National Institute for Public Health
and the Environment
Ministry of Health, Welfare and Sport

Improving laboratory-based surveillance of **infectious diseases** in the Netherlands

RIVM report 2022-0253
I. van Walle et al.

Colophon

RIVM attaches a great deal of importance to the accessibility of its products. However, it is at present not yet possible to provide this document in a completely accessible form. If a part is not accessible, it is mentioned as such. Also see www.rivm.nl/en/accessibility

DOI 10.21945/RIVM-2022-0253

A. Backx (author), RIVM
G. Haringhuizen (author), RIVM
G. Klous (author), RIVM
S. Koullali (author), RIVM
T. Kraaij (author), RIVM
A. Kroneman (author), RIVM
N. Lieftink (author), RIVM
N. Vonk (author), RIVM
I. van Walle (author), RIVM

Contact:
Ivo van Walle
Laboratory Response COVID-19 department
ivo.van.walle@rivm.nl

This investigation was performed by order, and for the account, of the European Centre for Disease Prevention and Control (ECDC), within the framework of GRANT/2021/PHF/23776

Technical Report for
ECDC GRANT/2021/PHF/23776

Enhancing Whole Genome Sequencing (WGS) and/or Reverse Transcription Polymerase Chain Reaction (RT-PCR) national infrastructures and capacities to respond to the Covid-19 pandemic in the European Union and European Economic Area

Published by:
National Institute for Public Health and the Environment, RIVM
P.O. Box 1 | 3720 BA Bilthoven
The Netherlands
www.rivm.nl/en

Synopsis

Improving laboratory-based surveillance of infectious diseases in the Netherlands

In the Netherlands, there are special laboratories that test whether people have an infectious disease. The analysis of how many people become ill and by which variant of a virus or bacterium is known as laboratory-based surveillance of infectious diseases. RIVM performs this analysis in cooperation with these 'medical-microbiological laboratories' and with the Municipal Public Health Services.

During the coronavirus pandemic, it became increasingly clear that laboratories, including those of RIVM, must be able to test for infectious diseases on a large scale. They must also be able to see which variants of a pathogen are involved, such as the Delta and Omicron variants of the SARS-CoV-2 virus. In addition, it must be possible to exchange laboratory data with RIVM safely and efficiently, so that RIVM can monitor how a pandemic is progressing in the Netherlands. For infectious diseases other than COVID-19, it is equally important that RIVM is able to monitor how they are evolving through laboratory tests.

Currently, different technical systems are used for different diseases. Some of these systems are outdated. To better cope with a possible new epidemic, RIVM will set up a new technical platform for laboratory-based surveillance. This means that the surveillance of various infectious diseases can be technically supported in the same way, both during data storage and data analysis. This will also make it possible to exchange data in the same way, which is more efficient and clearer for all laboratories concerned. Furthermore, the platform will be better able to process large amounts of test results and make it easy to store and analyse data about new pathogens.

In preparation for this improvement, RIVM has described the legal context for the exchange of data, such as data privacy legislation. It has also described the first full version of the technical specifications and functionalities that the platform must meet.

RIVM will develop and refine the most important components of the platform as from 2023. In order to develop the platform as carefully as possible, its individual components will be tackled one by one. This way, RIVM can learn from each experience. This approach will be beneficial for both the quality and cost of the ultimate platform.

Keywords: COVID-19, surveillance, laboratory, testing, preparedness, pandemic

Publiekssamenvatting

Verbetering van het toezicht op infectieziekten op basis van laboratoriumonderzoek

In Nederland testen speciale laboratoria of mensen een infectieziekte hebben. Dit geeft inzicht in hoeveel mensen ziek worden en van welke variant van een virus of bacterie. Het RIVM doet deze analyse in samenwerking met deze 'medisch microbiologische laboratoria' en met de GGD'en.

Tijdens de coronapandemie werd duidelijker dat laboratoria, inclusief die van het RIVM, goed in staat moeten zijn om op grote schaal op infectieziekten te testen. Ook moeten zij kunnen bepalen om welke varianten van een ziekteverwekker het gaat, zoals de delta- en omikron-varianten van het virus SARS-CoV-2. Daarnaast moeten de gegevens van de laboratoria veilig en efficiënt met RIVM kunnen worden uitgewisseld om te kunnen bepalen hoe een pandemie in Nederland verloopt. Ook voor andere infectieziekten dan COVID-19 is het belangrijk om via laboratoriumtesten in kaart te kunnen brengen hoe ze zich ontwikkelen.

Op dit moment worden voor verschillende ziekten verschillende technische systemen gebruikt. Een deel van deze systemen is verouderd. Om een eventuele nieuwe epidemie beter aan te kunnen, gaat het RIVM hiervoor een nieuw technisch platform opzetten. Hiermee kan het toezicht op verschillende infectieziekten op dezelfde manier technisch worden ondersteund, zowel bij de opslag als de analyse van de data. Zo kunnen de gegevens op dezelfde manier worden uitgewisseld. Dat is efficiënter en voor alle laboratoria overzichtelijker. Verder moet het systeem beter in staat zijn om grote hoeveelheden testresultaten te kunnen verwerken. Ook wordt het makkelijker om data van nieuwe ziekteverwekkers op te slaan en te analyseren.

Als voorbereiding op deze verbeterslag heeft het RIVM de juridische context beschreven voor de uitwisseling van de gegevens, zoals de privacywetgeving. Het heeft ook de eerste volledige versie beschreven van de technische specificaties en functionaliteiten waaraan het platform moet voldoen.

Vanaf 2023 ontwikkelt en verfijnt het RIVM de belangrijkste delen van het platform. Om het platform zorgvuldig te kunnen ontwikkelen, worden losse onderdelen een voor een aangepakt. Zo kan het RIVM telkens leren van de ervaringen. Deze aanpak is gunstig voor zowel de kwaliteit als de kostprijs van het uiteindelijke platform.

Kernwoorden: COVID-19, surveillance, laboratorium, testen, paraatheid, pandemie

Contents

Summary

In the wake of the COVID-19 pandemic, improving laboratory preparedness for an epidemic or pandemic has received a lot of attention. This report focuses on improving data sharing and preparedness for laboratory-based surveillance of pathogens (lab surveillance) within the Netherlands. This entails the collection and analysis of primarily human sample-based laboratory data for the primary purpose of signal detection. Signal detection includes detection of trends in SARS-CoV-2 variants, but also detection of both short-term and long-term signals that are relevant to other pathogens. In addition, preparedness implies, for instance, being easily adjustable to accommodate lab surveillance for a newly emerging pathogen.

Lab surveillance in the Netherlands is organised principally around three partners, i.e. the medical microbiological laboratories (MMLs), the municipal health services (MHSs) and the Dutch National Institute for Public Health and the Environment (RIVM). The legal aspects of lab surveillance are strongly linked with the fact that this data is considered personal data under the EU General Data Protection Regulation (GDPR). Taking into account GDPR-related and public health legislation in the Netherlands, we conclude that the legal framework could be improved to better support lab surveillance and surveillance of infectious diseases in general. This will also help to better organise the collaboration between the various partners, which is a crucial success factor. An important aspect of collaboration between the partners is also potential financial compensation of MMLs for their contribution to public health, which falls outside the scope of this report.

A second crucial success factor is having a good technical platform at RIVM to support lab surveillance, which increases the preparedness for any newly emerging pathogen and is able to better address the needs of both MMLs and MHSs. Currently, the Netherlands uses several systems combined with substantial manual work, and which are nearing their technical end-of-life. Therefore, we also describe here how the future platform for lab surveillance could be developed and embedded within a larger environment for infectious disease surveillance, based on a number of fundamental principles, including compliance with the legal framework. These are translated into an initial set of requirements or specifications for the platform.

Following this preparatory work, the next steps include drawing up Terms of Use for the platform and conducting a pilot phase with selected pathogens, in which components are rapidly developed and adjusted on the basis of user input. This also removes uncertainty and saves costs compared to starting off with a large-scale implementation. At the same time, the lab surveillance platform needs to be made into a production-grade platform that fits within the larger effort at RIVM to improve its surveillance environment in general.

1 Introduction

In the wake of the COVID-19 pandemic, improving laboratory preparedness for an epidemic or pandemic has received a lot of attention. On 17 February 2021, the European Commission launched 'HERA Incubator', which is a new EU bio-defence preparedness plan against SARS-CoV-2 variants. This report is supported by one of the five action areas within this plan, i.e. the rapid detection of SARS-CoV-2 variants, and focuses on improving data sharing and preparedness within the Netherlands for laboratory-based surveillance of pathogens (lab surveillance) in general. This includes the purpose of signal detection, such as detection of trends in SARS-CoV-2 variants, but also detection of signals that are relevant for other pathogens. In addition, data sharing should be easily adjustable to accommodate lab surveillance for any newly emerging pathogen. We also keep in mind the European perspective, by clearly distinguishing in the text, where relevant, what is specific to the Netherlands and what is more broadly applicable.

We distinguish two major factors that contribute to improved lab surveillance and preparedness. The first includes legal and organisational aspects, such as the legal basis for lab surveillance and other relevant aspects of compliance with applicable law, as well as the way lab surveillance is or can be organised practically among the three main (groups of) partners in the Netherlands. These partners include the medical microbiological laboratories (MMLs), the municipal health services (MHSs), who are responsible for the public health at the regional level, and the Dutch National Institute for Public Health and the Environment (RIVM), which is responsible for all public health surveillance at national level. All the legal aspects that apply to lab surveillance, both on EU level and within the Netherlands, as well as some of the issues observed, are covered in a separate chapter. The chapter on the future situation includes specific sections on collaboration.

The second major factor is the technical platform that supports the lab surveillance processes that needs to be improved. This fits in with a broader effort of RIVM to improve the technical environment for surveillance in general. In the long run, the intention is to have the current existing collaboration agreements and systems for lab surveillance, such as Type-Ned MRSA and CPE, virological weekly reports, SeqNeth-SARS-CoV-2 and Molecular Enterovirus and Parechovirus Platform replaced by a single structure for collaboration and an associated lab surveillance platform. These currently existing processes and systems are described in the corresponding chapter. Subsequently, the needs for an improved future lab surveillance platform are worked out in three stages, starting from fundamental principles for lab surveillance and ending with an initial set of requirements. This process includes feedback from MMLs and MHSs, and traceability to legal requirements is maintained as well.

2 Legal aspects

The purpose of this chapter is to map out the legal aspects that directly impact public health (lab) surveillance, both within EU Member States in general, and specifically within the Netherlands. In particular, since the collected data typically are – pseudonymised – personal data, data collection and processing for lab surveillance must be legitimated on one of the legal bases for processing personal and sensitive data as set out in the GDPR, as well as the legal bases set out in other national and international legislation. The legal bases will also depend upon the actors involved, the purposes of the data collection and the characteristics of the healthcare system.

The applicable law is described at the level of individual articles within each law. This also serves to link the compliance requirements for a future lab surveillance platform more clearly (see Chapter 4) to the corresponding articles in law. Where relevant, an assessment is added to the article with respect to its application to or impact on public health. This assessment is given in italics.

In addition to this overview of applicable law, further in-depth clarification is given on particular aspects. These include considering the different classes of data from a legal perspective, and the ability to link lab surveillance data to separately collected data on epidemiological cases.

2.1 European Union law

The EU's constitutional basis is set out by two core international treaties: the Treaty establishing the European Economic Community signed in Rome in 1957 (formerly known as the Treaty of Rome), and the Treaty on European Union signed in Maastricht in 1992 (formerly known as the Maastricht Treaty). Through them, the EU can create e.g. Regulations, which are legal acts that all EU Member States are bound to.

2.1.1 EU General Data Protection Regulation

Full name: EU Regulation (EU) 2016/679 General Data Protection Regulation (GDPR) (European Parliament and the Council of the European Union, 2016).

The EU GDPR regulates the protection of natural persons with regard to processing personal data. Here, we describe the main relevant elements of this regulation for public health in general and for (laboratory) surveillance in particular.

Article 4 (13) provides a definition of genetic data and refers to genetic data as a special category of personal data in Article 9. *The GDPR refers to genetic data but not genomic data. However, there is a certain level of uncertainty and disagreement as to whether genomic data is also covered by the definition of genetic data in the GDPR.*

Article 4 (15) defines data concerning health as personal data related to the physical and mental health of a natural person, including the provision of healthcare services, which reveal information about his or her health status. *In practice, however, health data is often understood as any personal data generated within healthcare systems, and some may include data concerning health that are collected by citizens and patients through wearable devices, apps and self-reported information.*

Article 5 imposes a number of requirements regarding the processing of personal data:

- Lawful, fair and transparent processing
- Purpose limitation
- Data minimisation
- Accuracy
- Storage limitation
- Integrity and confidentiality
- Accountability

With respect to purpose limitation, further processing for scientific research or statistical purposes is not considered to be incompatible with the initial purposes, provided that appropriate safeguards are taken (see Article 89).

Article 6 foresees six possible legal bases for the lawful processing of personal data. All data controllers must be able to point out the legal base being used for any act of data processing. This article lists two possibilities on the lawfulness of processing that are relevant for public health:

- Consent by the data subject
- Performance of a task carried out in the public interest. In this case, Member State law can determine specific requirements for the processing. *This is the case for the Netherlands, as further described in section 2.2.*

In accordance with Article 6(4) GDPR, data can only be further processed for a purpose other than the purpose stated at the time of collection if it is compatible with that purpose (known as the purpose limitation principle). *When it comes to research, however, this should be read in conjunction with Article 5(1)(b) which carves out a privileged position for research, stating that further processing for scientific research purposes in accordance with article 89(1) is not considered incompatible with the principle purpose.*

Article 9 defines special categories of personal data. These include categories relevant for public health, namely genetic data, data concerning health and data concerning a person’s sex life or sexual orientation. The processing of these special categories of data is generally prohibited, with one of the exceptions being for reasons of public interest in the area of public health and only on the basis of specific Union or Member State Law. *This is the case for the Netherlands, as further described in section 2.2.3.*

Article 16 defines the right of data subjects to rectification. While unlikely for public health and surveillance, the possibility that a data subject exercises this right must be taken into account.

Article 17 defines the right of data subjects to erasure. Personal data must be erased when they are no longer necessary for the purposes for which they were collected or otherwise processed, when an obligation to erase exists under Union or Member State law or when consent is withdrawn. *They do not need to be erased, however, for reasons of public interest in the area of public health.*

Article 18 defines the right of data subjects to restriction of processing. While unlikely for public health and surveillance, the possibility that a data subject exercises this right must be taken into account.

Article 21 defines the right of the data subject to object to processing. This includes processing on the grounds of performing a task carried out in the public interest. *In the Netherlands, this right is not restricted for processing for reasons of public interest in the area of public health.*

Articles 22 and 23 define the obligations of the controller. In particular, data protection by design and by default requires the implementation of technical and organisational data protection measures to safeguard the principles of Article 5, taking into account risk likelihood and severity for rights and freedoms posed by the processing. Certification can be used as an element to demonstrate compliance.

Article 32 defines the obligation of the controller and processor to implement technical and organisational measures to ensure a level of data security appropriate to the risk. These include (i) pseudonymisation and encryption, (ii) confidentiality, integrity, availability and resilience of processing systems and services, (iii) ability to timely restore availability and access and (iv) regular testing of security measures.

Articles 33 and 34 define the obligation of the controller to notify a personal data breach to both the supervisory authority and the data subject.

Article 35 defines the obligation of the controller to carry out a data protection impact assessment where there is likely to be a high risk to the rights and freedoms of natural persons. *This is required in case of large-scale processing of the special categories of personal data in Article 9, which is the case for public health lab surveillance.*

Article 36 defines the obligation of the controller to consult the supervisory authority prior to processing, in case the data protection impact assessment indicates a high risk in the absence of mitigating measures taken by the controller.

Article 87 on the processing of the national identification number gives Member States the right to define specific conditions on the processing of this number. *In the Netherlands, this is further defined in the Implementing Act GDPR.*

Article 89 defines safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes. In particular, Union or Member State law may provide for such derogations. *Public health surveillance activities also fall under scientific and statistical purposes.*

2.1.2 EU Regulation on the establishment of the European Centre for disease prevention and control

Full name: EU Regulation (EC) No 851/2004 of the European Parliament and of the Council of 21 April 2004 establishing a European Centre for disease prevention and control. (European Parliament and the Council of the European Union, 2004)

RIVM is a 'competent body' in the sense of Regulation EC 851/200416 and will submit anonymised data to the European Centre for Disease Prevention and Control (ECDC). The ECDC has no mandate to receive data directly from other organisations, such as MMLs or MHSs, finding a notifiable disease, and the latter have no corresponding duty to submit such data to the ECDC.

