samples. The targeted visualisations loosely reflected the Lab-Select data journey, in that they represent a drilling down of details for scale surveillance level profile to a highly stratified bespoke query (See Section 3.3 ‘Lab-Select Data and User Journey’. These representations also closely matched visualisations available via the WHO GLASS website.

6.2.2. Observations and Conclusions

The POC conclusions are summarised below; please also refer to section 7 ‘Conclusions’ of this document. For full details of the POC1 work and conclusions please refer to document “Report #2 POC1 Documentation”

High-level conclusions are as follows:

- The POC was able to federate AMR data between the Observatory and Data Nodes and support all the visualisation requirements
- Operation of the POC is a highly manual and procedure driven process however automation is possible via a number of options
- Transfer of data can be secured through use of standard networking and systems management tools and approaches
- Protection against running queries that extract case-level data is procedural. There are options for strengthening security by introducing scripts to provide automated scanning of SQL and warnings of potential data breaches to Data Node and Observatory owners
- Because WHONET Data was used as an input for Data Node 2 and GLASS visualisation used as design criteria for example output from the Observatory, the unintended conclusion from this work is that OHDSI-OMOP could feasibility be used as a technology to support the GLASS federated network

At the time of print access to POC1 is by arrangement only. Please contact Frank Leus - University Medical Center Utrecht F.R.Leus@umcutrecht.nl. with any access or demonstration requests.

7. Conclusions

7.1. Conclusions General

Our conclusions with regards to the 'Promises of Federated Networks', as based on our investigation of OHDSI-OMOP implementation are as follows:

Ability to handle the complexities of AMR diagnostic data: **Yes, with constraints.**

As highlighted in this document AMR AST/DST data is not natively supported by OHDSI- OMOP-CDM (V6.0). It is however, possible to accommodate AMR data using available CDM features. See section **Fout! Verwijzingsbron niet gevonden.Fout! Verwijzingsbron niet gevonden.** below.

Ability to accommodate diverse data sources: Yes

The Data Nodes in our POC (bioMérieux and FIND) used completely different data sources to populate the OMOP-CDM using processes and tools unknown to the Observatory. It is clear from the architecture and the experience of the POC that a wide diversity of data sources can be used, so long as the required data fields can be orchestrated by the Data Node owner and accurate mapping into the OMOP-CDM can be achieved. This does not remove the burden of middleware ETL but does distribute the burden of set-up and maintenance across the network and provides a scalable model for maintainability.

Avoid data sovereignty issues: Yes

Architecturally data sovereignty is preserved, in so much as the CDM dataset is completely under the control of the data owner and is hosted, maintained, secured and updated, purely at the discretion of the data owner.

Experiences shared by the EHDEN project leads Task 3.4 to understand that the legal burden is not necessarily vastly reduced by the avoidance of the data sovereignty issues. There is still however the need for data sharing agreements and for diligent data protection considerations in order to be compliant, for example with GDPR. More detailed investigation of the practicalities of establishing compliant data sharing agreements should be conducted as part of any ongoing feasibility study for Lab-Select.

Ability to secure sensitive data and avoid un-authorised sharing: Part true

OMOP-CDM and the Observatory's Lab-Select database, use common industry standard data storage technologies and can therefore be secured and encrypted, and can be subjected to access controls using best practice security measures and standards compliant technology solutions. Transmission of data queries from Observatory to data nodes and the return of aggregated results, can all be secured using standard internet security protocols (e.g. https / TLS) and network security solutions.

Regarding procedural security

Inspection of the architecture and experience from the POC clearly shows us that access to sensitive data is only possible if the data owner grants it. We describe this level of security as procedural. To illustrate consider - if the Observatory issues a data query requesting patient identifiable data, this sensitive data will only be shared if the Data Node accepts the query, runs the query and wilfully returns the sensitive data. Although it is not the policy or intended use of Lab-Select to share sensitive data, technically this can occur either as an authorised procedure or as an un-authorised procedural breach. Task 3.4 are not aware of native OHDSI-OMOP electronic features (in the community edition of Arachne) to protect from procedural breaches.

Electronic methods that enhance procedural security to prevent unauthorised case level queries from being issued and executed, needs further investigation. It is clear that a security wrap can be built, as an 'add on' possibly using the Arachne API, or by augmenting the open source code itself. Such extended functionality would look to scan query requests and provide warnings, confirmations, and enforcement of approvals. These mitigations would be invoked when confronted with any suspicious or clearly identifiable patient level queries, as well as providing sophisticated auditing and logging of queries and data exchange. Our investigation has shown us that there are hooks into the system to allow strengthening of procedural security, however demonstration of such methods of enhancement was out of scope of POC1.

