Healthdata.be

Healthdata.be enables all Belgian clinical laboratories to report – for multiple projects - in a standardized way towards a **centralized data management** and analysis platform used by researchers and surveillance teams. To improve performance and scalability, the new technical architecture HD4DP2.0 is offered, deployed and maintained by healthdata.be to all Belgian healthcare institutions. An application programming interface (API) for the exchange of the clinical, epidemiological and genomic data will be composed of Fast Healthcare Interoperability Resources (FHIR) as designed by the standards definition organization Health Level Seven International (HL7). Given the nature of the clinical, epidemiological and genomic data to be collected by healthdata.be, the services of an **independent trusted third party (eHealth)** will be used for pseudonymization and the **end-to-end encryption** of the transfer of the person identifiers. Translation and harmonization of currently used codes and values towards international adopted terminology systems (*i.e.* SNOMED CT, LOINC,…) will be performed. In the frame of this HERA-BE-Incubator project: healthdata.be will assure the collection of data and its transmission in the expected formats to various end parties serving surveillance and outbreak investigations purposes. The figure below gives an overview of the high level design (Figure 2)

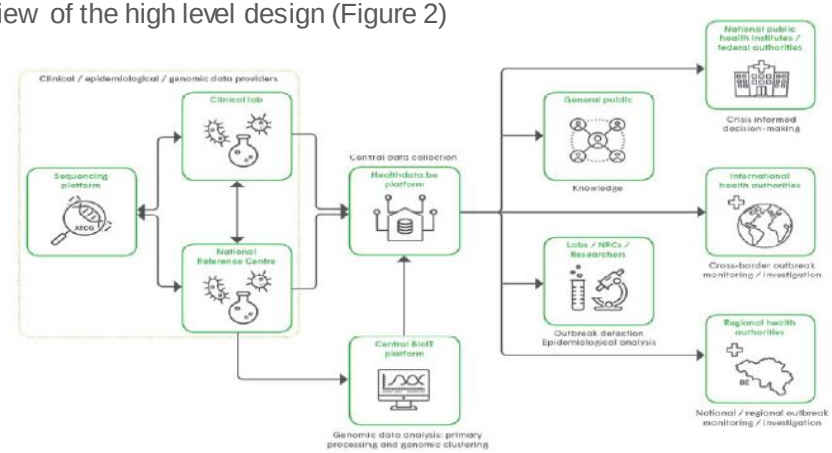

Figure 2: High level design of the national eHealth infrastructure for genomic-epidemiologic surveillance of infectious diseases. Epidemiologic and/or genomic data, originating from a national reference center or clinical laboratorium, will be collected with a standardised variable set on a centralised platform. Raw genomic data will be dispatched immediately to a central bioIT-platform where an automated and harmonised bioinformatics analysis will be performed in the primary processing, through the consensus pipeline per pathogen. Resulting or uploaded FASTA files and genomic indicators will be stored on the healthdata.be-platform. From the centralized genomic-epidemiologic data

More information?
Scan the QR-code

HERA-BE-INCUBATOR-WGS: TOWARDS A NEW (H)ERA

National infrastructure for genomic epidemiological surveillance of infectious diseases

Introduction HERA-BE-Incubator

As a response to the COVID-19 pandemic, the European Commission established the European Health Emergency Preparedness and Response Authority (HERA). It aims to strengthen the European Health Union with a better preparedness and response to future national and cross-border or pandemic outbreaks of infectious diseases. The HERA-BE-Incubator project was launched by HERA, and its implementation in Belgium makes it possible to develop a national infrastructure for the collection of Whole Genome Sequencing (WGS) data and to build a sustainable infrastructure in which genomic data can be linked to epidemiological and clinical data. This can then be used more broadly to anticipate outbreaks and manage related risks. The project aims at the transactional processing of larger data volumes in case of a major outbreak or pandemic. **Currently, the HERA-BE-Incubator project is set up as a proof-of-principle project for five national reference centres (NRCs) for human microbiology and it uses the healthdata.be platform. The project is coordinated by Sciensano and aims to welcome the participation of additional laboratories in the near future.**

National Reference Centers (NRCs)

Genomic analyses become more and more integrated in the workflows of the NRCs and even non-NRC clinical labs, as shown during the COVID-19 pandemic. Data is currently scattered amongst different organisations and will be scattered further in the coming years when sequencing will be implemented more easily by laboratories. By centralising genomic data of pathogens, and providing automated and validated bioinformatics pipelines, **it allows for the integration of data from multiple sources.** Through harmonised processing, **the data are more comparable and can be used for cluster detection**. In addition, the sensitivity to identify a possible outbreak increases and **monitoring will be extended** of for example, predicted resistance profiles.

"The project will strengthen the outbreak detection and investigation by centralizing genomic, clinical, and epidemiological surveillance."