2.2 Dutch law

2.2.1 Implementing act GDPR

Dutch name: *Uitvoeringswet Algemene verordening gegevensbescherming* (Staten-Generaal, 2016).

This law implements several aspects of the EU GDPR, as foreseen in the latter. Here, we describe the main relevant elements of this regulation for public health in general and for (laboratory) surveillance in particular.

Article 22 confirms the exception in Article 9 (e) of the GDPR that the special categories of personal data can be processed if this is necessary for the working of the organisation in question.

Article 24 confirms that special categories of personal data can be processed if collected for statistical purposes that serve a public interest, for which obtained consent requires an unreasonable effort and appropriate safeguards are in place to protect the rights and freedoms of the individual. *Public health surveillance activities meet these requirements as they are continuous and require a sound statistical basis for collection, and as such can process these special categories of personal data.*

Article 25 restricts the processing of the special personal data category of race or ethnicity. This data can only be processed with the purpose of reducing factual disadvantages that individuals of that race or ethnicity may be subject to. It is not allowed if the individual objects to the processing of this data. *Public health surveillance activities need to demonstrate that they meet these requirements on a pathogen-by-pathogen or disease-by-disease basis and as such can process the special category of personal data of race or ethnicity, since not all diseases are associated with race or ethnicity.*

Article 28 restricts the processing of the special personal data category of genetic data. This data can only be processed when consent is given and appropriate safeguards are in place to protect the rights and freedoms of the individual. *In the context of public health, it should be clarified that this refers to human genetic data, and does not cover genetic data of any other organisms, pathogenic or not, sampled from an individual. The latter genetic data is not considered personal data.*

Article 30 restricts the processing of the special personal data category of health data. This data can only be processed by healthcare organisations for the purpose of treatment or care for the individual or for management of the organisation.

Article 44 suspends several rights of individuals under GDPR when data is processed for scientific research or statistical purposes, provided that sufficient safeguards are in place that the data can only be used for these purposes. This includes the right of access, the right to rectification and the right to restriction of processing (Articles 15, 16, 18 of the GDPR). *Public health surveillance activities meet these requirements only if the data is not subsequently used for response (outbreak investigations or contact tracing), in which case individuals may need to be identified or contacted. In the latter case, these rights cannot be seen as suspended.*

Article 46 restricts the processing of a national identification number to cases where this is specifically defined by law. A Dutch Citizen Service Number (BSN) is an official national identification in the Netherlands. In the Netherlands, the BSN belongs to the so-called 'special categories of personal data'. Processing of special categories of data is prohibited unless a specific legal exception applies. Additional cases where a national identification number can be processed can be defined through executive action (*algemene maatregel van bestuur*). *For government entities, including RIVM, the processing of the national identification number is permitted, when necessary, through the General Provisions Citizen Service Number Act (article 10). An encrypted Citizen Service Number is still considered a national identification number.*

2.2.2 Public health act

Dutch name: *Wet publieke gezondheid* (Staten-Generaal, 2008).

This is the main law, which specifies how public health is organised and what rights and duties different actors have. Here, we describe the most relevant elements of this regulation for (laboratory) surveillance of infectious diseases.

Article 1 defines several categories of infectious diseases:

- Category A: MERS-CoV, smallpox, polio, SARS and viral hemorrhagic fever, monkeypox
- Category B1: human infection caused by an animal influenza virus, diphtheria, plague, rabies or tuberculosis
- Category B2: typhoid fever, cholera, hepatitis A, B and C, whooping cough, measles, paratyphoid fever, rubella, shigellosis, STEC/EHEC infection, invasive streptococcus A infection, or

infection likely resulting from the consumption of food when at least two persons have been infected.
- Category C: defined by the executive branch in accordance with Article 19.

Article 6 generally tasks municipalities with preventing and controlling infectious diseases. In case of a category-A disease and a new subtype of human influenza virus, the safety region is responsible instead.

Article 6(b) confers the responsibility for the execution, monitoring and evaluation of the vaccination programme to RIVM.

Article 6(c) generally tasks RIVM with preventing and controlling infectious diseases. To execute these tasks, it has the right to process the special personal data category of health data. RIVM may only process pseudonymised personal data, and may continue to process the personal data of an individual even when they make use of their right to object. Any reference laboratories whose services RIVM uses, have these same rights to process personal data.

Article 12(a) tasks RIVM with coordinating execution, monitor and evaluate population screening in the area of healthcare.

Article 14 establishes the municipal health services (MHSs).

Article 19 empowers the executive branch to determine which diseases belong to Category C.

Article 21 compels the treating physician to report cases of a likely infectious disease of unknown aetiology to the MHS.

Article 22 compels the treating physician to report cases of infectious diseases to the MHS.

Article 24 specifies the data that needs to be reported by the treating physician to the MHS for cases that fall under Article 21.

Article 25 compels a laboratory that performs diagnostic tests for a physician to report a positive result of an infectious agent to the MHS. The MHS can also request further diagnostic tests to be performed.

Article 27 compels the MHS to report cases of Category A diseases to the safety region.

Article 28 compels the MHS to report cases of Category A, B1 and B2 diseases to RIVM. It also specifies the data that must be reported. The national identifier is not included in this list.

2.2.3 The RIVM act

Dutch name: *Wet op het RIVM* (Staten-Generaal, 1996).

RIVM is an independent agency of the Dutch Ministry of Health, Welfare and Sport (VWS). The tasks and responsibilities of RIVM, and how RIVM

ensures its independence, have been laid down in the RIVM act of 21 October 1996.

Article 3 defines the tasks of RIVM. These include:

- Monitoring and surveillance, and applied research to support public health.
- Reporting periodically about the status of public health and future developments.
- Participating in international collaborations and research, and assuming a coordinating role in such cases, where desirable and related to its tasks.
- Other tasks as assigned by the minister of health.

RIVM is allowed to process the special personal data category of health data for the purpose of surveillance. This is also the case for the other tasks, which are assigned by the Dutch Minister of Health.

2.2.4 Medical treatment contracts act

This law is part of the Dutch Civil Code, Book 7, Part 5. Dutch name: *Wet geneeskundige behandelingsovereenkomst* (WGBO) (Burgelijk Wetboek Boek 7, 1994).

Article 457 generally forbids healthcare workers to provide patient data to others. The health professional can only breach this obligation when one of the four following conditions apply: the patient gave their consent, a statutory regulation obliges this, there is a conflict of duties or other, compelling reasons of interest.

Article 458 allows patient data to be provided to others for the purposes of statistics or scientific research for public health, as an exception to article 457. This can only be allowed if specific conditions are met.

2.2.5 Statute of the Kingdom of the Netherlands

Dutch name: *Statuut voor het Koninkrijk der Nederlanden* (Staten-Generaal, 2017).

Article 36. The Netherlands, Aruba, Curaçao and Sint Maarten aid each other.

Article 38(1). There is a reference to a mutual arrangement that supports cooperation in the field of implementation of the International Health Regulation (IHR). Dutch name: *Onderlinge regeling samenwerking implementatie Internationale Gezondheidsregeling Nederland, Aruba, Curaçao en Sint Maarten*.

2.3 Legal categorisation of public health data

2.3.1 Purposes for data collection

Data collection for public health purposes may consist of both personal and non-personal data. With respect to the latter, Article 5 of the GDPR imposes that processing of personal data must, among others, be for a limited purpose. It is, therefore, crucial to define the purposes for processing of data for public health and/or by a public health institute. Table 1 lists these public health purposes for RIVM in the Netherlands,

adapted from an assessment of EU Member States (Johan Hansen, 2021).

Table 1 Purposes for personal data collection by RIVM in the context of public health.

Name	Description	Legal basis
Patient care	RIVM has a clinical diagnostic lab, performing tests for patient care in some cases, whose results are normally usable for surveillance as well.	These services are based on informed consent and are covered primarily by the Dutch Medical treatment contracts act (2.2.4).
Surveillance	RIVM is a national health institution with legal competences in public health surveillance whereby it may carry out scientific studies without the consent of the data subject in situations of exceptional relevance and seriousness for public health. This may also consist of non-personal data	This legal competence in public health surveillance and monitoring is based in the GDPR in conjunction with National legislation (sections 2.2.2 and 2.2.3)
Research	Data processing for scientific or historical research by both academic (not-for-profit) and commercial organisations.	Article 9(1) of the GDPR notes that, in general, processing of data concerning health or genetic data shall be prohibited. Article 9(2)(a) provides that this prohibition will not apply if the data subject has given explicit consent, unless Member State law states that the prohibition in article 9(1) GDPR cannot be lifted by explicit consent

2.3.2 Data classes

A fundamental legal distinction between different types of data relevant for public health is that between personal and non-personal data. Personal data is defined in the GDPR as "*any information related to an identified or identifiable natural person*". Within personal data, special categories of sensitive personal data are defined, among which health data, ethnic or racial origin, genetic data and sexual orientation are relevant for public health (see section 2.1.1).

It is critical to describe the various categories of personal as well as non-personal data from a legal perspective since this determines if and how the data can be processed. There is also a potential grey zone between these two main categories, which arises from the fact that

personal data includes data through which an individual can be identified indirectly by a process known as statistical disclosure.

In Table 2, we describe these different legal categories of personal data as we see relevant for public health. The potential grey zone between personal and non-personal data is also addressed there, using the concepts of anonymisation, pseudonymisation and statistical disclosure as described in subsequent sections.

Table 2 Legal data classes.

Name	Examples relevant for public health
Regular personal data	
Directly identifiable data	Name and surname, date of birth, gender, sufficiently complete postal code, home address, national identification number or any other identifier of general application, email address, phone number.
Pseudonymised identifiers	Any identifier that is not a national identification number or an identifier of general application and that requires access to data either controlled by a third party to retrieve additional personal data, such as a human case or sample identifier.
Special categories of personal data	
Health data	Test results, diagnoses (including disease), injury, disability, clinical treatment, determinants influencing health status, information collected when registering for health services or accessing treatment, data on healthcare provision, risk behaviour.
Ethnic or racial origin	Race, ethnicity, membership of a minority population.
Genetic data	Complete human genome or partial human genome that is statistically unique for one individual.
Sexual orientation	Transmission mode if indicating sexual orientation.
Non-personal data	
Non-aggregated anonymous data	Data on single individuals, excluding directly identifiable data and pseudonymised identifiers, and containing statistically insufficient information for identification through statistical disclosure. Non-human origin samples, e.g. from food, animal feed, animal and the environment Sewage samples insofar human genetic material or data have been statistically removed.
Aggregated anonymous data	Data on groups of individuals, by definition excluding directly identifiable data and pseudonymised identifiers, and containing per group statistically insufficient information for identification through statistical disclosure.

Name	Examples relevant for public health
Microorganism data	Phenotypic or genotypic data on any microorganism, including those obtained from – but no longer linked to – human-origin samples, provided that they are statistically free of human genetic material or data.
Economically sensitive data	Human-origin samples directly linked to a specific clinical diagnostic laboratory or healthcare organisation. Non-human-origin samples positive for foodborne pathogens directly linked to a specific food or food business operator.

2.3.3 Pseudonymisation

As defined in the GDPR, article 4(5), "*pseudonymisation means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person*".

Pseudonymisation is normally achieved by assigning another identifier instead of directly identifiable information. These identifiers themselves are still considered personal data. While pseudonymised data is essential for public health for receiving (updated) data, the results derived from this data are normally disclosed in aggregated form or at least with these identifiers removed. Provided that these results do not allow identification through statistical disclosure, they are anonymous and therefore not personal data.

2.3.4 Statistical disclosure

Data that no longer contains directly identifiable data elements or pseudonymised identifiers may still enable identification when the data elements, potentially combined with other available information, statistically reduce the number of people that could match that description to two or one. For public health, such statistical identification may take the form of, for instance, a single person having a (rare) disease AND living in a particular area. Hereby, it should be assumed that there will be some people, such as relatives or friends, for whom this is sufficient information for identification. In addition, if two persons match the description, then one of them may be able to identify the other.

In itself, statistical identification of a person based on a number of data elements is not necessarily an issue, since the data used to identify the person is already known to the person performing the identification. However, if additional data elements are provided that contain, for instance, health information or information about sexual preference, that information would be disclosed. In addition, people could make inferences – justified or otherwise – about, for instance, the person's sexual preference based on known or perceived associations with a disease.

An example here would be that if a man with monkeypox could be identified statistically on the basis of disclosed data (not including the fact that he has or has had monkeypox), then the people or general public to whom his identity is subsequently disclosed could assume that he has a sexual preference for men, even though the actual transmission mode is through close person-to-person contact in general. This in turn may negatively affect that person.

Statistical disclosure may therefore infringe on the rights of the data subject. The risk (probability and impact) of disclosure also needs to be considered and weighed against the public benefit of releasing the corresponding results. Techniques to guard against statistical disclosure include not showing results for small numbers, combining groups, showing a bracket instead of an actual value and rounding values (Emily Griffiths, 2019). Another possible technique is differential privacy, which relies on adding random noise to results (Ficek, 2021).

2.3.5 Anonymisation

Personal data can be transformed into non-personal data through anonymisation. Non-aggregated data can be anonymised by removing directly identifiable data, pseudonymised identifiers and data that together are usable for statistical disclosure.

Aggregating data into groups of persons with the same characteristics implies by definition removal of directly identifiable data and pseudonymised identifiers. However, this does not necessarily mean that the data is anonymous since some groups may consist of only one or very few persons, for whom statistical disclosure may still be possible.

2.4 Linking samples and epidemiological cases

Anonymisation of the human infectious disease case-based (epidemiological) surveillance and sample-based lab surveillance data provided to the public health agency, as described further in section 4.1.2, is usually very detrimental to the effectiveness and efficiency of public health actions. This is because the case data, which normally contains additional epidemiological and clinical data, can then not be linked to the sample data, or only partially linked and with risk of errors. For example, the assessment of vaccine effectiveness or disease severity associated with different variants of the same virus species, can then no longer be assessed. In addition, the data subject may need to be identified, for instance in case an outbreak is detected, or more or corrected information would have to be provided to properly assess public health risk.

Even if both case-based surveillance and sample-based lab surveillance are pseudonymised, the same pseudonymised identifier would need to be shared in both data flows to enable linking them. This can represent another significant practical challenge. A potential solution there is to share the national identifier with the public health agency instead since this identifier is likely to be available to both submitters of data. However, the use of the national identifier for this purpose may be

prohibited by national law, as is the case in the Netherlands (see sections 2.2.1, 2.2.2).

Finally, without a pseudonymised or national identifier for the public health agency to communicate to the actual holder of directly identifiable information, identification of cases – by this holder, not necessarily by the public health agency – and linking case and sample data is possible only occasionally on the basis of statistical matching. This can, for instance, take the form of matching on age, postal code and disease onset with sample date. This process can be cumbersome and prone to delays that substantially reduce or even render ineffective subsequent public health actions, including for foodborne outbreaks. The ability to occasionally identify cases statistically is also contradictory in a legal sense, since it implies that the data provided to the public health agency was not fully anonymised in the first place but partially prone to statistical disclosure.

Unless national law specifically allows sharing national identification numbers, which it does not allow in the Netherlands, pseudonymisation rather than anonymisation of human case-based and sample-based data provided to the national public health institute is a critical means for maintaining an efficient and effective public health system. At the same time, it keeps the risk to data subjects regarding disclosure of their personal data low, since directly identifiable data is still not stored at the national public health institute for this purpose. Further conditions for this risk to remain low are that the data is stored in compliance with the GDPR and data protection-by-design in particular, and that statistical disclosure is guarded against.

3 Current situation

Currently, most pathogens which pose a threat to public health are under lab surveillance in the Netherlands. The current situation in terms of processes, systems and resources has developed over a long period of time. In order to put in place a lab surveillance platform that takes into account lessons learned from the past as well as the COVID-19 pandemic, we captured and analysed this information on the current situation in a consistent way, as a key input to make a successful transition towards the future. The latter is further described in Chapter 4.

3.1 Methodology

We used the Business Process Analysis (BPA) methodology to capture information about the current processes (Weilkiens, 2016). Three steps of this process were executed, mostly consecutively: (i) discovering the as-is process, (ii) documenting the as-is process and (iii) analysing the as-is process.

For the discovery step of the as-is process, a high-level process map was first created on the basis of introductory talks with subject-matter experts (SMEs) within RIVM's Centre for Infection Research, Diagnostics and Laboratory Surveillance (RIVM-IDS, Figure 1). This also outlines the scope of the processes under consideration and gives a good general representation of these processes. In particular, the processes performed within RIVM's Centre for Epidemiology and Surveillance of Infectious Diseases (EPI), which combine lab surveillance data with epidemiological case-based data from other sources, were not covered as part of this exercise.