Ability to federate complex custom queries across selected labs: Yes

The ability to hand code SQL queries and distribute them from the Observatory, to as many Data Nodes as desired, shows us that, in principle, any query, as complex as required, can be run against the network, with the same constrains as if the data were all gathered together in a single local database; both methods being constrained by the available data fields, data relationships and the capabilities of SQL logic. The next steps for investigation are to test this conclusion further using real world data and real-world analytical challenges.

Can federate queries run automatically? Yes with external automation support

OHDSI-OMOP processes (community edition) are heavily manual, from creation of SQL for data queries, to sending, executing and aggregation of results returned. Inspection suggests that there are numerous options for creating automated workflows, via use of the enterprise edition or development of external automation scripts, that drive Arachne. This can be achieved through interaction with the API and working with files exposed in the folder system.

There are several options for scheduling queries to run automatically, the essential feature being that Data Nodes can (in the enterprise edition) accept queries to be automatically run without manual approval or intervention. This allows the possibility for regular weekly or monthly update style queries, which can be run in order to build-up a near time surveillance or profile type view of a laboratories activities.

Can allow creation of a secure and usable service wrap to support Lab-Select? Yes

From observation of the OHDSI toolkit, Task 3.4 believes that it is possible to design a fully featured Lab-Select user experience that uses OHDSI-OMOP as a platform component providing the federated network services. In principle API hooks and automation can be used to populate the Lab-Select database and provide a method for delivering bespoke queries. All other Lab-Select functions (i.e. user registration and account management, inventory data collection and maintenance) can be provided by separate established industry standard technologies. Next steps are to model and demonstrate a user journey, across a number of different platform components including an integration with OHDSI.

7.2. Handling AMR data

7.2.1. OHDSI-OMOP-CDM Can support AMR data – with constraints

The OHDSI-OMOP-CDM (V6.0) does not natively support microbiology results, however by leveraging specific features of this version (6.0) the hierarchy of data associated to microbiology results can be represented. This may end-up by making database queries very complex (current OHDSI tools for queries do not yet support this CDM version!).

The constraints of the OHDSI-OMOP vocabularies where that each database field needs to be populated by codes associated to a particular domain (such as measurements fields to be populated by codes extracted from vocabularies of the measurement domain), while enforcing interoperability, forces use of additional tables to ensure compliance to these constraints.

7.2.2. Strategic long-term solution

As an outcome to this study, the team will integrate with the OHDSI-OMOP-CDM community in order that specific requirements (generated by the need to support Microbiology data in the OMOP-CDM model), can be integrated into future evolutions of the data model.

By taking these requirements into account, the OHDSI-OMOP data sets will have the capacity to analyse real world data, that would encompass all data captured from preliminary clinical signs up to final patient diagnosis including all supporting laboratory testing data.

Concerted efforts will be required to influence the community to adopt support for microbiology and AMR, since it represents a significant addition to the data model and requires therefore a powerful case to affect change. Value-Dx is therefore challenged to argue the case for AMR and to pursue this cause.

8. Appendix A: Guide to OHDSI

8.1. Organisations

- **OHDSI**: http://ohdsi.org – This is a collaboration focussed on creation of open source solution for analytics using the Observational Medical Outcomes Partnership's (OMOP) Common Data Model CDM.

- **OMOP's Common Data Model**: A standardized data model to which all participating network 'Data Nodes' must implement. This model is updated circa once a year. https://Github.com/OHDSI/commondatamodel

- **Project EHDEN:** An IMI funded project, promoting and actively using the OHDSI-OMOP Common Data Model to host and query hospital data.

8.2. OHDSI Toolkit Components

The federated network technology stack consists of wide range of opensource tools including (but not limited to) the following components:

- **Arachne:** Toolkit that provide workflow orchestration across federated network. 'Arachne central' provides the 'Observatory' with the ability to establish the network, profile Data Nodes, issue queries and collect results.

- **Arachne Central:** Software component to be use by the centralised Observatory

- **Arachne Data Node:** Software component to be use by all participating Data Nodes

- **Enterprise and Community Editions:** Arachne is available is both a paid service (Enterprise Edition) and Open Source version (Community)

- **ATLAS:** A web-based integrated platform for database exploration, standardized vocabulary browsing, cohort definition, and population-level analysis. This tool is used to generate queries, Arachne is used to distribute the queries. Note that queries can also be constructed manually or using other tools.

- **Implementation Tool Kit:**
  Although the user's own tools can be used to prepare and map data for hosting in the OMOP-CDM, OHDSI provides a number of open source tools for this purpose. These tools include the following:

- **Athena:** http://athena.ohdsi.org/search-terms/terms The Standard Vocabulary 'foundational tool' initially developed to enable transparent and consistent content across disparate observational databases.
- **ACHILLES** – A standardized database profiling tool for database characterization and data quality assessment.