Table 1: Opportunities for the public health actors

What is in it for me?
Epidemiological and genomic data can be **linked** and analysed together
Data is collected **near real-time**
Timely and sensitive variant/outbreak detection
Improved **outbreak management** and investigation by enhanced data availability
Supporting public health activities and increasing surveillance
Harmonized and automated bioinformatics pipeline for all labs involved
Extended monitoring of predicted **antimicrobial resistance profiles** through genomics
Automated uploading of large data files possible
Possibility to **upload FASTQ** files to ENA through the system

"Establishing, strengthening and transforming the infrastructure developed during the COVID-19 pandemic into broader preparedness and response to pandemic outbreaks"

Bioinformatics

A central bioinformatics environment will be established that will provide the required bioinformatics tools, pipelines, and databases to allow **standardized, harmonized, and traceable processing of all WGS data for pathogens** to allow full genomic typing and characterization including a wide range of indicators (*e.g.* cgMLST, predicted antimicrobial resistance (AMR) characterization, virulence gene prediction, serotyping, …). Support for all relevant pathogens will gradually be implemented. Results will be reported to the data providers in clear and intuitive reports. Genomic indicators obtained from samples will be fed into a centralized national molecular database. The data of the samples will be **reanalyzed periodically with updated reference databases** to ensure new genetic information (*e.g.* new AMR genes, cgMLST markers, …) is incorporated into the national genomic-epidemiologic framework. Additionally, procedures for automated cluster detection will be established based on genomic relatedness to trigger automated alerts for potential clusters to guide and facilitate more in-depth cluster investigation.

"The centralized and standardized approach opens opportunities to start a new era of health emergency preparedness in Belgium"

Epidemiology of infectious diseases and public health

The Belgian network of the 41 NRCs ensures the expertise, diagnostics and monitoring of a range of important pathogens for public health. The data collection is currently centralized, processed and made available for the public health actors and the public ("classic" NRC data collection). However, a **patient-identifier** is lacking. Transition to a data collection system based on a unique patient-identifier opens up public health opportunities, such as the possibility of data linkage. This allows retrospective and prospective cohort, case-case and case-control studies, and will all together increase our understanding of the impact and epidemiology of infectious diseases. Data will be collected using standardized **data collection definitions** (**DCDs**) and covers both the "classic" NRC data as well as the genomic data, including WGS data. This will result in a combined data **collection of genomic, clinical and epidemiological data** facilitated by healthdata.be. This centralized data collection will serve the various public health actors, according their needs and mandates. The national framework aims to provide better data during all 3 phases of signal-detection and follow-up by enhanced data (Figure 1).

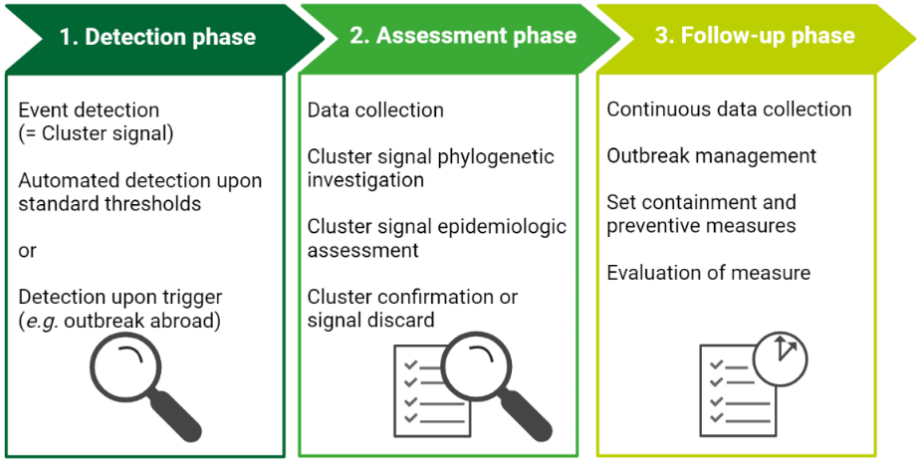

Figure 1. Visual representation of the three phases of epidemiological signal detection and monitoring, each important for pandemic preparedness. Even amplification of all these phases will increase pandemic preparedness. The three phases include: 1) detection of possible outbreaks, 2) assessment of possible outbreaks and 3) follow-up of the outbreak, i.e. management of the outbreak. Central genomic data allow automated cluster detection (phase 1): based on similarities in the genome, alerts are generated for possible clusters (= cluster signal). However, these possible clusters require an individual evaluation (phase 2), including a detailed phylogenetic study and a contextual and epidemiological assessment. After confirmation of the cluster or outbreak, preventive and containment measures can be taken if necessary (phase 3) and further follow-up can be planned. However, the HERA-BE incubator project currently focuses on creating the necessary infrastructure and does not implement all the above phases in detail.