For the documentation step of the as-is process, several one-on-one interviews were held with each SME, to eventually capture the entire end-to-end lab surveillance process for the respective pathogen or group of pathogens in the standard Business Process Model and Notation (BPMN) notation, version 2.0.2 (Weilkiens, 2016). A generic example of such a process is given in Figure 2. In addition to capturing each of these processes, both the incoming information flows from external parties to RIVM, as well as the outgoing information flows from RIVM to external parties, were entered in a spreadsheet and example files were collected. The purpose of this was to gain further insights into the actual data collected, both from a technical data modelling perspective and from a legal perspective, as an input for the future situation. The metadata elements captured to describe data provided to RIVM are described in Table 3.

Lastly, interviews were held with MMLs, who provide the majority of the data and samples to RIVM. These interviews covered both the current situation, including what works well, what works less well, and what they would expect from the future. These outcomes are directly integrated in the requirements for the future situation, see also sections 4.1.2 and 4.3.2.

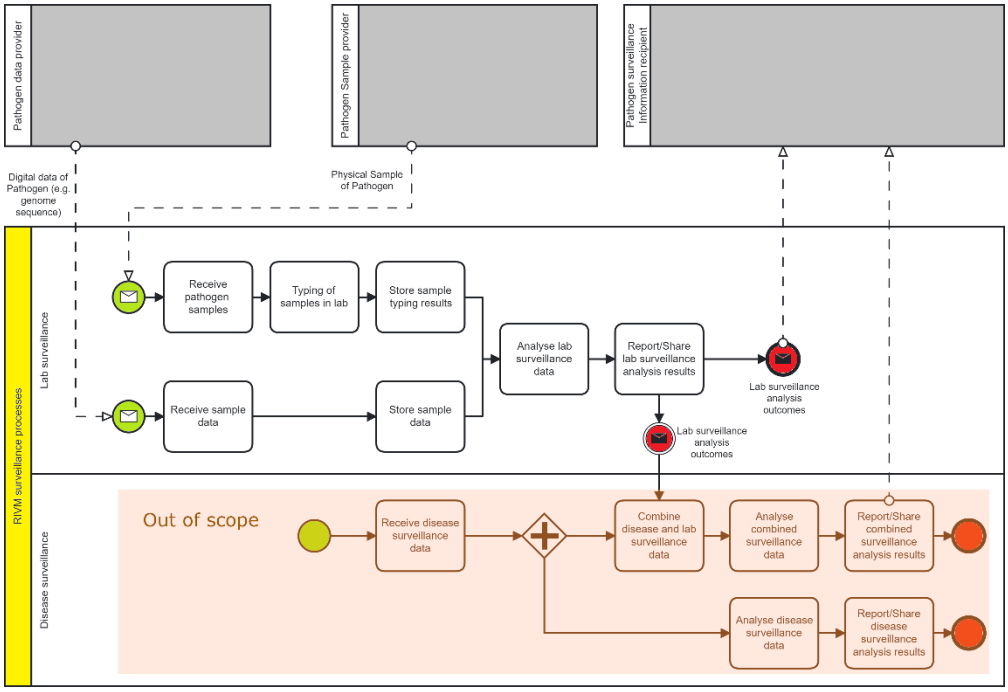


Figure 1 High-level end-to-end lab surveillance process and broader surveillance context.

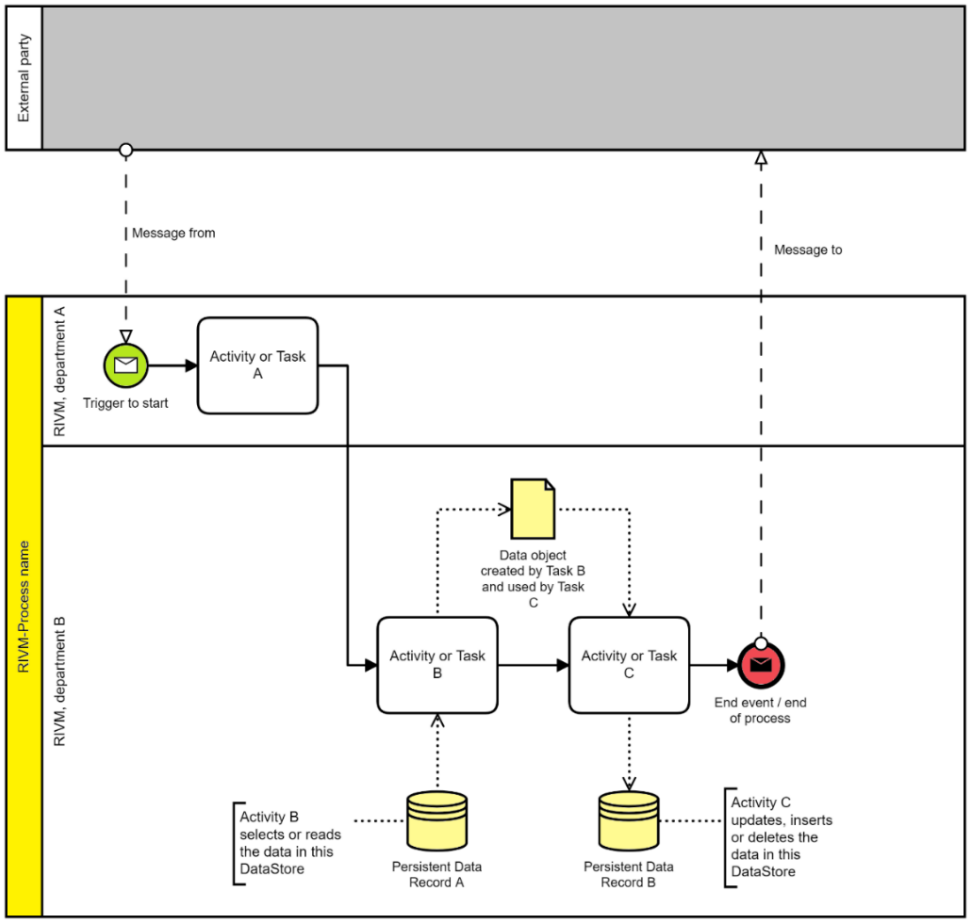


Figure 2 Example of a process in BPMN notation.

Table 3 Metadata elements captured to describe data provided to RIVM.

Metadata element	Description
ID	[#] Identifier in incoming data stream
Description	[text field] Description of incoming data stream
Legal purpose of data processing	[multi select] Legal purpose (e.g. Surveillance of notifiable diseases)
Pathogen(s)	[text field] Which pathogen the data stream is about
Type of data provider	[single select] Type of legal entity providing data
Legal data classification	[multi select] GDPR data classification (e.g. Personal Data – Pseudonymised identifiers, Special personal data – Genetic data, etc.)
Technical data elements	[text field] Attributes of incoming data stream (e.g. Postal code, sex, age, sampling date, sample ID, etc.)
Legal basis	[text field] Reference/explanation of legal basis on which the incoming data can be legally processed.
Comment	[text field] Comments/clarifications

Metadata element	Description
Data subject selection criteria	[multi select] Rationale of how the data is selected to be sent to RIVM (e.g. random positive sample)
Scope of data sample population	[text field] Scope of received data sample (e.g. by type of disease, age group, etc.)
Applied security measure for personal data protection?	[text field] Measures taken to secure the (privacy of) data (e.g. pseudonymisation)
If there are applied security measures, please specify on which 'Technical data elements'	[text field] Specify to which *Technical data element* the security measures apply

3.2 Results

This section represents the analysis step of the as-is process, i.e. the third and last step of the applied BPA methodology. An overview of the pathogens for which SMEs were interviewed to capture the current processes is given in Table 4. After conducting 24 process discovery sessions with 10 SMEs, 73 end-to-end processes were drawn up. In addition, 87 data files were collected that were used within the processes to register, analyse or share surveillance data. The stakeholders involved in those end-to-end processes are listed in Table 5.

Suggestions for improvements by SMEs are included as well, which are also an input for the requirements for the future situation (Chapter 4). The sections below follow the same structure as the process discovery sessions with the SMEs:

- What samples are received from whom
- What analyses are performed and reported on individual samples
- Which surveillance analyses are performed and with whom results are shared

Table 4 Overview of pathogen for which SMEs have been interviewed.

Disease	Pathogen
COVID-19	SARS-CoV-2
CPE infection	Carbapenemase-producing Enterobacteriaceae (CPE)
Enteroviral disease	Enteroviruses
Hepatitis B	Hepatitis B virus
Influenza infection	Influenza viruses
Measles	Measles virus
Meningococcal disease	*Neisseria meningitidis*
MRSA infection	Methicillin-resistant *Staphylococcus aureus* (MRSA)
Mumps	Mumps virus
Pertussis	*Bordetella pertussis*
Pneumococcal disease	*Streptococcus pneumoniae*

Disease	Pathogen
Rubella	Rubella virus
Salmonellosis	*Salmonella* spp.
Tuberculosis	*Mycobacterium tuberculosis*

Table 5 Overview of stakeholders.

Stakeholder	Description
RIVM	Dutch National Institute for Public Health and the Environment
CIb	Centre for Infectious Disease Control. Part of RIVM.
LCI	National Coordination Centre for Communicable Disease Control. Part of CIb.
EPI	Centre for Infectious Diseases, Epidemiology and Surveillance. Part of CIb.
IDS	Centre for Infectious Diseases Research, Diagnostics and Laboratory Surveillance. Part of CIb.
MMLs	Medical Microbiology Laboratories. Conduct research on micro-organisms that can cause infections in humans
Healthcare Institutes	Hospitals, General Practitioners
VWS	Dutch Ministry of Health, Welfare and Sport
WHO	World Health Organisation
ECDC	European Centre for Disease Prevention and Control

3.2.1 Receipt and processing of samples

RIVM-IDS receives mainly human-origin samples, for the purposes of infectious disease diagnostics and public health surveillance. It is equipped with bacteriology, virology and parasitology laboratories that perform, among others, detection, sequencing and relevant phenotypic testing, such as for antimicrobial susceptibility.

Agreements have been made with multiple healthcare partners and laboratories to share samples that are positive for particular pathogens, to establish, for instance, a sentinel lab surveillance system based on random sampling for that pathogen. In some cases, aggregate numbers of samples with negative detection results for a particular pathogen are shared as well. Excluding SARS-CoV-2, approximately 70% of the samples received at RIVM-IDS are sent in for diagnostics. The majority of samples RIVM receives are sent in by MMLs. For SARS-CoV-2, a substantial subset of the samples are sequenced by the MML itself and do not require further typing at RIVM. In these cases, only data including the finished consensus genome is sent to RIVM instead.

For most pathogens, a paper request form for diagnostics to be performed accompanies each received sample, containing information on the requesting party, the requested test(s) and some brief clinical and epidemiological information about the patient. Information from the request form is entered into RIVM's Laboratory Information Management System (LIMS). The paper form itself is scanned and stored. SARS-CoV-2, MRSA and CPE are exceptions and do not have a paper sample request form. Instead, the information is supplied digitally to RIVM

through respectively a filled in spreadsheet template supplied by email and a web interface where each individual sample is entered.

SMEs at RIVM have suggested the following improvements:

- It needs to be clear in advance to both parties, i.e. RIVM and the lab surveillance sample provider (typically an MML), what is expected in terms of sample and accompanying data. This includes the legal aspects, especially regarding data privacy, the time investment and what can be expected in terms of feedback on test results.
- Request forms for performing diagnostics are mostly in a paper format, the content of which has to be transferred manually into the LIMS. This is prone to errors compared to digital receipt of the data.

3.2.2 Analysis and reporting on individual samples

The received samples are further prepared for (diagnostic) testing, typically through steps such as extraction and culturing. After this, the actual tests can be performed in the RIVM-IDS lab. Test results, in particular those with a bioinformatics component, can also be generated by separate software, such as BioNumerics. For the processes that were considered here, (diagnostic) test results were then usually saved again in the LIMS.

Multiple persons validate the test results several times at different steps of the process. A SME typically approves the final test result. They, or a bioinformatics pipeline, may also further interpret the result to produce derived results, such as a classification as a particular variant or as being resistant, or not, to a particular antimicrobial agent.

The final approved test result(s) are communicated back to the requesting party. In the case of Type-Ned MRSA and CPE, there is an automated upload function in BioNumerics to upload a selection of test results to Type-Ned, where they can be viewed by the requesting party. For other pathogens, communication takes place via email.

3.2.3 Reporting and sharing surveillance data

Different reports are produced to inform the public domain as part of lab surveillance. These publications are normally based on aggregated data. Typically, surveillance reports are published with a given frequency such as annually or weekly, and they are centered around a specific pathogen or pathogen group, such as respiratory tract infections or food-borne diseases.

Analyses may include the number of positive samples, the proportion of samples with antimicrobial resistance, and trends and stratifications by subtype, age, gender and place. For most surveillance reports, the lab surveillance data is requested by RIVM-EPI, the epidemiology department, and by RIVM-LCI, the infectious disease coordination department. They request information on positive samples and circulating variants from RIVM-IDS SMEs, such as virologists and bacteriologists, and from MMLs. Reports are usually co-written by experts from the different centres. Typical publishing channels are RIVM, ECDC and WHO.

RIVM's LIMS is the primary data source for the lab surveillance results. In addition, samples may already have been fully tested by the MML or they may come from the public domain. This data is then stored separately. Independent from that, derived data may also need to be taken from additional databases, such as BioNumerics. In addition, it is highly desirable, if not necessary, for surveillance reports to be able to link the sample data from lab surveillance with the case-based data from epidemiological surveillance. The latter is typically stored in the Osiris database at RIVM-EPI for notifiable diseases. However, at present, this matching is frequently not possible or only partially possible and error-prone based on statistical matching.

SMEs at RIVM have suggested the following improvements:

- For some pathogens, manual pseudonymisation and anonymisation of gathered data is needed for reporting. Maybe this can be automated at the source with a dedicated export function.
- It should be easier to share a set of test results directly with MHSs and MMLs.
- Often, the total number of detection tests performed, i.e. including those with negative result, is not known. External parties should also provide RIVM with this denominator.
- Currently, separate software is used to create overviews (species, date of collection) from the test findings, resulting in exports in the form of excel files containing the combined data from different databases. This leads to switching tasks and manual actions, such as formatting, which should be avoided.
- When exporting data from different sources, there is a risk of using outdated data. Sample test results may be rectified by additional testing/judgement, and this rectification may not have been included in the export. There should be only one source to extract from.
- Case data from Osiris and pathogen data from LIMS/BioNumerics are manually matched and combined in spreadsheets. This should ideally be automated to a large extent.

3.3 General findings

Fragmentation. In many interviews, the phrase "*that is pathogen-specific*" was used. The meaning of this phrase is often threefold:

- Separate arrangements/agreements with healthcare partners are made per pathogen or group of pathogens
- Besides a core of common variables that are generally provided, there are some variables that are only provided for a subset or even for one pathogen
- Lab tests and further analyses are to some degree pathogen-specific

Spreadsheets. Because there are multiple external sources that are relevant to lab surveillance, employees from multiple departments make use of (Excel) spreadsheets to bring together information and keep track of information through the various databases. These spreadsheets contain, among others, the IDs of the corresponding database(s) for the notified case, the sample in the LIMS, and external databases, such as

ECDC’s TESSy database. Because these overviews are basically multiple ‘snapshots’ of the sources, there is a configuration management risk of data being not up-to-date or out of sync with the original sources. In addition, traceability of results is impacted and the amount of manual work high.

4 Future situation

Public health surveillance systems, and in particular laboratory surveillance systems, are often created with just a few or even one disease or health issue in mind, such as antimicrobial resistance. From there, they often grow organically, increasing in complexity and maintenance requirements.

Some of the issues that may occur over time are, for instance, insufficient data comparability due to insufficient initial attention to choosing appropriate variables and allowed values, and evolving into a complex system of multiple data stores including files and ad-hoc processes for analyses that very few if any people still understand well. In addition, surveillance systems usually have a lot of stakeholders with different needs and capabilities, including those that provide data. Finally, since the GDPR came into effect in 2018, the legal aspects of public health surveillance systems have arguably become fundamentally more complex, as described in Chapter 2.

Here, we attempt to address a major part of this challenge, having the freedom of defining a new and generic laboratory-based surveillance platform (lab surveillance platform) and surrounding processes entirely from scratch. We start by describing a few fundamental principles that we apply throughout, and then go through a defined stepwise process of translating and refining its requirements to eventually arrive at an initial set of requirements that can be used for initial design and implementation.