Other tools include:

- White Rabbit – Software tool to prepare the ETL process from longitudinal healthcare databases into the OMOP CDM.
- Rabbit in a Hat – Software tool that generates documentation for the ETL process. Can use reports generated by White Rabbit
- Usagi – Tool to help create mappings between coding systems and the vocabulary standard concepts

9. Appendix B: EHDEN organisations

Extract from project EHDEN factsheet https://www.imi.europa.eu/projects-results/project-factsheets/ehden

EFPIA companies

- Abbvie Inc, North Chicago, Illinois, United States
- Astrazeneca AB, Södertälje, Sweden
- Bayer Aktiengesellschaft, Leverkusen, Germany
- Celgene Management SARL, Couvet, Switzerland
- Eli Lilly And Company Limited, Basingstoke, United Kingdom
- Institut De Recherches Internationales Servier Iris, Suresnes, France
- Janssen Pharmaceutica Nv, Beerse, Belgium
- Novartis Pharma AG, Basel, Switzerland
- Pfizer Limited, Sandwich, Kent , United Kingdom
- Sanofi-Aventis Recherche & Developpement, Chilly Mazarin, France
- UCB Biopharma SRL, Brussels, Belgium

Universities, research organisations, public bodies, non-profit groups

- Erasmus Universitair Medisch Centrum Rotterdam, Rotterdam, Netherlands
- Forum Des Patients Europeens, 1040, Belgium
- National Institute For Health And Care Excellence, Manchester, United Kingdom
- Stiftelsen WHO Collaborating Centre For International Drug Monitoring, Uppsala, Sweden
- Tartu Ulikool, Tartu, Estonia
- Universidade De Aveiro, Aveiro, Portugal
- University Of Oxford, Oxford, United Kingdom

Small and medium-sized enterprises (SMEs) and mid-sized companies (<€500 m turnover)

- International Consortium For Healthoutcomes Measurement LTD, London, United Kingdom
- Odysseus Data Services Sro, Praha, Czech Republic
- Synapse Research Management Partners SL, Barcelona, Spain
- The Hyve BV, Utrecht, Netherlands

Patient organisations

- European Patients' Forum (EPF), Brussels, Belgium

10. Appendix C: List of Common LIS software

Below is a list of common LMIS software used as part of the connectivity section of the honest prep questionnaire.

- Accelrys LIMS from Accelrys
- BaseSpace Clarity LIMS from Illumina
- BIOVIA LIMS from Dassault Systèmes
- CCLAS from ABB Group
- Clinsis from Clinsis
- ELab from LabLynx
- Exemplar Biomarker Discovery from Sapio Sciences
- Exemplar Dx LIMS from Sapio Sciences
- Exemplar Research LIMS from Sapio Sciences
- Hach WIMS from Hach Company
- LABbase from Analytik Jena
- LabCollector from AgileBio
- LabWare LIMS from LabWare, Inc.
- MetaField Lab from Agile Frameworks
- Nautilus LIMS from Thermo Fisher Scientific
- NuGenesis 8 from Waters Corporation
- OmicsHub from Integromics
- readyLIMS from Analytik Jena
- SampleManager LIMS from Thermo Fisher Scientific
- SampleTrack from Bruker
- SIMATIC IT R&D Suite from Siemens
- SLIMS from Agilent Technologies
- STARLIMS from Abbott Laboratories
- TrakCare Lab Enterprise from InterSystems
- Watson LIMS from Thermo Fisher Scientific
- webLIMS from LabLynx

11. Bibliography

European Medicines Agency *(2018)* A common Data Model for Europe? – Why? Which? How?, Workshop report from a meeting held at the European Medicines Agency, London, United Kingdom.

EDM Forum (2016) Data Extraction And Management In Networks Of Observational Health Care Databases For Scientific Research - . Gini, M. Schuemie, J. Brown, P. Ryan

PLOS one (2019) Data model harmonization for the All Of Us Research Program!: Transforming i2b2 data into the OMOP common data model. J.G. Klann, M.A. Joss, K. Embree, S.N. Murphy

OXFORD University press (2015) Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. E.A Voss, R. Makadia, A. Matcho, Q. Ma, C. Knoll, M. Schuemie, FJ DeFalco, A. Londhe, V. Zhu, PB. Ryan

The book of OHDSI, *Observational Health Data Sciences and Informatics,* https://github.com/OHDSI/TheBookOfOhdsi

OMOP's Common Data Model: Standardized data model underpinning the OHDSI federated network. http://github.com/OHDSI/commondatamodel

EHDEN Project - https://www.ehden.eu/
Project to address the current challenges in generating insights and evidence from real-world clinical data at scale.

WHO GLASS Country Profiles - https://apps.who.int/gho/tableau-public/tpc-frame.jsp?id=2004
Online global surveillance observatory, reporting on WHO Glass priority AMR pathogens

www.imi.europa.eu www.value-dx.eu