TOWARDS A DATA VISION AND STRATEGY @SCIENSANO

A data maturity assessment and implementation plan

—

Data strategy working group

.be

WHO WE ARE

—

Sciensano can count on more than 900 staff members who are committed to health every day.

As our name suggests, science and health are central to our mission. Sciensano's strength and uniqueness lie within the holistic and multidisciplinary approach to health. More particularly we focus on the close and indissoluble interconnection between human and animal health and their environment (the "One health" concept). By combining different research perspectives within this framework, Sciensano contributes in a unique way to everybody's health.

For this, Sciensano builds on the more than 100 years of scientific expertise.

Sciensano

Services of the managing direction **-** Strategy and external positioning
Data Strategy

Drafted: April 2022 • Brussels • Belgium
Adapted: September 2022 • Brussels • Belgium
Approved by: Board of directors

—

Strategy and external positioning

Contact person: Sofie De Broe • T+32 478 843194 • sofie.debroe@sciensano.be

TABLE OF CONTENTS

—

organisationorganisation

ABBREVIATIONS

—

AHEAD	Towards the development of a national health data platform
AHLICS	Animal Health Laboratory Information Centralisation System
AI	Artificial Intelligence
AirAllergy	Belgian Aerobiological surveillance network
BeBOD	Belgian National Burden of Disease Study
CESSDA	Consortium of European Social Science Data Archives
CINECA	Common Infrastructure for National Cohorts is Europe, Canada and Africa
CoZo	Collaboratief Zorgplatform
DIPoH	Distributed Infrastructure on Population Health
DMP	Data Management Plan
DPO's	Data Protection Officers
DXP	Digital Experience Platform
EAN	European Aeroallergen Network
ECMWF	European Centre for Medium-Range Weather Forecasts
EHDS	European Health Data Space
EJP RD	European Joint Programme on Rare Diseases
ELSI	Ethical Legal and Social Implications
EOSC	European Open Science Cloud
Epistat	Epidemiology of Infectious Diseases
EU HERA	European Health Emergency Preparedness
EUMETNET AutoPollen	European consortium on aerobiology
FAIR data principles	Findable, Accessible, Interchangeable, Reusable
FDP	FAIR Data Point
GDPR	General Data Protection Regulation
HDA	Health Data Authority
HISIA	Belgian Health Interview Survey Interactive Analysis
HRIC	Health Research and Innovation Cloud
IANPHI	International Association of National Public Health Institutes
IMI	Innovative Medicines Initiative
InfAct	Joint Action on Health Information
ISMS	Information Security Management Systems
LIMS	Laboratory Information Management Systems
NPHI	National Public Health Institutes
PHI	Public Health Institutes
PHIRI	Population Health Information Research Infrastructure
RDM	Research Data Management
REDCap	Research Electronic Data Capture
SCRA	Small Cell Risk Analysis
SOP	Standard Operating Procedure
SPMA	Standardised Procedures for Morality Analysis
TDI	Treatment Demand Indicator
TEHDAS	Toward the European Health Data Space
TOP	Technology Organisation Process

TTP	Trusted Third Party
UXP	User Experience Platform
VDI	Virtual Desktop Infrastructure

INTRODUCTION

—

1. What is a data strategy and why is it important?

Data are playing an increasingly important role in running our daily lives, the public sector and companies. The potential of data for providing policy-relevant information is large. At the same time, the amount of irrelevant data is also increasing. This makes it an even greater challenge to properly protect the valuable data, to manage it and to distinguish it from 'noise'. Public institutions and companies are therefore forced to make choices about what data they (want to) manage and analyse, how they protect and share the data, what information they want to make available in the future.

A data strategy is a highly dynamic process employed to support the acquisition, organisation, analysis, and delivery of data in support of business objectives. Specifically for Sciensano its data strategy is a long-term plan that describes how Sciensano will realize the FAIR data principles, which ensure that our data is findable, accessible, interchangeable and reusable; what steps Sciensano will take to make its data available and manageable; how Sciensano will generate and use new data streams and information in the future to fulfil the mission of the organisation. All this depends on how data mature the organisation is on various indicators of interoperability (1); the purpose of our data; the future vision of Sciensano in terms of the role it wants to play in the data landscape and the extent to which it wants to innovate as part of its future mission.