4.1 Fundamental principles

The fundamental principles that apply to setting up the platform are as follows:

- The purpose is clearly defined.
- The collaboration between the involved parties is good.
- The platform and processes comply with applicable laws, rules and regulations.
- The accuracy, precision and timeliness of the data stored, processed and reported are known and score sufficiently high to be able to make relevant decisions that positively impact public health.
- Standardisation is applied where possible.
- Different concerns are separated within the platform by design.

In the following sections, each of these principles is addressed in detail.

4.1.1 Purpose

In principle, good systems must have a clear purpose. For this lab surveillance platform, expert opinion was sought both internally at RIVM as well as externally at MMLs and MHSs to define the various goals. This was effected first through individual sessions with twenty MMLs and a session with MHSs, which resulted in an initial formulation of the goals of the system as part of a broader discussion paper that also contained

points on legal basis, planning and improving collaboration. Subsequently, the goals – and the entire paper – were first reviewed internally at RIVM's IDS centre by fourteen experts, then by the centre's management team and finally by experts from the EPI and LCI centres. Finally, they were discussed in a joint session organised on 2022-09-15 by the IDS centre with all MMLs and MHSs invited.

The goals of the lab surveillance platform, as formulated after these rounds of discussions and review, are as follows:

1) The purposes of the lab surveillance platform are signal detection, including trend analysis, quality monitoring and supporting signal follow-up. Unless agreed otherwise or as required by law, RIVM restricts signal detection to national or supra-regional signals (with regional referring to MHS regions). The platform can also facilitate local signal detection between MHSs and MMLs, since both would have access to the corresponding data.
2) In principle, signal detection only requires the data that is typically available in the LIMS of the MML, plus the possible additional typing data generated by RIVM when a sample is sent to RIVM or when a sequence is sent and further typed through bioinformatics analysis. Negative typing data (detection data) may also be required for signal detection, e.g. for calculation of proportion positive test results. Signal detection can be based on other surveillance data sources as well. However, these fall outside of the scope of the lab surveillance platform.
3) In this case, quality monitoring covers, among others, the detection of reduced clinical sensitivity of PCR tests due to natural evolution of the pathogen, and detection of significant differences in average antimicrobial resistance between laboratories, possibly due to quality issues. While the platform is capable of such quality monitoring, its desirability and the exact manner of implementation would need to be determined case by case.
4) Support of signal follow-up is a goal of the lab surveillance platform, but signal follow-up also requires additional clinical and/or epidemiological data that are normally collected by the MHSs and then forwarded to RIVM. Legally, the responsibility for signal follow-up of notifiable non-category-A diseases lies in first instance with the MHSs. RIVM may also play a role here, as established in the law on public health (see section 2.2.2).
    a. There are some exceptions here, such as local signals of antimicrobial resistance, where follow-up may fall within the responsibility of a hospital's infection prevention department. In some cases, RIVM also supports this process.
    b. In any case, the data stored by the lab surveillance platform need to remain available for signal follow-up. To do so efficiently, the MHSs can receive separate access rights, possibly including the ability to link within the platform a sample to the epidemiological case that they collect. In itself, however, this does not solve the problem of linking the data, which would have to be done manually by the MHS. This would probably only occur in case of outbreaks.

   c. For non-notifiable diseases, for which the MHSs do not send epidemiological data to RIVM, the lab surveillance platform can – if needed – facilitate signal follow-up by also storing these additionally required data.
5) While scientific research is a structural part of the tasks of RIVM and MMLs that are part of a university medical centre, it is not a purpose of the lab surveillance platform per se. Just as for surveillance research that is broader than lab surveillance, scientific research with relevance for public health that uses typing data, often requires additional clinical or epidemiological data, such as symptoms, vaccination status, severity of the disease and/or more elaborate phenotypic tests. The platform can, however, facilitate scientific research as follows:
   a. The lab surveillance platform can facilitate initiation of scientific research because interesting samples as well as the parties that have them or sent them to RIVM can be identified within it.
   b. Where relevant and feasible, the lab surveillance platform can facilitate the execution of scientific research by storing the additionally required data for the relevant samples. That makes it possible to work with a central dataset, which can be extracted from the platform when needed, to be analysed by the parties conducting the research. In this case, the collection of these additional data would be on a different legal basis than data collection for signal detection, and require, for instance, informed consent. Technically, this difference can be implemented in the platform, e.g. with access rights and the concept of a 'data collection' or, more generally, a set of samples.

Apart from this, the lab surveillance platform itself is part of a broader effort within RIVM-CIb to develop an integrated surveillance environment for infectious diseases. That includes the case-based epidemiological surveillance through the MHSs, and the linking between those cases where possible. In addition, it includes (linking to) surveillance of food, animal feed, animals and the environment within the One Health principle. The One Health aspect of lab surveillance is, however, kept out of scope from this report, because it still needs further discussion with a broader set of stakeholders and including it with currently available resources would delay progress significantly.

4.1.2 Collaboration

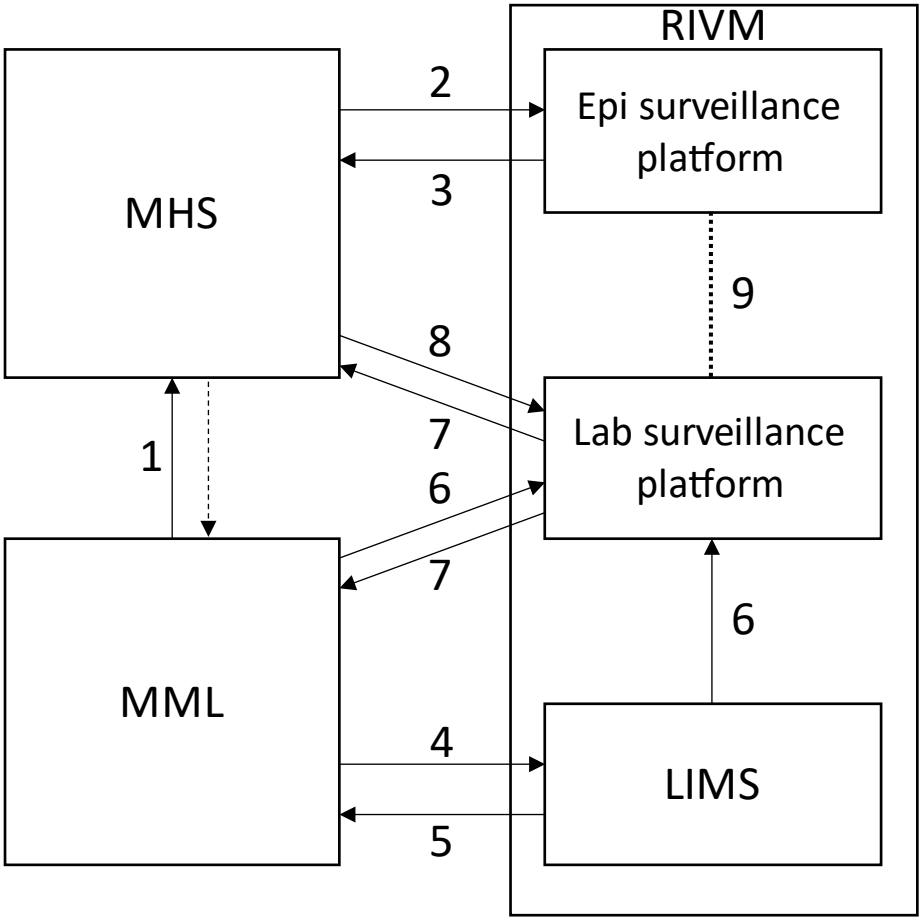
Good collaboration within the MML-MHS-RIVM triangle is critical to realising the goals of the lab surveillance platform (section 4.1.1). This collaboration has existed for a long time and is based both on legal requirements (see sections 2.2.2 and 2.2.3 in particular) and on a general willingness to contribute to public health. Nonetheless, there are difficult issues in this relationship, such as financial compensation for MMLs for this aspect, the authorship of any scientific papers arising from the collaboration and exchanged data, and the role that some MMLs see for themselves in public health surveillance relative to that of MHSs and RIVM, while this is not established by law.

In order to extend the existing collaborations to the new lab surveillance platform, RIVM initiated a dialogue with MMLs and MHSs, also involving the Dutch Society for Medical Microbiology (NVMM, a professional association for individuals) and the Dutch association of medical microbiology laboratories (VMML, a sector association for the laboratories as organisations). So far, this dialogue has taken place on four major occasions: a kick-off meeting in February 2022 to which all MMLs and MHSs were invited, sessions with a total of twenty individual MMLs, a group session with several MHSs all in the period of April-May 2022, and a group session in September 2022 to which all MMLs and MHSs were invited.

The sessions with individual MMLs and the group session with some MHSs were used to generate a Discussion Paper to prepare for the September 2022 meeting, as described earlier in section 4.1.1. In addition, they were used to extract as many external user requirements for the platform as possible. The result of this will need to be refined over time. Further use of these requirements is described in section 4.3. The main dataflows between the triangle of MMLs, MHSs and RIVM, including the future lab surveillance platform, are described in Figure 3.

A crucial instrument to sustain and improve the collaboration around the new platform is its Terms of Use. The suggested key points of these Terms of Use were put forward in the Discussion Paper and subsequently discussed during the September 2022 meeting. These particularly include the areas where RIVM can make decisions on its own, the areas where decisions are to be made together with MMLs and MHSs, and how scientific authorship and publication of data in general are handled. The parties concerned will need to jointly set them out in more detail to achieve a complete and acceptable Terms of Use (ToU) document.

One of the key points in the ToU is the establishment of a commission for joint decision-making. Among the tasks of the commission is the updating and creation of new specifications for sharing the data submitted to the platform between its members. This is crucial to making joint decisions, since it specifies how other parties can see and potentially make use of one’s data, but rarely affects the core of the platform.



Dataflows and interactions:

1) Notifying cases of notifiable diseases.
2) Entering cases of notifiable diseases.
3) Viewing user-accessible results on cases of notifiable diseases.
4) Sending physical samples for diagnostics or typing, accompanied by relevant additional epidemiological or clinical data.
5) Reporting of diagnostics and typing results on individual samples.
6) Submission of typing data and relevant additional epidemiological or clinical data.
7) Viewing user-accessible results on (aggregated) typing and relevant epidemiological and clinical data.
8) Possible future option: submission of epi surveillance platform case identifiers to enable linkage.
9) Linking of epidemiological cases to samples through shared sample identifier, shared case identifier or shared civil service number (BSN) if permitted in the future.

Figure 3 Interactions with the future lab surveillance platform. Not mentioned: other parties that submit data than MMLs (e.g. general practitioners, healthcare professionals), non-human origin samples (e.g. from national food safety authority), public samples and typing data, use of the data by different departments within RIVM and NL reference laboratories, further submission to public databases and EU/WHO, samples submitted directly from MHS to RIVM.

A final aspect of collaboration is contributing to international public health. Apart from the collaborations that RIVM already has in this area, one way of contributing is by making relevant software components available as open-source code. This allows other organisations to avoid investing similar amounts of resources into similar functionality, whenever the source code in question is not specific for RIVM and of good quality. Therefore, where possible, such components should be developed as open source. In addition, the development as open source also generally increases the quality level since the code is visible to all.

4.1.3 Legal compliance

The legal basis and framework for lab surveillance in the Netherlands is described in Chapter 2. Here, we focus on the provisions of the applicable law that have a substantial and specific impact on the platform.

Lab surveillance data is largely based on human-origin samples, which, in order for public health actions to remain effective and efficient, should be provided to RIVM as pseudonymised rather than anonymised data (2.3.2). As such, the platform will contain personal data and the EU GDPR (2.1.1) and NL Implementing Act GDPR (2.2.1) are applicable to this data.

In addition, the platform will be used to collect data on different pathogens that will have a somewhat different legal status in the Netherlands, such as being a category A, B1, B2 or C pathogen as defined in the NL Public health act (2.2.2). The platform may also facilitate scientific research through collection of additional data under a different legal basis. As such, it must be able to distinguish, for each record and variable that contains personal data, on which legal basis it was collected. Table 6 lists all the articles in law that are likely to directly impact the design of the platform, and as such should be traceable in requirements.

4.1.4 Accuracy and precision

Effective and efficient public health actions require data that is sufficiently accurate, precise and timely. For the definition and difference between accuracy and precision, or equivalently bias and variability as used in statistics, see e.g. (Rothman & Lash, 2008)

In short, accuracy is the degree of systematic errors for continuous variables. Many variables in public health are categorical though, including those for lab surveillance. The latter include the laboratory test used, some laboratory test results and especially further interpreted results, such as antimicrobial resistance or genotypic classifications. Accuracy also applies here, in the sense that the definition and use of the values should not introduce systematic errors. Finally, accuracy of lab surveillance is affected by the selection of which samples to include, i.e. the statistical sampling method.

Precision is the degree of random errors for continuous variables. For time points or periods, not considering random errors on the measurement itself, it is the resolution at which the data is available, e.g. year only, month as well, or full date. For location, not considering

random errors on the measurement itself, it is usually an administratively defined entity that determines the resolution, such as postal code, municipality, province, or country. For categorical variables, precision arises from the granularity of the categories as well as from the possibility to also use sets of categories as valid values, e.g. for laboratory tests, antimicrobials or diagnoses. A final form of (loss of) precision is when only observations that are already aggregated by one or more variables are available for analysis, such that, for instance, dates are reduced to weeks or months, or age in years is reduced to age categories spanning several years.

The variability in precision of data, e.g. by time, performing or submitting laboratory and pathogen, must be taken as a given, together with the possibility that the variables collected for the same pathogen may change over time. The platform must be able to handle this variability by design, so that it does not decay over time into, for instance, a collection of different implementations per disease. This requires substantial up-front design efforts, including the choice of appropriate data standards (see section 4.1.6). To further clarify the concept of precision, as well as how varying degrees of precision can be accompanied within the same table, some examples are provided in Table 7.

The extraction and transmission of lab surveillance data from laboratory to public health institute, as well as its storage there, may lead to the degradation of accuracy and precision. This can be due to various factors, e.g. errors in the extraction, or conversion to a format for transmission, or subsequently during the conversion from the format of transmission to the storage format. Manual conversion is particularly prone to random errors affecting precision, while automated extractions or conversions are prone to systematic errors affecting accuracy.

The detection of issues with both accuracy and precision, and any subsequent preventive or corrective actions, are critical to ensuring the quality of the data, and thus the reliability of any conclusions derived from them. Manual preventive actions can take the form of, for instance, clear requirements for submission and thorough testing of data transformations. Automated preventive actions are typically in the form of validation rules that are applied when data is transmitted from laboratory to public health institute and may prevent submission of data that is of insufficient quality. However, there is also a large grey zone of potential errors that may nonetheless be correct, or that can only be detected after submission of additional data. It must therefore be possible to apply validation rules both during the submission of data and following its storage. In both cases, there must also be an effective mechanism in place to notify the relevant user so that preventive or corrective actions can be undertaken if necessary.

4.1.5 Timeliness

Timeliness is the processing time between recording the measurement and its actual availability – usually in further derived form as part of an analysis – for decision making. Each step in between these starting and end point adds to this processing time. The lab surveillance platform

covers many of these processing steps, and their individual or combined maximum processing times should be part of performance requirements.

The timeliness of the steps before submission to the lab surveillance platform is just as important but cannot be addressed by the platform. The corresponding frequency of data submission should be agreed upon in advance. An exception to that can be that issues with the actual submission to the platform cause delays, which can be due to the quality, user friendliness or availability (i.e. uptime) of the interfaces for data submission. At the same time, the combined processing time of these steps before submission can normally be recorded in the platform, since the sampling date is normally part of the submitted data and the date of submission is straightforward to record.

Another type of timeliness is the speed at which the relevant components can be adapted to new or extended needs, e.g. for a new disease or more enhanced surveillance. This is a function of the overall system and component design as well as of the change management processes. The latter can be complex and slow, involving several organisational units, which should be avoided where possible. This is especially the case during a crisis situation, such as a major outbreak, a pandemic or a newly emerging pathogen.

Concretely, this means that it must be possible to add new diseases and pathogens, as well as typical enhancements to the surveillance without change to at least the data storage component. Similarly, a decoupling of how data can be submitted from how they are stored is required, so that evolving needs in how to submit data can be met without necessarily changing the data storage component.

4.1.6 Standardisation

The use of data standards where possible is considered as critical to quality for any public health surveillance system, including this lab surveillance system. They should, therefore, be used on principle. Using appropriate standards means making use of the usually large amount of experience that has gone into crafting a standard, in a way that is more generally applicable than only the current needs foreseen by the organisation. It also facilitates communication within and outside the organisation, and may make it easier to train people, find people with appropriate expertise, and make use of existing software. Finally, having the data readily available in a standard format provides a viable exit strategy if the chosen system is no longer able to address the current needs (Kodra, 2018).