The development of a data strategy is therefore essential. As Sciensano generates and uses an ever-increasing amount of data and data is its biggest asset, a data strategy is a necessity. The data strategy project started in our institute in late 2021 and will be implemented in 2022. This follows the evolution in Europe which has taken a strong position in the data landscape and has her own data strategy.

Not having a data strategy risks undermining an organisation's reputation in terms of trustworthiness, transparency, quality and security in dealing with (personal) health data; secondly it also reflects insufficient attention to FAIR principles; thirdly it means missing opportunities by insufficiently addressing future data potential for the goals and vision of the organisation (2). A solid and well-designed data strategy ensures that the data Sciensano uses, generates and stores is valued to the best extent during its life cycle. In the absence of a data strategy and data governance, data is scattered, unidentified, invisible to most people because of unilateral access: access is difficult, random, undocumented. Sciensano's data strategy must include a set of decisions that formulate a high-level framework that facilitates the leveraging of data from a holistic view of the data: storage, access, security, processing environment, back-up, ... It must adhere to the legal role of Sciensano and meet user's needs: allow the use of data in an operational and optimized ecosystem. It must ensure visibility, accessibility and traceability as well as the quality and description of the data to guarantee their valorisation-reuse. The data strategy does not incorporate the data related to for example financial performance indicators or human resources.

1.1. MOTIVATIONS FOR A SCIENSANO DATA STRATEGY

There are many good reasons why Sciensano needs to have a vision and data strategy. In addition to the arguments stated previously, the following reasons can be added:

- There is more demand and supply of data than ever before but people/researchers face hurdles accessing the data they need for research and policy advice;
- Sciensano's expertise lies in surveillance, monitoring and research of human health, animal health and the environment from a One Health perspective which requires data collection and linkage;
- There is increasingly a need for transparent, factual, real-time, detailed and evidence based (public) health advice and information. In order to meet this information need, Sciensano must be able to collect and/or receive/share data that meets our output finalities (f.e. data must be accessible in a timely manner);
- Many organisations and researchers want to work with Sciensano so there is the need to share the data bases and be able to link different databases from internal and external researchers;
- A solid data strategy also gives Sciensano credibility in its expertise position in regard to the data that it generates. Well managed databases often brings data providers to the centre of scientific collaboration projects/networks.
- Data are spread across different organisations in the health system in Belgium and a more cost-effective way of sharing data is needed. Sciensano's data will have to fit in a future « Belgian » health data governance structure;
- There is increasingly a need for data science and data engineering to be embedded in organisations to deal with increasing amount of (large) data sets;
- Covid has made issues around data governance and data management very obvious and the data has exploded in just 1,5 years. Digital health has been accelerated by Covid-19, but also sensitivities regarding sharing and re-use of data.

There are also clear trends in the world around us:

- There is the creation of a European Data Space as one of the priorities of the Commission 2019-2025, including the health sector (3). A common European Health Data Space (EHDS) will promote better exchange and access to different types of health data (electronic health records, genomics data, data from patient registries etc.), not only to support healthcare delivery (so-called primary use of data) but also for health research and health policy making purposes (so-called secondary use of data). It is under the EHDS that there is the demand for a Health Data Authority (HDA). The entire data system will be built on transparent foundations that fully protect citizens' data and reinforce the portability of their health data, as stated in article 20 of the General Data Protection Regulation;
- The European Data Governance Act aims to strengthen various data-sharing mechanisms and to promote the availability of data that can be used to power applications and advanced solutions in artificial intelligence (AI), personalized medicine, green mobility, smart manufacturing and numerous other areas.
- There is a clear trend in the world around us towards Cloud. The EU has its own policy with regards to cloud computing as we can find on this website;
- Open science has become the norm for all subsidies and grants at EU and federal level;
- High Performance Computing, Artificial Intelligence, latest technologies and methods in the area of Privacy Preserving Techniques are relevant for Sciensano's research activities;
- The EU has made a lot of funding available for the Digitalisation of Europe in order to make more use of available big data sources and new methods and techniques. Some of the topics are of great relevance to the (public) health research activities of Sciensano and are part of Sciensano's data strategy and management plan;
- With the implementation of the GDPR many organisations are moving towards federated or distributed analysis. This means the data does not move out of the institute managing it but becomes available locally and can be analysed locally in a safe space for the researcher. This requires technological developments and change in data handling procedures.