4.1.7 Separation of concerns

Separation of concerns, or modularity, is a well-established principle in computer science, whereby distinct individual components each encapsulate different data and functionality accessible through a well-defined interface (Rob Conery, 2009). As such, components can be maintained largely separately from each other, thereby making the overall system more flexible and maintainable.

Two crucial separations of concerns are (i) between how data can be submitted to the platform in various formats and how it is stored in a

single standardised format, and (ii) between this standardised format and derived results that are, among others, provided back to users. Without these, the system has a critical lack of flexibility for maintaining a consistent approach across pathogens: constraints on how and what data can be submitted per pathogen would lead to per-pathogen design changes, and so would requests on what data can be shared per pathogen. The consequence of this is that there must be at least three conceptual components: Data Submission, Data Storage and Data Sharing.

A third important separation of concern is between bioinformatics pipelines for genetic sequence analysis, and the subsequent use of the results (Black, 2020). Three different types of analyses are generally performed on sequences: primary analyses to convert sequence reads into a finished genome, secondary analyses to classify genomes phylogenetically and to predict phenotypic properties, and tertiary analysis to determine (phylogenetic) relatedness within a set of sequences (Gargis, 2015). Typically, only the primary and secondary analyses are performed using an automated pipeline that is also amenable to distributed computing, whereas the tertiary analyses require additional data and are tightly coupled to (interactive) visualisations. The tertiary analyses are therefore fundamentally different from the primary and secondary analyses. Finally, within the tertiary analyses, the uncoupling of calculation of the results – which may, for instance, take specific user access rights into account – from their visualisation is also an important one to make.

Table 6 Articles in law that have a direct impact on the design of the platform.

Law	Article number and short description[1]
EU GDPR	(5) Transparent processing, purpose limitation, data minimisation, accuracy, storage limitation, integrity and confidentiality, accountability (6) Performance of a task carried out in the public interest, consent by the data subject. (9) Special categories of personal data. (15) Right of access by the data subject. (16) Right of the data subject to rectification. (17) Right of the data subject to erasure. (18) Right of the data subject to restriction of processing. (20) Right to data portability. (24) Responsibility of the controller. (25) Data protection by design and by default. (35) Data protection impact assessment.
NL Implementing act GDPR	(25) Restricted processing of the special category of personal data of race or ethnicity. (28) Restricted processing of the special category of personal data of genetic data. (30) Restricted processing of the special category of personal data of health data. (44) Suspension of GDPR data subject rights in specific cases. (46) Restricting processing of the national identification number to cases where this is specifically defined by law.
NL Public health act	(1) Different categories of infectious diseases: A, B1, B2, C. (6) Municipality and safety regions. (14) Municipal health services (MHSs). (19) Category C disease specification. (24) Data requirements for reporting to MHS. (25) MHS request for additional tests. (27) Reporting of Category A diseases to safety region. (28) Reporting of Category A, B1 and B2 diseases to RIVM.

[1]For a more in-depth description of these articles, see Chapter 2.

Table 7 Examples of how different degrees of precision can occur, how they can be put in the same table (per example), and how precision can affect analyses.

Example	Degree of precision and description	Storage in same table of different degrees of precision
Pathogen detection through PCR with 2 targets	**Very low.** Detection implicit, i.e. the data is submitted or stored in a table dedicated to a particular pathogen, containing only overall positive results. This enables, among others, calculation of a crude count.	Single row TEST equal to PCR_PATHOGEN_X TARGET equal to ALL RESULT equal to POSITIVE
	Low. The PCR test and overall result (positive or negative) are recorded. This additionally enables, among others, calculation of a positivity ratio, and depending on the statistical sampling method potentially a decline in clinical sensitivity of the overall result.	Single row TEST equal to PCR_PATHOGEN_X TARGET equal to ALL RESULT either POSITIVE or NEGATIVE
	Medium. The PCR test and result per target (positive or negative) are recorded. Depending on the statistical sampling method, this additionally enables, among others, potential detection of the risk of decline in clinical sensitivity as assessed by dropping out of one of the two targets.	One row per target TEST equal to PCR_PATHOGEN_X TARGET respectively LOCUS1 or LOCUS2 RESULT either POSITIVE or NEGATIVE
	High. The PCR test and result per target (Ct-value) are recorded. Depending on the statistical sampling method, this additionally enables, among others, potential detection of inter-laboratory variability.	One row per target TEST equal to PCR_PATHOGEN_X TARGET respectively LOCUS1 or LOCUS2 RESULT equal to provided Ct-value
Birth date	**Low.** Provided as age in years part of a table with laboratory test results including sampling date. This involves the risk that age in years was calculated from birth date and by rounding to the nearest integer rather than rounding down. This enables calculations by age in years.	BIRTH_MIN_DATE equal to sampling date minus (age + 1) BIRTH_MAX_DATE equal to sampling date minus age
	Medium. Provided as year and month rather than full date for reasons of data protection. This enables calculations by age in months rounded to the nearest integer, which may be relevant for infant cases.	BIRTH_MIN_DATE equal to YEAR/MONTH (partial date) BIRTH_MAX_DATE empty

Example	Degree of precision and description	Storage in same table of different degrees of precision
	High. Provided as a full date. This enables calculations by age in months, weeks or days, which may be relevant for infant cases.	BIRTH_MIN_DATE equal to provided date BIRTH_MAX_DATE empty
Travel prior to or during onset of disease	**Low.** Provided as a single variable ("YES" or "NO") part of a table with laboratory test results including sampling date. Travel occurred implicitly abroad and recently before the sampling date. Recently is interpreted as in the last N weeks.	TRAVEL_TO equal to ABROAD TRAVEL_START_DATE equal to sampling date minus N weeks TRAVEL_END_DATE equal to sampling date
	Medium. Provided as a single variable with a list of destinations part of a table with laboratory test results including sampling date. Travel occurred implicitly recently before the sampling date. Recently is interpreted as in the last N weeks.	One row per destination TRAVEL_TO equal to DESTINATION_X TRAVEL_START_DATE equal to sampling date minus N weeks TRAVEL_END_DATE equal to sampling date
	High. Provided as a separate table with, per row, the actual destination and start and end date.	One row per destination TRAVEL_TO equal to DESTINATION_X TRAVEL_START_DATE equal provided start date TRAVEL_END_DATE equal to provided end date
Vaccination status for COVID-19	**Low.** Provided as a single variable ("YES" or "NO") part of a table with laboratory test results including sampling date that are implicitly positive for SARS-CoV-2. Vaccination is thus with a vaccine that protects against COVID-19 and was carried out on or before the sampling date. This enables, among others, calculation of association between individuals vaccinated at least once and the laboratory test results.	Single row VACCINE equal to COVID-19 DOSE_COUNT empty VACCINATION_DATE equal to <= sampling date
	Medium. Provided as a single variable with the actual vaccine name and part of a table with samples including sampling date	Single row

Example	Degree of precision and description	Storage in same table of different degrees of precision
	that are implicitly positive for SARS-CoV-2. Vaccination was thus done on or before the sampling date. This additionally enables, among others, calculation of this association for each specific vaccine.	VACCINE equal to COVID-19_VACCINE_X DOSE_COUNT empty VACCINATION_DATE equal to <= sampling date
	High. Provided as a separate table with, per row, the actual vaccine name, vaccination date and dose count. Sample results table provided separately. This additionally enables, among others, calculation of this association for each specific vaccine and number of doses.	One row per vaccination event VACCINE equal to COVID-19_VACCINE_X DOSE_COUNT equal to provided value VACCINATION_DATE equal to <= sampling date

4.2 Data standards

Data standards are handled separately from the rest of the requirements because of the complexity of the subject, which warrants its own section. We looked extensively for existing data standards relevant for public health, both in terms of the data model and in terms of ontologies, including controlled terminology, to use for populating its records. These are shown in Table 8. Any standards specific for molecular biology, e.g. for genes or subtyping, are not included in this list, since it is a wide, complex and evolving field that warrants its own in-depth exploration. In addition, the formats for storage of the genetic data itself are rather well-established: fastq for raw sequence reads and fasta for finished sequences, contigs or genomes.

We did not find a standard methodology for evaluating standards. Instead, we evaluated them on the basis of own experience according to the following criteria:

1) Applicability. For data model standards, this includes data elements that are (not) covered and how well other standards can be mapped onto it.
2) International adoption
3) Sustainability, including maintenance and forward and backward compatibility
4) Documentation
5) Redundancy within the standard
6) Levels of precision that are not covered

Only two international data model standards were found for the operational data: the Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM) (CDISC, 2021) and openEHR (OpenEHR Foundation, 2022). In addition, a Dutch information model called health and care information building blocks (*zorginformatiebouwstenen*, ZIBs) maintained by the Dutch Foundation national ICT institute for healthcare (*Stichting Nationaal ICT Instituut in de zorg*, Nictiz) was found (Nictiz, 2022). There are also a number of international standards for transmission of data, including Health Level 7 (HL7) (Health Level Seven International, 2022) and ISO 13972 Health informatics – Clinical information models – Characteristics, structures and requirements (ISO, 2022). HL7 is particularly important because of its use – albeit limited at present – to transmit data to current lab surveillance systems at RIVM.

The data model standards and ontologies are described in further subsections, followed by a conclusion subsection. Given the separation of concerns between the data submission and storage (section 4.1.7), any data transmission standard or format can be used, provided that the data can be converted into that of the chosen data model standard and ontologies. For this reason, the data transmission standards are not covered further, although it is important to note that for HL7 FHIR, a mapping exists with CDISC SDTM that was jointly created by both standards organisations.

4.2.1 *openEHR and Nictiz ZIBs*

The openEHR standard is intended for semantically querying and transmitting the Electronic Health Record (EHR) of a single patient between EHR systems or between an EHR system and clinical applications. It has a formal process for updates to the different components of the standard, with the first version dating from 2006 and the current one from June 2021. Similar to openEHR, the Nictiz ZIBs are intended for semantically querying and transmitting healthcare information – including the EHR – about individual patients. It also has a formal process for updates, with the first version dating from 2015 and the current one from 2020 (prereleases for 2021 and 2022 are in place). Neither prescribe a concrete data model, typically leaving that up to the developer of the EHR system. The Nictiz ZIBs are only intended as a national, i.e. Dutch, standard.

General aspects of the openEHR standard include:

- Reference Model with six clinical classes: folder, composition, section, entry, cluster and element
- Entries: Observation (e.g. laboratory test result, symptom, vital sign, medical image), Evaluation (e.g. diagnosis, risk, prognosis), Instruction (e.g. medication prescription), Action. Instruction and Action refine Intervention in terms of what should happen and what actually happened.
- Archetypes: particular configurations of Reference Model classes, e.g. a patient summary.
- Can be used in combination with ontologies, such as LOINC, SNOMED-CT and ICD.
- Data types: basic (Boolean, identifier, state), encapsulated (multimedia, parsable), quantity (date/time, count, interval, ordinal, proportion, quantified, quantity, scale, reference range), text (terminology code, free text), time (timepoint, periodic time).

General aspects of the Nictiz ZIBs include:

- A total of more than hundred ZIBs, i.e. building blocks, for various groups: administrative, clinical context, measurements, medication, partial information models, patient context, process patterns, scales and screening tools, selfcare and treatment.
- Each ZIB has an information model, linking it to different data and container entities with a particular cardinality. Data entities can be linked to a code list and there may be constraints on the entities depending on the actual data they contain.
- Is used in combination with ontologies such as LOINC, SNOMED-CT and ICD. Specific subsets of these standards have been defined for this purpose, see section 4.2.3.

4.2.2 *CDISC Study Data Tabulation Model*

The CDISC SDTM has its origins in clinical studies for all therapeutic areas, where it is widely used and is, for one thing, the required standard for new drug applications to the US Food and Drug Administration (CDISC, 2021). It has existed for over twenty years and is continuously maintained through a robust procedure including different stages for a new version and public review. Updates to the standard are regularly made. Over 500 organisations are members of

the consortium. It has a general SDTM Implementation Guide (SDTMIG) as well as specific Therapeutic Area User Guides (TAUGs). There are TAUGs for several infectious diseases, as well as for vaccine administration and virology in general. There are also associated standards, e.g. for transmission of data (CDASH), and analysis datasets (AdaM).

General aspects of the CDISC SDTM standard include:

- Classification of data into observation classes: Interventions (e.g. treatments), Events (e.g. adverse events) and Findings (e.g. laboratory tests, ECG, questionnaire)
- Domains: specific implementations, as tables, of each observation class. Relevant domains for infectious diseases.
- Common identifier and timing variables for all observation classes. Common variables per observation class, i.e. for all of its domains. A small number of domain-specific variables. A large set of controlled terminology.
- Associated persons can be included, e.g. family members
- Study level data, e.g. trial arms, in/exclusion criteria, disease milestones
- Study references, e.g. devices, non-host organisms.
- Can be used in combination with ontologies such as LOINC, SNOMED-CT and ICD.

A more in-depth analysis of four TAUGs for infectious diseases, COVID-19, influenza, tuberculosis, vaccine administration and virology, revealed the following positive aspects:

- Can likely capture all laboratory operational data, as well as probably a large part of all other epidemiological and clinical data in several domains (tables). Relevant domains for laboratory data include BioSpecimen Findings, Biospecimen Events, Microbiology Findings, Microbiology Susceptibility and Genomics Findings.
- Uses the concept of a Study, analogous to Data Collections, except that for the latter, one record may belong to more than one Data Collection. Other Study aspects, such as trial arms, visits, reference periods and dispositions, are not applicable.
- Can handle sample aliquoting, culturing and isolation.
- Can store both original values and standardised values, thereby improving traceability.
- Can probably handle all relevant degrees of precision for time points and time periods.
- Supports questionnaires.

There are also a number of issues that require further attention. None of these are expected to be critical, especially since it is possible to add custom variables. Each of these does need actual testing or piloting to be able to resolve them well:

- Geographic location is supported to a very limited extent, and may require custom variables.
- The Genomics Findings domain is intended mainly for recording polymorphisms, one polymorphism per row. Series of polymorphisms as well as entire sequences, or references to them, would also have to be storable.

- There is a CDISC SDTM standard controlled terminology for many variables. The use of this versus other more preferred controlled terminologies needs to be evaluated, but is unlikely to require custom variables.
- There is some redundancy in sample description.
- Linkage of related records through a separate domain seems more complex than necessary.

4.2.3 Ontologies

There are clearly established international ontologies – among which controlled terminologies are a special case without any semantic links or hierarchy among terms – for coding different aspects of the data content. These include Logical Observation Identifiers Names and Codes (LOINC) (Regenstrief Institute, 2022), Systematised Nomenclature of Medicine – Clinical Terms (SNOMED-CT) (SNOMED International, 2022), Medical Dictionary for Regulatory Activities (MedDRA) (ICH Assembly, 2022), International Classification of Diseases (ICD) (WHO, 2018) and Anatomical Therapeutic Chemical (ATC) classification (WHO, 2022).

LOINC is an extensive set of codes and names for electronic reporting of clinical laboratory test results. There are codes available to describe a test with high level of detail, including the anatomical origin of the sample, or to specify the subtype of a microorganism for which the test is used, and other codes for a more generic description of the same test. SNOMED-CT is a comprehensive, hierarchically organised collection of medical terms providing codes, terms, synonyms and definitions used in clinical documentation and reporting.

Nictiz maintains a subset of LOINC and SNOMED-CT codes, together called the NL LabCodeset. These have been established as part of the Dutch *Eenheid van Taal* (Unity of Language) project and are used within the content of lab-to-lab and lab-to-public health data HL7 messages. The NL LabCodeset is limited to materials, laboratory tests and units of measurement and is restricted to antimicrobial susceptibility testing, but may be extended for broader use. Making use of these (sub)sets, therefore, has the advantage that it takes into account all the work that has gone into defining them and reaching consensus, e.g. with MMLs, while international standards are still used.

Apart from the concepts covered by the NL LabCodeset as a subset of LOINC and SNOMED-CT, there are also other concepts, such as medicinal products, taxonomy and geographical locations. For medicinal products, at least the WHO Anatomical Therapeutic Chemical (ATC) classification needs to be supported because it is already being used at RIVM. We do not expect the need to use an additional or different ontology for medicinal products. For taxonomy, SNOMED-CT, National Center for Biotechnology Information (NCBI) taxonomy, International Committee on Taxonomy of Viruses (ICTV) taxonomy for viruses, Integrated taxonomic information system (ITIS) and an organism list composed by Nictiz are all possibilities and it is likely that a combination of these needs to be used. For geographical units, a combination needs to be used as well, since there are different geographical hierarchies within the Netherlands, and outside of the Netherlands, to the extent

needed. There is, for instance, the ISO 3166 country code list and the Eurostat Nomenclature of territorial units for statistics (NUTS).