1.2. ENABLING AND EXISTING INITIATIVES

Some of the existing initiatives, projects which Sciensano coordinates or in which we participate, and enabling situations for the data strategy at Sciensano are given below. The further development of these initiatives will build on our main vision/strategy and will serve the needs of our external stakeholder for (public) health policy making and research. However, Sciensano is (partially) dependent on collaborative partners for data sharing/delivery and for research developments.

- Sciensano plays a key role at international level in shaping the future of Europe with regards to the secondary use of health data through the activities of the EU Health Information System Unit. Sciensano is a driving partner in the European Health Data Space, the European Open Science Cloud and the Health Research and Innovation Cloud. Sciensano facilitates the exchange of health information across Europe by leading and participating in seven major European projects in health information. It has built its expertise in European data governance mechanisms and has created a wide network across Europe through projects such as Joint Action on Health Information (InfAct) (2018-2021), Population Health Information Research Infrastructure (PHIRI), the European Population Health Information Portal, Distributed Infrastructure on Population Health (DIPoH), Joint Action Towards the European Health Data Space (TEHDAS) supporting the implementation of the European Health Data Space (EHDS);

- HealthyCloud examines the feasibility of sharing individual level health data across different European countries for research purposes and identifies challenges and options to overcome these. This brings extensive knowledge on FAIR principles in order to identify the best way to evaluate the FAIRness level of a data infrastructure;

- EGI-ACE aims to empower researchers from all disciplines to collaborate in data- and compute-intensive research through free-at-point-of-use services. EGI-ACE works closely with European Open Science Cloud (EOSC) to provide storage facilities, technical services, analytics tools and support;

- BY-COVID aims to mobilise and connect well-established data resources in line with the FAIR principles and deliver access to heterogeneous yet interlinked and organised data across scientific, medical, public health and policy domains. This will be done via the COVID-19 data portal while engaging with the Horizon Europe European Open Science Cloud (EOSC) and European Health Data Space (EHDS);

- The Belgian aerobiological surveillance network (AirAllergy), managed by Sciensano, has agreed on a European cooperation with a number of national aerobiological networks in order to share scientific research data on the presence of allergens in the outdoor air. Historically, Belgium is a notable founding member of this agreement which gave birth to the European Aeroallergen Network (EAN) in 1987;

- Sciensano is involved in a European consortium on aerobiology (EUMETNET AutoPollen) that aims to define from-the-start the main standards for the new era of automatic real-time bioaerosol monitoring;

- BelCoVac Consortium: in the context of the COVID-19 vaccines which are currently marketed or will soon be marketed in Belgium, the consortium institutions wish to participate in a consortium program in cooperation with and under the auspices of Sciensano with the purpose of organizing several non-commercial clinical trials and follow-up research projects in respect of said COVID-19 vaccines in order to generate information and data that might be useful to patients, the clinical practice and/or policy decision makers

- Sciensano works on the consolidation of his current Laboratory Information Management Systems (LIMS) and several projects aiming to improve the exchange of LIMS data with external partners and networks and make LIMS data more easily accessible to internal users

- Sciensano is a central player in the Belgian health information landscape
  1. Numerous surveys including the Belgian Health Interview Survey, food consumption survey, etc.
  2. Active follow-up of the COVID-19 pandemic through existing and new data sources
  3. Coordination of the Focal Point International Data Collection
  4. Platform HealthData.be for the technical facilitation of scientific projects on health and health care (see more information below)
  5. Lead of the HERA BE Incubator project for enhancing the national infrastructure for genomic and epidemiological data exchange in Belgium