Regardless of the choice of standards for storage, it may occur that data is submitted to the platform using other standards that are used in the Netherlands, such as International Classification of Primary Care 2 (ICPC-2) or International Classification of Diseases (ICD). To support conversion of this data, a mapping of the terms between ontologies should also be stored among the reference data, so that the component that performs the conversion can make use of it. Similarly, some laboratory tests, particularly in the area of PCR or sequencing, may not yet be in any standard or not in sufficient detail (precision). In those cases, it must be possible to add these tests to the reference data as well.

4.2.4 Selection of standards

The openEHR, Nictiz ZIBs and CDISC SDTM standards are broad and can capture not only laboratory data but also clinical and epidemiological data. They can be used in combination with the relevant ontologies. However, only CDISC SDTM offers an actual tabular data model that is readily interpretable, implementable and supported with extensive and clear documentation for several infectious diseases as well as vaccine administration. The documentation also links clinical knowledge with data models and includes a wide range of examples. It has clearly demonstrated its usability for analysis of healthcare data in a clinical trial setting.

Having a robust and proven data model is very important to maintain consistency in implementation across diseases or pathogens, and as such to benefit from this consistency in terms of efficiency and quality. Without a pre-specified tabular data model, inconsistencies naturally tend to arise over time when there are a substantial number of experts dedicated to only a subset of diseases or pathogens weighing in on the data model. Additionally, CDISC offers a standard for the derived analysis datasets, the Analysis Data Model (ADaM), that is optional in this context but may bring further consistency in how analyses are performed.

Clinical trials are also very similar to public health case-based surveillance in terms of the required types of data, except that in public health surveillance there are normally no predefined trial arms, case report forms or investigational drugs, for instance. Public health case-based surveillance of infectious diseases, on the other hand, typically needs location data as well as variable questionnaire data on travel, contact and/or food consumption history when outbreaks are subsequently detected and investigated. While it is unclear if the latter would fit in the CDISC SDTM standard, this aspect of questionnaires also clearly falls outside the scope of lab surveillance. The SDTM can also be expanded with custom variables if needed, which likely allows capturing any additional concepts not currently in the standard.

For the reasons stated above, we intend to store operational data, with the possible exception of genetic data in the structure of the CDISC SDTM standard. That is, the domains (tables) and variables prescribed

by the standard should be used, taking into account the issues described in section 4.2.2. The ontologies used to populate the data should include at least LOINC and SNOMED-CT, and the relevant subsets thereof provided by Nictiz, as well as ATC. Other required ontologies are yet to be refined. Also, these broad choices generally need a lot of practical verification and refinement, which we think is best done during a pilot phase.

Table 8 Relevant data standards for lab surveillance. Some if the links with geographical scope the Netherlands refer to Dutch language websites.

Name	Content scope	Geographical scope	Maintained by
CDISC Study Data Tabulation Model (SDTM)	Clinical, laboratory and epidemiological entities. Concrete information model for master data.	International	Clinical Data Interchange Standards Consortium (CDISC)
CDISC Analysis Data Model (ADaM)	Clinical, laboratory and epidemiological entities. Concrete information model for analysis data.	International	Clinical Data Interchange Standards Consortium (CDISC)
Open Electronic Health Record (openEHR)	Electronic health record data. Information model for mainly data transmission, semantic interoperability and querying.	International	openEHR Foundation
Nictiz Clinical information building blocks (zorginformatiebouwstenen, ZIBs)	Information used in the care process, including electronic health record data. Information model for mainly data transmission and semantic interoperability.	NL	Foundation national ICT institute for healthcare (*Stichting Nationaal ICT Instituut in de zorg*, Nictiz)
Health Level 7 (HL7)	Clinical, laboratory and epidemiological entities. Data transmission standards (V2, CDA, HL7 FHIR).	International	Health Level Seven International (HL7)
ISO 13972 Health informatics – Clinical information models – Characteristics, structures and requirements	Clinical, laboratory and epidemiological entities. Information model for mainly data transmission.	International	International Organization for Standardization (ISO 13972)
Systematised Nomenclature of Medicine – Clinical Terms (SNOMED-CT)	Broad, e.g. body structure, clinical finding, organism, pharmaceutical product, social context, specimen, substance	International	International Health Terminology Standards Development Organisation (IHTSDO), trading as SNOMED International
Medical Dictionary for Regulatory Activities (MedDRA)	Broad, e.g. symptom, sign, disease diagnosis, therapeutic indication, investigation, and medical social or family history characteristic	International	ICH MedDRA Maintenance and Support Services Organization (MSSO)
International Classification of Primary Care 2 (ICPC-2)	Broad, e.g. symptom, procedure, treatment, test result	International	World Health Organization (WHO)

Name	Content scope	Geographical scope	Maintained by
International Classification of Diseases (ICD)	Disease	International	World Health Organization (WHO)
Human disease ontology	Disease	International	Institute of genome sciences, Baltimore. USA
NL Diagnosis thesaurus (*diagnosethesaurus*)	Symptom, disease. Subset of SNOMED-CT and ICD.	NL	Foundation national ICT institute for healthcare (*Stichting Nationaal ICT Instituut in de zorg*, Nictiz), Dutch Hospital Data (DHD)
Nictiz Labcodeset materials	Specimen material. Subset of SNOMED-CT.	NL	Foundation national ICT institute for healthcare (*Stichting Nationaal ICT Instituut in de zorg*, Nictiz)
Logical Observation Identifiers Names and Codes (LOINC)	Laboratory test type	International	Regenstrief institute USA
Nictiz Labcodeset methods	Laboratory test type. Subset of LOINC.	NL	Foundation national ICT institute for healthcare (*Stichting Nationaal ICT Instituut in de zorg*, Nictiz)
Identification of Medicinal Products (IDMP)	Medicinal products, pharmaceutical products, substance identification, dosage form and route of administration	International	International Organization for Standardization (ISO 11615, 11616, 11238, 11239, 11240)
Anatomical Therapeutic Chemical (ATC) classification	Medicinal products, indication	International	World Health Organization (WHO)
CBG Medicinal products information database (*Geneesmiddeleninformatiebank)*	Medicinal products	NL	Dutch Medicines Evaluation Board (*College ter beoordeling van geneesmiddelen*, CBG)
ISO 3166 country codes	Geographical location	International	International Organization for Standardization (ISO 3166)
Global Administrative Unit Layers (GAUL)	Geographical location	International	United Nations Food and Agriculture Organization (FAO)

Name	Content scope	Geographical scope	Maintained by
Nomenclature of territorial units for statistics (NUTS)	Geographical location	EU, EFTA, EU candidate and potential candidate countries	Eurostat
NL Postal code regions	Geographical location	NL	Dutch Postal Service
NL MHS regions (GGD regio's)	Geographical location	NL	Dutch government
NL Safety regions (Veiligheidsregio's)	Geographical location	NL	Dutch government
NL AMR care regions (ABR zorgregio's)	Geographical location	NL	n.a.
NL Regional Medical Consultant regions (RAC regio's)	Geographical location	NL	RIVM
NL Regional Epidemiological Consultant regions (REC regio's)	Geographical location	NL	RIVM
BES Islands (Bonaire, Sint Eustatius, Saba)	Geographical location	NL	Defined by law
CAS Islands (Curaçao, Aruba, Sint Maarten)	Geographical location	NL	Defined by law
ISO 8601 date and time format	Date and time	International	International Organization for Standardization (ISO 8601)
Unified Code for Units of Measure (UCUM)	Result unit	International	Regenstrief Institute
Nictiz Labcodeset units	Laboratory test result unit	NL	Foundation national ICT institute for healthcare (*Stichting Nationaal ICT Instituut in de zorg*, Nictiz)
Nictiz Labcodeset ordinal results	Laboratory test result unit	NL	Foundation national ICT institute for healthcare (*Stichting Nationaal ICT Instituut in de zorg*, Nictiz)
NL Civil person number (Burger Service Nummer, BSN)	Natural person	NL	Dutch Ministry of the Interior and Kingdom Relations

Name	Content scope	Geographical scope	Maintained by
NL Legal entities and partnerships number (*Rechtspersonen en Samenwerkingsverbanden Identificatie Nummer*, RSIN)	Legal entity, partnership	NL	Dutch Trade Registry
NL NVMM Microbiology laboratory list	Medical microbiological laboratory	NL	Dutch Society for Medical Microbiology (*Nederlandse Vereniging voor Medische Microbiologie*, NVMM)
NL AGB Registry	Healthcare provider	NL	Vektis
Codex alimentarius	Food	International	United Nations Food and Agriculture Organization (FAO)
Food ontology	Food	International	Academic consortium
LangualL	Food	International	Danish Food Informatics
FoodEx2	Food	International	European Food Safety Authority (EFSA)
National Center for Biotechnology Information (NCBI) taxonomy	Taxonomy	International	National Institutes of Health (NIH), USA
International Committee on Taxonomy of Viruses (ICTV) taxonomy for viruses	Taxonomy	International	International Committee on Taxonomy of Viruses (ICTV)
Integrated taxonomic information system (ITIS)	Taxonomy	International	Integrated Taxonomic Information System (ITIS) partnership
Nictiz Organism list	Taxonomy. Subset of SNOMED-CT.	NL	Foundation national ICT institute for healthcare (*Stichting Nationaal ICT Instituut in de zorg*, Nictiz)

4.3 Generation of requirements

Specifying good initial requirements is a crucial step towards designing and building an effective and efficient system or collection of systems (i.e. a platform). We use the term 'initial requirements', since these inevitably change over time, including during the implementation of the first release, as areas of uncertainty are clarified and misconceptions are corrected.

In order to create a good set of initial requirements, a gradual process of refinement was followed. The platform is not only described from an end-user perspective, but also more in depth, e.g. in terms of conceptual components, legal compliance and data standards (the latter are handled separately in section 4.2). We consider this more holistic multi-disciplinary approach that includes working with conceptual components as crucial for the development of a good system or platform. It reduces the chances of missing an aspect and provides a common framework that includes insights from different disciplines.

4.3.1 High-level requirements and legal compliance

As a first step in the requirements generation process, the fundamental principles described in section 4.1 were translated into a number of high-level fundamental minimum requirements. These also trace back to all the legal compliance requirements of section 4.1.3, to ensure clarity on where and how these are taken into account. This is in itself a legal requirement for systems that handle personal data, as is the case here, stemming from GDPR article 25 (data protection by design and by default).

The high-level minimum requirements are listed in Table 10, together with the references to applicable law. Identifiers follow the naming scheme HLR-#. The requirement on separation of concerns, HLR-1, defines a number of (categories of) conceptual components that are further defined in section 4.3.4.

4.3.2 User requirements

In a second step, user requirements were elicited from different groups of stakeholders, including internal and external SMEs. Requirements from external SMEs were gathered during a total of twenty sessions with individual MMLs (see also section 4.1.2). In practice, this step was executed in parallel with the first one of the high-level requirements (section 4.3.1). The requirements gathered mostly covered data submission and data sharing, including user interface functionality for accessing the results derived from submitted data. In addition, recommendations on how to organise the collaboration around the platform, regardless of its implementation, were also collected, and these formed the basis for a significant part of the Discussion Paper (section 4.1.2). This Discussion Paper was then again discussed with external experts.

Requirements and remarks from MMLs as well as internal users from RIVM (section 3.2) are directly integrated into the conceptual components (section 4.3.4) and the consolidated initial requirements for

the platform (section 4.4). Table 9 gives an overview of the different activities performed for user requirement elicitation (Alfeche, 1997).

Table 9 Activities performed for user requirements elicitation.

Source	Activity
RIVM IDS Subject Matter Experts	Process discovery sessions (24x)
MML Subject Matter Experts	1-on-1 structured interviews (20x) Plenary session based on Discussion Paper
MHS Subject Matter Experts	Focus group discussion Plenary session based on Discussion Paper

4.3.3 *Corporate software requirements*

RIVM imposes these requirements on all software that is used by the organisation, and as such are not specific for the lab surveillance platform. They include using supported application platforms and database servers, having a (three-)tiered architecture, making use of open standards, and a series of security-related requirements. These requirements are not listed here since they are not specific for the lab surveillance platform, but we do categorise the consolidated requirements of section 4.4, not only in terms of the conceptual components but also in terms of which of the three tiers of the application they fit in.

4.3.4 *Conceptual components and their responsibilities*

Following the definition of the (categories of) conceptual components in requirement HLR-1 (section 4.3.1), these conceptual components are further defined here in terms of their main responsibilities. This definition was made mainly based on the other high-level requirements, as well as the user requirements, choice of data standards and general corporate software requirements of sections 4.2, 4.3.2 and 4.3.3. For the purposes of this report, we do not cover the Authentication component since it purely encapsulates IT-technical functionality, independent from the data content and related legal aspects. Table 11 describes the conceptual components and their main responsibilities, and Figure 4 gives a visual overview of the conceptual dependencies.

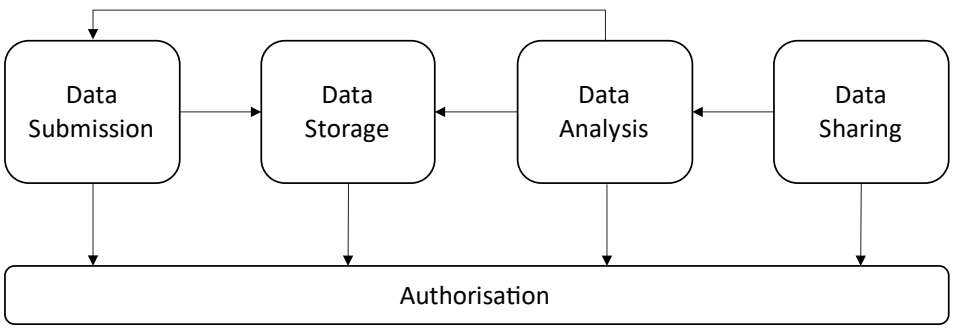


Figure 4 Conceptual components of the new system. Arrows express "x makes use of y", and as such denote dependencies between these components.

4.4 Consolidated initial platform requirements

The previous sections on gathering fundamental principles and requirements, including data standards (sections 4.1, 4.2, 4.3), represent the first two major steps in gradually developing an understanding of what is required from the lab surveillance platform and also framing it in a broader context of purpose and collaboration between different partners. In this section, we take a third and final step by consolidating all these requirements into a single initial list that forms a good starting point for the next stages of actual implementation.

This consolidation includes deriving as much as possible mutually exclusive and commonly exhaustive requirements. To reduce the probability of omissions, they were written down following the main data flow through the system, from data submission to the platform and finally sharing results again with users. The requirements are described in Table 12 and have the following variables:

- A unique identifier.
- The conceptual component and application tier to which the requirement applies. The conceptual components are those of HLR-1, i.e. Data Submission, Data Storage, Data Analysis, Data Sharing and Authorisation. The three application tiers used are Presentation layer, Business Layer and Data Layer (Rob Conery, 2009)
- The description of the requirement. In line with the Principle of Least Knowledge, also known as the Law of Demeter, requirements applying to one conceptual component may not refer to internal functionality of another component.
- The traceability of the requirement to those gathered in sections 4.2 and 4.3, and thus implicitly the traceability to and from legal requirements. An additional description of the rationale is added, in particular in case of no immediate traceability.

Table 10 Fundamental high-level minimum requirements for the lab surveillance platform. The order of these requirements generally follows the main operational data flow.

Id	Description	Rationale
HLR-1	The platform must be governed by a Terms of Use that is agreed with MMLs and MHSs and covers at least a means to jointly decide relevant changes to the platform and specific uses of the data. This includes access to results derived from submitted data, and a check on the legal compliance of providing access to these results.	Provides transparency on how the platform should be used, along with the means for addressing suggested changes and regulating allowed data use. Considered required for GDPR compliance: Art 5 (Integrity and confidentiality), Art 24, Art 25. Facilitates Data Protection Impact Assessment (Art 35).
HLR-2	The platform must consist of independent components with a clear interface for access. At least the following (categories of) components must exist: Data Submission, Data Storage, Data Analysis, Data Sharing, Authorisation and Authentication. Component(s) must preferably be developed as open source.	Reduces complexity of the platform and thereby the probability of conceptual errors as well as implementation errors (bugs). Contributes to international public health through potentially reusable code, and improves code quality.
HLR-3	Submitted operational data, including data for correction or deletion of existing values or records, must be stored as received along with relevant metadata of the submission such as the time and account used.	Enables tracing back any derived data to its origin, e.g. for the purpose of finding the root cause of data errors. Considered required for GDPR compliance: Art 5 (Transparent processing, Accuracy, Accountability), Art 16, Art 17.
HLR-4	Submitted operational data must be transformed into and stored in a format that is compatible with selected existing external standards, both with respect to structure and to content. These standards must allow storage of data with different levels of relevant precision as well as expansion to other pathogenic species without changes in storage format.	Makes use of the expert knowledge and consensus that went into developing the standards, thereby reducing the probability of conceptual errors. Contributes to preparedness by making the platform ready for pathogen/disease X. Potentially enables use of existing third-party software libraries. Potentially facilitates the recruitment of appropriate personnel. Improves speed of implementing changes. Considered required for GDPR compliance: Art 20.