- Sciensano explicitly wants to support the scientific valorisation of the Belgian health information landscape
  1. BELCOHORT has actively explored the possibilities of setting up a cohort in Belgium, and has formulated a series of recommendations together with the parties involved
  2. AHEAD will proactively enhance the visibility of the Belgian health information landscape and will explore the technical, legal and ethical conditions to be able to work towards a national health data platform to improve the accessibility to the different data sources and to study more complex questions (e.g. relation between health and socio-economic indicators)
  3. Sciensano plays a crucial role in the setting up of a Belgian HDA

- Sciensano focuses on creating links between different sources:
  1. HISLINK studies the quality of health care data through links between the health survey and the health care data of the Inter Mutualist Agency
  2. HELICON studies social inequalities in COVID-19 and the indirect and long-term effects of the corona crisis through links between Sciensano COVID data and Statbel demographic data, and Intermutual Agency health data
  3. TDI-IMA studies the use of health care and medication among people in treatment for drug use by linking the Treatment Demand Indicator (TDI) register to the Intermutual Agency health care data.
  4. SPADIS will study the employment and social participation of people with chronic diseases. For this purpose, it will link existing surveys and registers of Sciensano with the Crossroads Bank and health care data of the Intermutual Agency. A similar study of cancer patients (Cancer Register) has already been carried out.
  5. LINK-VACC: linking vaccination registers with hospital data, testing data, as well as external data sources such as COHBRA.

- There are several initiatives that Sciensano supports or hosts that make data Findable and Accessible (the F and A of the FAIR principles) into practice
  1. Healthdata.be maintains the fair.healthdata.be portal with an overview of health and healthcare data. This project originated as Action Point 18 of the eHealth plan 2013-2018: an inventory of all available health and healthcare registry and surveillance data in Belgium. Depending on the data source, there is information on purpose, licence, variables,... Besides the possibility of searching for data, the portal also offers an API for integrating the available information in other systems. Adding and maintaining the information is the responsibility of the data owner.
  2. https://data.gov.be is part of the federal open data policy with the aim of making governmental data more open and making it available for reuse by citizens, researchers, businesses and governments. The website itself reuses metadata from a series of domain-specific portals. For health data, the information is currently relatively limited (a.o. the COVID-19 data sources via fair.healthdata.be, Statbel data sources, ...).

2. Data vision Sciensano

A data strategy should be preceded by a data vision. Below follows a brief data vision for Sciensano. It consists of the legal frameworks within which Sciensano has to operate and its vision for the future on data governance. These data can be owned by Sciensano or by other parties and serves to fulfil Sciensano's role as a scientific institute in delivering its mission (4). The data strategy concerns *all* data that are at Sciensano (including non- health data).

In carrying out its tasks, Sciensano sees in the 21st century the emergence of (new, large) data sources and the rapid development of new methods, tools and technologies such as AI for using both existing (survey, register and census data) and new types of data sources (big, real world, open data). As an almost insurmountable result, these new and existing data sources are scattered across different health (care) systems in different organisations and countries. At the same time, policy and decision makers, researchers, entrepreneurs and the general public want to have high-quality information on (public) health issues and large initiatives are being set up (such as e.g. the EHDS). Citizens need transparent and factual information, governments and businesses are looking for up-to-date, accurate and detailed information so they can design their policies and governance in an evidence-based manner, and scientists use the data as the basis for their research.

Traditionally, Sciensano has undertaken its tasks along two lines. First, Sciensano monitors health through its monitoring and surveillance networks, platforms and labs and provides advice to policy makers on (public) health issues. Secondly, Sciensano undertakes research to deliver evidence based (public) health advice. In order to enable Sciensano to deliver this advice and undertake research, health information is collected through surveys and monitoring systems, and access to certain registers. Getting access to data is often a time intensive and complex activity due to the absence of transparent and standard procedures where no separate data entry, validation and collection unit exists. The re-use of health (care) data collected by Sciensano and other government agencies is certainly a data collection path that should be further prioritized, structured and governed. There is a need to lobby for changes in the laws in terms of data access and protection, data stewardship and separation of data management roles. There is also a lot to gain if the monitoring of hospital, laboratory, general practitioner data could be a structured network of data delivery that is adapted to Sciensano's output requirements (with respect to the burden of the data providers e.g. syndromic health surveillance). Finally, the roles of the different Belgian health data holders (healthdata.be, Statbel, Sciensano, Riziv…) should be specified. The creation of a Belgian HDA would offer an ideal opportunity for this.

In these new times, Sciensano wants to (re)define its role and position in an increasingly data- and technology driven society respecting deontological and ethical regulations (see strategic objectives of the management plan of Sciensano). In its management plan, Sciensano has expressed its ambition to increase its impact for health policy makers by making use of available (open) data, adopt an Open Science policy and deal with data using the FAIR principles (Findable, Accessible, Interoperable, Reusable). FAIR data starts at the data collection stage and FAIR principles are included in agreements and contracts with data providers (Can the data be reused? Can the data be open?) and collaborators. With the expansion of the data interchange, Sciensano wants to promote the visibility and accessibility not only of its own data but of Sciensano's data in combination with data from other departments/organisations. It is the ability to combine and link data that is in great demand and creates the added value of Sciensano's data for research and policy making. At the same time, Sciensano wants to make its own processes more efficient and transparent as stated in its management plan.

This section presents the Data Vision of Sciensano, which describes the desired role of Sciensano in a national and international data ecosystem in more detail.

2.1. LEGAL FRAMEWORKS

Sciensano was created by the law of 25 February 2018. Among the legal tasks conferred by this law are the processing of data as described in articles 4, § 4 and § 4/1 of the law:

" § 4. Sciensano staat, met inachtneming van de ter zake toepasselijke wetten, in voor de behandeling, daarin inbegrepen de verzameling, validering, analyse, rapportering en archivering van gegevens van persoonlijke aard, met name met betrekking tot de volksgezondheid of in verband met de gezondheid en andere wetenschappelijke informatie met betrekking tot het gezondheidsbeleid. Sciensano maakt daartoe kwantitatieve en kwalitatieve wetenschappelijke analyses op basis van de verwerkte informatie ter ondersteuning van het gezondheidsbeleid. Sciensano kan ook verwerkte gegevens en informatie ter beschikking stellen met toestemming van de bevoegde sectorale comités.
[1 § 4/1. Sciensano heeft in het kader van de beheersing van volksgezondheidscrisissen als opdracht om de wetenschappelijke aspecten die hieraan verbonden zijn te coördineren en uit te voeren, de risico's te bewaken en deze te evalueren door specifieke analyses van de verzamelde gegevens, adviezen en aanbevelingen te verstrekken aan de verschillende gezondheidsautoriteiten van het land, en de communicatie te organiseren ten behoeve van de overheden, de zorgverstrekkers en het publiek.]"

As a public law institution, Sciensano is subject to various laws concerning data:
- Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European data governance (Data Governance Act):
- REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation):
- DIRECTIVE (EU) 2019/1024 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 20 June 2019 on open data and the re-use of public sector information. The directive has already entered into force and must be transposed into national (Belgian) law by 17/07/2021 at the latest. This deadline has not been met.
- 4 May 2016. – the law concerning the reuse of governmental data
- 5 AUGUST 2006. – the law concerning access of the public to environmental information
- 11 APRIL 1994. – the law concerning the openness of government
- 2 JUNE 2019. — Royal decree concerning reuse of governmental information

Sciensano falls within the scope of the two laws that apply cumulatively to our institute: the law on Open Government (11/04/1994) and the law on the re-use of Public Sector Information (known as "Open Data" (04/05/2016)). The scope of the Law on Open Government is very broad. The law was drafted to ensure the transparency of information held by public authorities. It provides for a right to information. The exceptions that can be invoked to refuse a request for information are exhaustively listed in the law. It must be noted that when the law was drafted, there were no research institutions similar to Sciensano in terms of its status.