Id	Description	Rationale
HLR-5	Deviations from storage compatible with the external standards of (HLR-4) must be documented and rationalised to the extent that compliance with the standard for the type of data in question is demonstrated to be infeasible.	Reduces the probability of conceptual errors as well as implementation errors (bugs). Reduces non-portability of data.
HLR-6	The standardised operational data of (HLR-4) must be traceable to the submission(s) of (HLR-3) and the individual records or data elements within those submissions.	Enables tracing back any derived data to its origin, e.g. for the purpose of finding the root cause of data errors. Considered required for GDPR compliance: Art 5 (Transparent processing, Accuracy, Integrity), Art 25.
HLR-7	An audit trail must store previous versions of records within the standardised operational data, so that changes to records can be inspected and the state of the standardised operational data can be reconstructed at any given point in time.	Same as (HLR-6).
HLR-8	Submitted operational data must include the data collection protocol(s) under which each record has been collected. These protocols must be agreed in advance and stored as reference data, including at least (i) the purpose and corresponding legal bases for the data collection, (ii) species and the variables that may be collected, (iii) the statistical sampling method, including convenience sampling as a method, and (iv) the minimal reporting frequency.	Enables validation against allowed variables. Enables selection of (statistically) relevant subsets of observations, thereby improving accuracy. Improves verification of timeliness. Considered required for GDPR compliance: Art 5 (Transparent processing, Purpose limitation, Data minimisation, Integrity), Art 9, Art 25. Facilitates Data Protection Impact Assessment (Art 35). Considered required for NL Public health act: Art 1, Art 19, Art 24.
HLR-9	Operational data must be verified against automated validation rules both prior to completion of submission and after storage. The submitter must be notified of any issues found prior to completion of submission. Operational data with severe issues must be prevented from submission.	Increases detection of data quality issues. Prevents submission of low-quality data. Considered required for GDPR compliance: Art 5 (Accuracy, Integrity).

Id	Description	Rationale
HLR-10	Variables used to store operational and reference data must be stored themselves as metadata and be annotated with at least a classification as to whether they can contain personal data and if so which type (general, national identifier, special categories).	Considered required for GDPR compliance: Art 9. Considered required for NL Implementing act GDPR: Art 25, Art 28, Art 30, Art 46.
HLR-11	Data may only be submitted and accessed by duly authorised people. Authorisation must take into account at least (i) the organisation, the type of organisation (including at least MML, MHS and RIVM), (ii) the data collection protocol and (iii) the type of access.	Prevents misuse of data. Considered required for GDPR compliance: Art 5 (Confidentiality, Accountability), Art 25. Considered required for NL Law on public health: Art 1, Art 6, Art 14, Art 19, Art 27, Art 28. May facilitate Art 25 processes.
HLR-12	Stored data must be logged with timestamps for when the record was created and last modified. The dynamic generation of non-persistent results derived from operational and reference data must be logged and must include a start and end timestamp.	Enables assessment of timeliness of processes and individual steps and thereby finding the root cause of any performance issues.

Table 11 Conceptual components of the new system and their main responsibilities. Relevant concepts within those responsibilities are upper cased and in italics.

Component	Main responsibilities
Data Submission	1) The interfaces for submission of operational data (*Submitted Operational Data*) in various formats to RIVM from external parties or other RIVM systems, either through machine-to-machine communication or manually through a web interface. This includes subsequent submission of relevant outputs from bioinformatics pipelines for primary and secondary analyses such as (links to) consensus genomes, genotypic classifications and phenotypic predictions. 2) The conversion of *Submitted Operational Data* into *Standardised Operational Data* and subsequent storage into the Data Storage component. 3) The provision of *Validation Results* to the submitting user to ensure the quality of the *Submitted Operational Data*. 4) Maintaining the traceability of *Data Submissions* and the corresponding *Submitted Operational Data*, *Validation Results* and *Standardised Operational Data*.

Component	Main responsibilities
Data Storage	1) The storage and interfaces for *Standardised Operational Data*. 2) The storage and interfaces for *Reference Data*, including *Geographical Entities*, relevant ontologies, *Organisations*, *Reference Genetic Data*, *Data Collections* and *Result Types*. 3) The storage and interfaces for *Set Data* of a particular type (*Entity*), e.g. *Sample Sets*. 4) The storage and interfaces for *Metadata* for all *Entities*, including their *Variables*, and *Validation Rules*. 5) Maintaining the traceability from *Data Submissions* to *Standardised Operational Data*, and the traceability of changes to stored data. 6) The performance of *Data Validation*, generating *Validation Results.*
Data Analysis	1) Searching for samples on genotypic or phenotypic similarity, potentially generating *Sample Sets*. 2) Generating dynamic *Result Data* and relevant data structures for sharing, for each *Result Type*, a given user and *Sample Set*, and taking into account the user's *User Rights* in terms of which data they are allowed to access. 3) Conversion of bioinformatically derived laboratory results into *Standardised Operational Data* and subsequent storage of the relevant results in the Data Storage component through the Data Submission component. 4) The Data Analysis component is *not* responsible for execution of bioinformatic pipelines that perform primary and secondary analyses such as genome assembly, reference genome mapping, phenotypic predictions and genotypic classifications.
Data Sharing	1) The interfaces for accessing and exporting *Result Data*. 2) The generation of relevant visualisations of *Result Data*.
Authorisation	1) The storage and interfaces for *User Right Types,* linked to Organisations, Organisation Types, Geographical Entities, Data Collections and Result Types. 2) The storage and interfaces for an individual user's *User Rights* in terms of *User Right Types*.
Authentication	1) The authentication, i.e. verification of the identity, of users.

Table 12 Consolidated initial lab surveillance platform requirements.

Id	Component	Description	Rationale
REQ-1	Data Submission - Presentation layer	It must be possible to submit file-based operational data in different formats. Formats include at least the Standard Format, Type-Ned MRSA, Type-Ned CPE, Type-Ned Entero-Paracho, Virological weekly updates, and COVID-19 lab surveillance.	HLR-2. Submission in different formats. Decouples from storage.
REQ-2	Data Submission - Presentation layer	It must be possible to extract relevant operational data from the RIVM iLES Laboratory Information System and submit it.	HLR-2. Submission in different formats. Decouples from storage.
REQ-3	Data Submission - Presentation layer	It must be possible to provide, along with the initially Submitted Operational Data, additional required variables with corresponding values for the submission, which are applicable to, i.e. constant for, all submitted records. This may include the Data Collection(s) under which the data is submitted. Together, these form the complete Submitted Operational Data.	HLR-8. Submission formats often contain implicit, required information such as the pathogen or Data Collection concerned. This may need to be rendered explicit if data can be submitted in various formats.
REQ-4	Data Submission - Presentation layer	It must be possible to validate the Submitted Operational Data, thereby initiating a Data Submission.	HLR-9
REQ-5	Data Submission - Presentation layer	It must be possible to show Validation Results applied to the Submitted Operational Data both upon validation and after completion of the Data Submission.	HLR-9
REQ-6	Data Submission - Presentation layer	It must be possible to show the result of conversion of the Submitted Operational Data into Standardised Operational Data during and after completion of the Data Submission	Provides an extra opportunity for the user to inspect the data before submission, thereby increasing the chance of detection of issues and in particular those that cannot be detected through validation rules.
REQ-7	Data Submission - Presentation layer	It must be possible to submit operational data for previously uploaded records identically to other Data Submissions. Data for both new and previously existing variables may be provided and in the latter case, the new value may be identical, different or deleted.	HLR-3

Id	Component	Description	Rationale
REQ-8	Data Submission - Presentation layer	It must be possible to submit file-based operational data for previously uploaded records in order to delete those records in their entirety, identically to other Data Submissions.	HLR-3
REQ-9	Data Submission - Presentation layer	Users with no User Rights to submit for any Data Collection may not gain access to any web interface	HLR-11
REQ-10	Data Submission - Presentation layer	The first time that users log in, they must accept the Terms of Use of the system. They also need to do so when there is an update to the Terms of Use.	HLR-1
REQ-11	Data Submission - Presentation layer	It must be possible to show the descriptions of the different Data Collections	Informs the user about which Data Collections data can be submitted for, and what their specifications are.
REQ-12	Data Submission - Presentation layer	Any web interface must use the same interface for conducting a Data Submission as a Data Submission through machine-to-machine communication	Enforces reuse.
REQ-13	Data Submission - Business layer	There must be an interface(s) for receiving the Submitted Operational Data, thereby initiating a Data Submission. These must return the Standardised Operational Data as well as the Validation Results.	HLR-4, HLR-6, HLR-9
REQ-14	Data Submission - Business layer	It must be possible to convert Submitted Operational Data of different formats into the Standardised Operational Data. In addition, metadata must be included that links each Standardised Operational Data record to its source record(s) in the Submitted Operational Data.	HLR-4, HLR-6
REQ-15	Data Submission - Business layer	It must be possible to retrieve Validation Results for a set of Standardised Operational Data from the Data Storage component	HLR-9
REQ-16	Data Submission - Business layer	It must be possible to store Standardised Operational Data in the Data Storage component	HLR-4, HLR-6

Id	Component	Description	Rationale
REQ-17	Data Submission - Business layer	A Data Submission must be persisted in the Data layer at each step of the process, including at least (i) the submitting user, (ii) the time of initiation, (iii) the time and reason of (un)successful completion, (iv) the Submitted Operational Data, (v) the Standardised Operational Data if available, (vi) the Validation Results if available	HLR-3
REQ-18	Data Submission - Business layer	Users without the User Right to submit for a particular Data Collection must be blocked from proceeding from the Data Submission at the latest during the validation stage	HLR-11
REQ-19	Data Submission - Business layer	Ongoing Data Submissions may not have any records in common	Avoids possibility of Validation Results not being valid for the Data Submission in question.
REQ-20	Data Submission - Business layer	The acceptance of the Terms of Use and which version, must be stored in the Data layer	HLR-1
REQ-21	Data Submission - Business layer	It must be possible to temporarily deactivate making Data Submissions containing records for a particular Data Collection	Allows Data Collection-specific maintenance, e.g. on Reference Data. Also allows extraction and validation of an analysis dataset while it is guaranteed that the underlying data has not changed between the two steps.
REQ-22	Data Storage - Presentation layer	It must be possible to create a set of instances of the same type based on a set of instance identifiers and to define and update data for this set (Set Data). In particular, Sample Sets must exist and it must be possible to submit Sample Set Data.	Sets of instances, and in particular Sample Sets are essential and operational data, regardless of how they are defined. Sample Sets can store sets of samples that are of epidemiological interest, such as signals, outbreaks and all samples during a particular time period for analysis. They can be used in their own right as well as serve as a clear means for selecting the respective samples, e.g. for further analyses.

Id	Component	Description	Rationale
REQ-23	Data Storage - Presentation layer	It must be possible to add, update or delete different types of Reference Data records	Ease of use, versus, for instance, command line or direct SQL updates
REQ-24	Data Storage - Presentation layer	Users with no Admin User Rights may not gain access to any web interface	HLR-11
REQ-25	Data Storage - Business layer	There must be interface(s) for receiving Standardised Operational Data for Data Validation. These must return the Validation Results. This also applies to Reference Data, Set Data and Metadata.	HLR-9
REQ-26	Data Storage - Business layer	There must be interface(s) for storing Standardised Submitted Data. This also applies to Reference Data, Set Data and Metadata.	HLR-4, HLR-6
REQ-27	Data Storage - Business layer	Data Validation of Standardised Submitted Data must be performed after merging the Standardised Submitted Data with any already stored data. This also applies to Reference Data, Set Data and Metadata.	HLR-9. Detects problems for the entire record in question. If not done this way, it opens up the possibility for data with critical errors that would otherwise be rejected, to enter the database. In addition, users reviewing the validation results can also see the broader impact of their submission.
REQ-28	Data Storage - Business layer	The implementation of individual Validation Rules must be versioned	HLR-9
REQ-29	Data Storage - Business layer	Validation Results must include a hash code per result on already stored records and that remains identical as long as the record values used to perform the validation, as well as their internal identifiers, remain identical	HLR-9. Allows distinguishing new validation results derived directly from the database from previous validation results derived directly from the database. This in turn allows excluding non-issues or known but unresolvable issues when inspecting the database.
REQ-30	Data Storage - Data layer	Metadata must be stored for at least Entities, Variables and Validation Rules, and cover both Reference Data and Operational Data	HLR-9, HLR-10

Id	Component	Description	Rationale
REQ-31	Data Storage - Data layer	Reference Data must be stored for geographical entities, any applicable ontologies, Reference Genetic Data, Data Collections, Result Types and Organisations	HLR-8
REQ-32	Data Storage - Data layer	Operational Data, potentially including data on samples of non-human origin, must be stored in a format compatible with the Standard Format, with the exception of Set Data. The Standard Format is CDISC SDTM, with the exception of Genetic Data. Domain specific ontologies used, where possible, are LOINC and SNOMED-CT subsets defined by Nictiz, units of measurement defined by Nictiz, ATC and a combination/subset of taxonomy by NCBI, ICTV and SNOMED-CT.	HLR-4, HLR-5. Potential future compatibility with non-human origin samples for One Health lab surveillance.
REQ-33	Data Storage - Data layer	Metadata, Reference Data and Operational Data must be stored in a Variable-Value format, i.e. not in separate physical columns per Variable. In addition, creation timestamp, last modified timestamp and (virtually) deleted status must be stored.	HLR-10. Allows evolution of the data in the system without requiring any costly and slow physical changes to the database. One particular example is the addition of an emerging pathogen/disease X, requiring the ability to adapt fast in order to maintain a good level of preparedness.
REQ-34	Data Storage - Data layer	An audit trail must be present for all Metadata, Reference Data and Operational Data	HLR-7
REQ-35	Data Storage - Data layer	Each instance of an Entity must have its own integer identifier assigned. Multiple external identifiers, each coupled to the type of identifier, must be storable per instance. For example, a sample may have an identifier assigned by the processing laboratory as well as by RIVM. There may even be additional identifiers, such as one assigned by the system for further external use.	There are often several external identifiers for a particular instance, which must normally all be stored to enable finding an instance based on their value. Having a numeric internal integer used subsequently throughout the system as a (foreign) key increases transparency and efficiency.
REQ-36	Data Storage - Data layer	It must be possible to implement derived persisted views or equivalents on the master data, including through partial updates that take into account only more recently modified master data records.	Improves performance by (i) decoupling reading, which can be done mostly in the persisted views or equivalents from writing and (ii) making the update process of these views efficient

Id	Component	Description	Rationale
REQ-37	Data Analysis - Presentation layer	The Data Analysis component has no Presentation layer	Not foreseen to be needed. Added for completeness.
REQ-38	Data Analysis - Business layer	It must be possible to retrieve input data for analyses from the Data Storage component	The Data Storage component is the source of input data for analyses.
REQ-39	Data Analysis - Business layer	It must be possible to store the most relevant output from analyses in the Data Storage component as derived laboratory test results (e.g. consensus genome or identifier, wgMLST, phenotypic predictions, phylogenetic classifications, quality assessments) or sample sets (e.g. cluster detection). Reference(s) that allow tracing how the analysis was performed must also be included.	Some derived results, in particular from genetic sequence data through bioinformatics pipelines, need to be stored since they are operational data in their own right. Adding references enables traceability of how results were generated.
REQ-40	Data Analysis - Business layer	It must be possible to derive the Result Data for each Result Type from input data	Generation of all relevant pre-determined results for sharing.
REQ-41	Data Analysis - Business layer	It must be possible to filter or further adjust the Result Data for each Result Type depending on the User Rights of the user.	Application of user rights to access to results for sharing. For instance, depending on the region the user is a member of, data from that region might not be aggregated, filtered or reduced in precision, whereas data from other regions is; instead of individual organisations, an indicator might be derived that conveys whether the sample is submitted by the user's own organisation or by another.
REQ-42	Data Analysis - Business layer	It must be possible to perform genotypic similarity searches based on wgMLST allele distance (different pre-defined schemes) and on SNP distance among selected samples and up to a maximum distance	Enables the two most relevant genetic similarity searches.
REQ-43	Data Analysis - Business layer	It must be possible to perform partial phenotypic similarity searches for antimicrobial resistance to different antimicrobials	Enables searching for specific resistance patterns, for instance.
REQ-44	Data Analysis - Business layer	It must be possible to generate different types of trees, including trees based on hierarchical clustering and phylogeny, for a selected set of samples	Enables the two most relevant tree generation methods.