The scope of application of the law on the re-use of public sector information is different, as it focuses on the re-use of public sector information for a purpose other than that for which the information was collected. The legal exceptions provided for in this Act differ from those that exist for the Government Information (Public Access) Act. The application of both laws work disadvantageous for Sciensano. It would be sufficient to invoke one of these two laws to demand information. A draft law is currently being prepared to replace the 'Open data' law. The aim is better protection of research data. Unfortunately, this bill does not (at this stage) prevent citizens from invoking the 1994 law.

The current legal situation can be called the "perfect storm" as many laws are in need of updating: the law on patient rights; the quality law (electronic health records should be available for each healthcare provider); the law on cooperation between the regions for data sharing; the privacy law. There are also several laws that are upcoming: law on data analysis (amount of health data, organisation of health data, process, objectives for the future) and there is the Data Governance Act which mentions the requirement to set up an information security committee.

2.2. MAKING USE OF DATA INSIDE AND OUTSIDE OF SCIENSANO

Data sources inside and outside of Sciensano

Sciensano sees the importance of data linkage between existing (health care) data and the emergence of new (large/open) data sources with a rapid development of new methods, tools and technologies to create the added value for health (care) research and policy relevant output. The data that Sciensano possesses is diverse. The data strategy concerns *all* data that are at Sciensano (including non- health data). Below we give a non-exhaustive overview of the data sources and systems that give access to data:

- Healthdata.be
- Surveys: see this link for an overview of projects that collect survey data
- Covid monitoring and surveillance using data inside and outside of Sciensano:
  1. Health data and PCR /antigen tests
  2. Collect data on Hospital surge capacity surveillance and clinical surveillance
  3. Collect data on incidence/mortality in care homes
  4. Surveillance at GP practices
  5. Surveillance at the workplace in terms of absence
  6. Contact tracing through IFC testing and tracing
  7. Cluster detection with data from several instances
  8. Passenger Location Forms for returning passengers
  9. Monitoring vaccination uptake and degree
  10. Monitoring effectiveness and safety of the vaccination
  11. Monitoring of building up of anti-bodies again SARS-COV-2 (sero-prevalence studies among different groups: blood donors, children, …)

- Other data systems used at Sciensano
  1. Laboratory Information Management Systems (LIMS)
  2. REDCap (Research Electronic Data Capture)
  3. SLIMS: Laboratory Information Management System (LIMS) combined with an Electronic Laboratory Notebook (ELN).
  4. BIOTECH platform
  5. TEAKNET
  6. ORPHANET
  7. Databases with chemical information: e.g. Scanner project/ New psychoactive substances:
  8. EAN
- Open data:
  1. https://epistat.wiv-isp.be/covid/covid-19.html
  2. https://fair.healthdata.be/sources/covid19
- Interactive databases
  1. https://www.healthybelgium.be/en/health-status/about-the-health-status-   report/interactive-databases-sciensano
  2. https://hisia.wiv-isp.be/SitePages/Home.aspx
  3. https://www.sciensano.be/en/projects/standardized-procedures-mortality-analysis/spmah

Next to own data Sciensano has access to and uses other (open) data such as:
- Biodiversity data

- NEVO
- NUBEL
- Food environment: Locatus data
- Air pollution data:
- Land cover and remote sensing data
- Belgian Geographical Institute
- Regional geographical data repositories:
  1. https://geoportail.wallonie.be/home.html
  2. https://leefmilieu.brussels/ & https://geobru.irisnet.be/
- https://www.geopunt.be
- Meteorological data

The above overview is not complete and it will be part of a data strategy to have an inventory of all data at Sciensano.

The desired role of Sciensano in a national and international data ecosystem

The figure below gives a first outline and exploration of the different types of activities that Sciensano performs making use of data and the types of data that are available for these activities. By law, Sciensano is not permitted to use personal microdata for purposes that are not related to scientific or policy advice; all data activities that are **not** covered by the legal framework where Sciensano is operating in, fall in the **1st quadrant.**
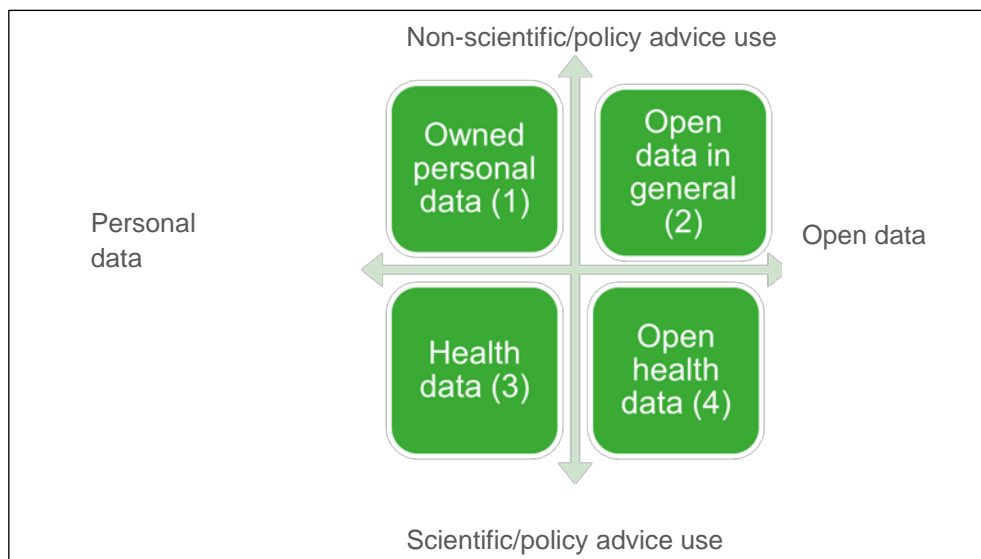
The data cited in the **2nd quadrant** are not provided by Sciensano but by other parties such as the government and private companies. The data are made available by other parties than Sciensano as open data and may contain quantitative data, but also other data, such as data on topographic maps etc. They may also contain data on individual companies, persons or institutions.

Sciensano's activities are currently situated in the lower two quadrants. The output of policy and research reports (health information, advice and scientific research) of Sciensano is in the **4th quadrant**, where the open data are derived from the microdata in the **3rd quadrant**. The microdata (or sometimes

raw data)[1] for science based policy advice and research are in the 3rd quadrant. A reasonably common terminology of data use is:

- Secure use: Confidential microdata that can be used within Sciensano.
- Scientific use: You can offer data to externals in a secure and controlled way. It should be clear if the data 'leaves' the organisation of stays inside.
- Public use: Open data, safe and not confidential.

Secure use and scientific use belong to the third quadrant and public use to the fourth quadrant.
Below we provide a description and an ambition Sciensano has in relation to the data and its use for each quadrant. The data in the **4th quadrant** are aggregated and not traceable to individual persons (in the case of companies or institutions the GDPR does not apply unless natural persons working at these are identifiable, but other economical laws apply), and made accessible for use as open data. The content in this quadrant is provided by Sciensano. This content consists of publications in report form or interactive tools or dashboards of results from analysis of surveys, registers or open data as end products of custom made advice or scientific research for external parties. The health users (policy makers, citizens, researchers) of this quadrant sometimes need (quick) information on a wide range of health subjects. It is the ambition of Sciensano to satisfy this need as much as possible and to take steps forward in this direction from its current services in order to address even better current policy questions by making data openly available.

The **3rd quadrant Health Data** contains microdata that may be used for advisory and scientific purposes and are therefore made accessible to internal and external (to Sciensano) scientists under strict conditions (Secure Use). We distinguish several types of internal use of the microdata:
- to describe population health. As such the data collection needs to comply with certain aspects such as representative sample of the population – this is the main goal of the monitoring/surveillance of health;
- to examine associations between risks and health status. Since the majority of the data collections is cross sectional, no conclusions on causal interference can be made. The above described projects (BELCOHORT, AHEAD…) are steps towards longitudinal data collection and research to allow for causal inference;
- to support public health advices: e.g. integration of distal determinants as enabling or limiting factors, to examine life course impact of lifestyle/diseases on health and/or social participation, to estimate the impact of potential policy measures on health outcomes;
- to predict the future trends in health and disease in humans and animals, etc…

Sciensano's ambition is to make data findable and accessible by helping researchers link the open data of Sciensano (4th quadrant) with the open data in the 3rd quadrant. Here Sciensano can apply its knowledge of data infrastructure, data linkage, metadata management and privacy protection. In this way, Sciensano would contribute to a non-fragmented Belgian landscape with well-described, usable data that does not reveal individual information, by combining data sources from different health administrations.