Id	Component	Description	Rationale
REQ-45	Data Analysis - Data layer	The Data Analysis component has no Data layer	Component layer not foreseen to be needed. Added for completeness.
REQ-46	Data Sharing - Presentation layer	The user must be able to see which Result Types they have access to.	Gives an overview to the user.
REQ-47	Data Sharing - Presentation layer	Each Result Type must be accessible through a separate URL that may be parameterised, irrespective of the user.	Allows including the link in communication, such as emails, which improves the efficiency of using the system.
REQ-48	Data Sharing - Presentation layer	At least the following categories of Result Types must be possible: own submitted data, trend of positive samples stratified by geographic region and/or subtype, annotated tree for a particular Sample Set.	These are general categories of Result Types, to be refined and agreed together with users through the processes agreed in the Terms of Use for the platform.
REQ-49	Data Sharing - Presentation layer	The user must be able to export relevant Result Data in an easy-to-use format	Users may want to use the data outside the system.
REQ-50	Data Sharing - Presentation layer	The user must be able to interact with the representations of the Result Data in relevant ways, e.g. sorting and filtering tables, zooming and panning maps, mouse-overs in a graph, zooming and highlighting nodes on a tree, selecting sample(s).	Improves general usability from static representations.
REQ-51	Data Sharing - Presentation layer	The user must be able to search for matching samples based on relevant genotypic or phenotypic (antimicrobial resistance) similarity for relevant Result Types.	Allows users to search for relevant samples on their own.
REQ-52	Data Sharing - Business layer	It must be possible to request the Data Analysis component to generate the Result Data for a particular Result Type, for a selection of samples, and further filtered and adjusted according to the user's User Rights.	The Data Analysis component is the source of Result Data.
REQ-53	Data Sharing - Business layer	The requests for viewing Results, as well as any user-performed searches within these Results, must be persisted in the Data layer, including at least (i) the user, (ii) the time of the request and (iii) the nature of the request.	HLR-22. Allows understanding user behaviour and generating statistics
REQ-54	Authorisation - Presentation layer	It must be possible to add, update or delete different types of User Rights for a particular user.	Basic functionality for this component

Id	Component	Description	Rationale
REQ-55	Authorisation - Business layer	Users with no Admin User Rights may not gain access to any web interface.	HLR-11. Admin rights are required for viewing and maintaining authorisation data.
REQ-56	Authorisation - Business layer	There must be interface(s) to retrieve a list of users.	HLR-11
REQ-57	Authorisation - Business layer	There must be interface(s) to retrieve a list of User Rights that a user has.	HLR-11
REQ-58	Authorisation - Data layer	Data must be stored for all Users, User Right Types and User Rights.	HLR-11
REQ-59	Authorisation - Data layer	User Right Types must have at least the dimensions Organisation, Organisation Type (including at least MML, MHS, country within the Kingdom), Data Collection, Result Type and Region.	HLR-8. Also allows future access by CAS countries.
REQ-60	Authorisation - Data layer	An audit trail must be present for all stored Entities.	HLR-7

5 Conclusion

This report concludes the main part of the preparatory phase for the establishment of a new national lab surveillance platform in the Netherlands, to be embedded in a wider public health surveillance environment at RIVM. It covers the legal basis and elements of compliance as they pertain to lab surveillance, the current processes and systems and lessons learned from them, and the desired future situation both in terms of the lab surveillance platform and the corresponding collaboration with MMLs and MHSs. Both MMLs and MHSs will have access to the platform. Issues that have been found during the process, but are out of scope for this project will be dealt with in the follow-up on this project with the MMLs and MHSs.

The legal basis for lab surveillance and compliance is a complex issue. This is mostly due to the fact that a large part of the collected operational data is classified as personal data, even if pseudonymised rather than directly identifiable. An improvement of the legal framework within the Netherlands may substantially increase clarity and efficiency, in particular for MMLs in terms of their role in public health surveillance.

A substantial number of the current processes and systems in use at RIVM were described consistently and assessed for similarities. These processes have often been developed and refined over a long period of time. However, they do suffer from fragmentation, whereby a somewhat different approach is chosen per pathogen or pathogen group, while at the same time the collaborating partners – in particular the MMLs – are generally the same. This fragmentation also introduces manual work that is not easily scalable in a pandemic situation and results in data quality issues and more verification steps.

Several systems are also technically reaching their end-of-life. Finally, recurring issues stated by MMLs, apart from those mentioned in the previous paragraph, include the time required to extract and submit data to RIVM, authorship on scientific papers using the lab surveillance data, and being able to access or receive more results from RIVM on the data they have submitted.

The future lab surveillance platform is expected to replace the current systems over time, applying a consistent way of handling data for all pathogens while taking into account their inherent differences. This includes the areas of data submission, data storage, data analysis, data sharing and user access rights. In addition, more of the submitted data is intended to be made accessible to, among others, other MMLs and MHSs. The details will have to be worked out together. Finally, the future platform is also designed to be prepared for newly emerging pathogens.

A pilot phase for a selected number of pathogens – such as SARS-CoV-2 – needs to be conducted, using agile software development and rapid user feedback. This is needed to handle the uncertainty inherent to developing any complex system in an efficient way both in terms of time

and cost. It is also the reason why the consolidated requirements for the complete platform that are specified in this document are called initial requirements (see section 4.3). These requirements may change as a result of the experiences in the pilot phase and after.

The pilot phase is to be executed incrementally, with requirements prioritised on the basis of their expected added value. One particular element that still needs to be worked out further is the One Health aspect and combining public health lab surveillance with that of the food and veterinary and environment sectors, and for which future compatibility needs to be taken into account. This may be worked out in parallel with the pilot phase, but should not unduly delay its progress. Finally, following the pilot phase, or if possible to some extent in parallel, the production platform needs to be set up, which may reuse components from the pilot phase in order to save costs.

A final crucial factor for future success of the pilot phase is the collaboration between RIVM, MMLs and MHSs. This needs to be further strengthened, including agreeing on a first version of the Terms of Use for the platform. Capabilities developed should deliver benefits for all cooperating parties and this will be tested in the pilot phase.

Acknowledgements

We would like to thank in particular the twenty medical microbiological laboratories that participated in individual sessions to discuss and contribute to the future of laboratory-based surveillance in the Netherlands, as well as the Dutch Society for Medical Microbiology. In addition, we would like to thank everyone that attended the group session on 15 Sep 2022 for their contributions.

References

Alfeche, R. a. (1997). *Requirements Engineering A good practice guide.* John Wiley and Sons.

Black, A. M. (2020). Ten recommendations for supporting open pathogen genomic analysis in public health. *Nature medicine, 26(6)*, 832-841. Retrieved from https://doi.org/10.1038/s41591-020-0935-z

Burgelijk Wetboek Boek 7. (1994, November 17). *Wet van 17 november 1994 tot wijziging van het Burgerlijk Wetboek en enige andere wetten in verband met de opneming van bepalingen omtrent de overeenkomst tot het verrichten van handelingen op het gebied van de geneeskunst*. Retrieved from overheid.nl: https://wetten.overheid.nl/BWBR0005290/2022-08-02/0#Opschrift

CDISC. (2021, November 29). *SDTM v2.0.* Retrieved from cdisc.org: https://www.cdisc.org/standards/foundational/sdtm

ECDC. (2014). *Data quality monitoring and surveillance system evaluation.* Stockholm, Sweden: European Centre for Disease Prevention and Control (ECDC).

Emily Griffiths, C. G. (2019). *Handbook on Statistical Disclosure Control for Outputs.* United Kingdom. Retrieved from https://securedatagroup.files.wordpress.com/2019/10/sdc-handbook-v1.0.pdf

European Parliament and the Council of the European Union. (2004, April 21). *EU Regulation (EC) No 851/2004 of the European Parliament and of the Council of 21 april 2004 establishing a European Centre for disease prevention and control*. Retrieved from EUR-Lex: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32004R0851

European Parliament and the Council of the European Union. (2016, April 27). *Regulation (EU) 2016/679 of the European Parliament and Of the Council*. Retrieved from EUR-Lex: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

Ficek, J. W. (2021). Differential privacy in health research: A scoping review. *Journal of the American Medical Informatics Association : JAMIA, 28(10)*, 2269–2276. Retrieved from https://doi.org/10.1093/jamia/ocab135

Gargis, A. S. (2015). Good laboratory practice for clinical next-generation sequencing informatics pipelines. *Nature biotechnology, 33(7)*, 689–693. Retrieved from https://doi.org/10.1038/nbt.3237

Health Level Seven International. (2022, October 13). *HL7 Standards.* Retrieved from hl7.org: https://www.hl7.org/implement/standards/index.cfm?ref=nav

ICH Assembly. (2022, September). *MedDRA Version 25.1.* Retrieved from meddra.org: https://www.meddra.org/basics

ISO. (2022). *ISO 13972:2022(en) Health informatics — Clinical information models — Characteristics, structures and requirements.* Retrieved from iso.org: https://www.iso.org/obp/ui/#iso:std:iso:13972:ed-1:v1:en

Johan Hansen, P. W.-B. (2021). *Assessment of the EU Member States' rules on health data in the light of GDPR.* Luxembourg: Publications Office of the European Union.

Kodra, Y. W.-d.-l.-P.-F. (2018). Recommendations for Improving the Quality of Rare Disease Registries. *International journal of environmental research and public health, 15(8), 1644*. Retrieved from https://doi.org/10.3390/ijerph15081644

Nictiz. (2022, June 8). *Health and Care Information models (HCIM) Prerelease 2022-1.* Retrieved from zibs.nl: https://zibs.nl/wiki/HCIM_Mainpage

OpenEHR Foundation. (2022, October 13). *openehr.org.* Retrieved from openEHR Specifications: https://specifications.openehr.org/

Regenstrief Institute. (2022, February). *LOINC Version 2.72.* Retrieved from loinc.org: https://loinc.org/kb/loinc-release-notes/

Rob Conery, S. H. (2009). *Microsoft Application Architecture Guide* (2nd ed.). USA: Microsoft Press. Retrieved from https://learn.microsoft.com/en-us/previous-versions/msp-n-p/ee658124(v=pandp.10)

Rothman, K. J., & Lash, T. L. (2008). *Modern Epidemiology* (3rd ed.). Philadelphia, USA: Lippincott Williams & Wilkins.

SNOMED International. (2022, September 30). *SNOMED CT.* Retrieved from snomed.org: https://www.snomed.org/snomed-ct/why-snomed-ct

Staten-Generaal. (2017, November 17). *Statuut voor het Koninkrijk der Nederlanden*. Retrieved from overheid.nl: https://wetten.overheid.nl/BWBR0002154/2017-11-17

Staten-Generaal. (1996, Oktober 21). *Wet op het RIVM*. Retrieved from Overheid.nl: https://wetten.overheid.nl/BWBR0008289/2020-03-19

Staten-Generaal. (2008, Oktober 9). *Wet publieke gezondheid*. Retrieved from Overheid.nl: https://wetten.overheid.nl/BWBR0024705/2021-09-01

Staten-Generaal. (2016, April 27). *Uitvoeringswet Algemene verordening gegevensbescherming*. Retrieved from Overheid.nl: https://wetten.overheid.nl/BWBR0040940/2021-07-01

Verner, L. (2004, May 1). *The challenge of Process Discovery.* Retrieved from BPTrends.com: https://www.bptrends.com/the-challenge-of-process-discovery/

Weilkiens, T. (2016). *OCEB 2 Certification Guide, Business Process Management - Fundamental Level, second edition.* Cambridge, United States: Elsevier.

WHO. (2018, June 18). *ICD-11 - International Classification of Diseases 11th Revision.* Retrieved from who.int: https://icd.who.int/en

WHO. (2022, January 1). *Anatomical Therapeutic Chemical (ATC) Classification.* Retrieved from who.int: https://www.who.int/tools/atc-ddd-toolkit/atc-classification

Glossary

Business process

The glossary of Workflow Management Coalition (WfMC) describes the business process as a set of one or more linked procedures or activities that collectively realise a business objective or policy goal, normally within the context of an organisational structure defining functional roles and relationships.

Information system

An information system comprises a series of processes or protocols involving the collection, processing and presentation of information which may or may not contain elements of technology. (ECDC, 2014)

Laboratory-based surveillance (Lab surveillance)

Laboratory-based surveillance for public health, or in short lab surveillance, entails the collection and analysis of primarily human sample-based laboratory data for the primary purpose of signal detection, such as the detection of short or long term trends in pathogen variants. See also Signal detection.

Laboratory-based surveillance (Lab surveillance) platform

The laboratory-based surveillance platform, or lab surveillance platform for short, consists of a set of information systems or components that enable users to efficiently and effectively submit, analyse and share sample-based data on infectious diseases for the primary purpose of lab surveillance. In addition, the platform may facilitate scientific research and must support preparedness by being easily adjustable to accommodate a newly emerging pathogen.

Clinical data.

This data contains elements that are typically recorded in the medical record of a patient with respect to, for example, medication, vaccination, (severity of) symptoms and diagnoses. For the purpose of this document, we consider microbiological typing data as a separate category from clinical data.

Epidemiological data.

This data contains different elements from typing data and clinical data that may be of importance for epidemiological analyses, such as age, sex, place of residence, history of travel, contacts and food consumption, habits and risk factors.

National/supra-regional signals.

In the context of lab surveillance, these are signals that geographically cover two or more MHS regions, and/or have an international component, and/or – in case of zoonoses – have a link with samples of non-human origin.

Signal.

In the context of lab surveillance, this is a single sample or cluster of samples with more genotypic and/or phenotypic similarities than expected and/or within a particular time period and/or geographic region and/or within a particular age/gender category, and with possible impact on public health. Long-term trends are also considered to be signals.

Signal detection.

See signal.

Signal follow-up.

The process of following up on detected signals, including the decision to take or not take further prevention and control actions and conducting scientific research for further clarification.

Typing data.

This data covers the results of all possible genotypic and phenotypic tests that are performed in a microbiology laboratory to type possible pathogens in a sample. This includes sequencing, as well as what is often called detection and identification, such as PCR tests, antigen tests and serological tests. The latter can also be referred to as diagnostic data since they are often used for patient care. For the purposes of this document, however, there is no need to distinguish between both.

Abbreviations

Abbreviation	Name
ADaM	Analysis Data Model
ATC	Anatomical Therapeutic Chemical
BPMN	Business Process Modelling Notation
CDASH	Clinical Data Acquisition Standards Harmonization
CDISC	Clinical Data Interchange Standards Consortium
CIb	RIVM Centre for Infectious Disease Control
COVID-19	Coronavirus disease 2019
EHEC	Enterohemorrhagic E. coli
EHR	Electronic Health Record
EPI	RIVM-CIb Centre for Epidemiology and Surveillance of Infectious Diseases
EU	European Union
GDPR	General Data Protection Regulation
HL7	Health Level 7
ICD	International Classification of Diseases
IDS	RIVM-CIb Centre for Infection Research, Diagnostics and Laboratory Surveillance
LCI	RIVM-CIb Centre for National Coordination of Communicable Disease Control
LIMS	Laboratory Information Management System
LOINC	Logical Observation Identifiers Names and Codes
MedDRA	Medical Dictionary for Regulatory Activities
MERS-CoV	Middle East respiratory syndrome coronavirus
MHS	Municipal Health Service
MML	Medical Microbiological Laboratory
NVMM	Dutch Society for Medical Microbiology
RIVM	Dutch National Institute for Public Health and the Environment
SARS	Severe acute respiratory syndrome
SDTM	Study Data Tabulation Model
SDTMIG	Study Data Tabulation Model Implementation Guide
SME	Subject matter expert
SNOMED-CT	Systematised Nomenclature of Medicine - Clinical Terms
STEC	Shigatoxigenic Escherichia coli
TAUG	Therapeutic Area User Guide
VMML	Dutch Association of Medical Microbiological Laboratories
VWS	Dutch Ministry of Health, Welfare and Sport
WHO	World Health Organization

Published by:

National Institute for Public Health and the Enviroment, RIVM
P.O. Box 1 | 3720 BA Bilthoven
www.rivm.nl/en
The Netherlands

february 2023

Committed to
health and sustainability