---

[1] Microdata are sets of records containing information on individuals, households or businesses. Raw data (sometimes called source data, atomic data or primary data) is data that has not been processed for use.

COMPONENTS OF A DATA STRATEGY FOR SCIENSANO

—

For the different components of the data strategy we follow the subdivisions of the data maturity assessment publication (1). According to the proposed approach four different types of interoperability are used as a base for the maturity assessment i.e

- Organisational: Organisational interoperability can refer to how data governance and data management functions that touch upon interoperability are distributed across an organisation. It can also be viewed as the ways in which whole organisations, or individual departments within them, engage with the broader data ecosystem to decide the degree of interoperability they would like to achieve between their collective data assets.
- Human: Human interoperability can refer to the need to ensure that as data is made interoperable across data systems, it remains readable and usable by human users. It also refers to the ability of individuals, groups, teams, and departments of individuals within organisations to be able to communicate and work together in ways that foster interoperability across their work streams and data assets
- Data: Data interoperability is the need to ensure that data systems and data sets are designed to enable interoperability ensuring that they can collate, store, and process data in machine-readable formats and that interoperability needs are considered when data are modelled and classified.
- Technological: Technological interoperability requires appropriate data infrastructure to enable data, human, and organisational interoperability in meaningful ways.

Each of the interoperability dimensions encompasses several indicators of maturity and to each of these indicators a maturity level is assigned, ranging from undefined, to emerging, learning, building and consolidating. The maturity levels are well described and the goal is to identify at what level Sciensano is situated.

For the assessment, an internal working group was set up within Sciensano with representatives of all the scientific directorates. Each indicators of maturity was discussed and the level of maturity for Sciensano was determined. The same exercise was repeated to reach an agreement on the level of maturity Sciensano wants to reach in the next 5 years. The summary and explanation of maturity levels of the current and future data maturity assessment for Sciensano is available in Annex 1 and 2. At Sciensano healthdata.be is located under the directorate biological health risks[2]. Data maturity of healthdata.be is generally higher than for the data in the rest of Sciensano (referred to as Sciensano in the text); for the benefit of this exercise the distinction with healthdata.be will be made in some places.

---

[2] The platform healthdata.be, developed by Sciensano (formerly the WIV) and financed by the National Institute for Health and Disability Insurance, offers new perspectives on e-Health by simplifying the registration and storage of health data sent by various healthcare providers. Healthdata.be eventually ensures better quality health research. In 2006, the Federal Health Care Knowledge Centre (KCE) had access to almost 50 health data databases which could be accessed by the scientific community on the KCE website. The aim then was to give researchers quick access to the information they needed for their research in the various fields of public health, health care and epidemiology. By joining forces as part of a new inventory project, the KCE and Sciensano succeeded in creating 150 health data databases. Healthdata.be is governed by the steering committee of the healtdata platform, the privacy commission, the working group architecture of the e-health platform and the internal guidance committee if healthdata.be.

1. Organisational interoperability

1.1. STRATEGIC OBJECTIVES

"Data management can loosely be defined as the development, implementation, and monitoring of strategies and plans that allow the value of data to be safely unleashed… For organisations that manage, collate, or process diverse data sets, interoperability should be a key tenet in their data management strategy… When setting strategic objectives relating to interoperability, it is important to allocate budgetary and human resources for them in accompanying implementation plans…" (1).

Healthdata.be is located in the directorate biological health risks; the data in the rest of Sciensano are managed by the scientific directorates supported by ICT and the Laboratory Information Management Systems (LIMS). Healthdata.be achieves high levels of interoperability and a close collaboration between scientific directorates and healthdata.be is part of the Sciensano data strategy. It is Sciensano's ambition to adhere to the FAIR data principles.

For Sciensano as a whole, as of April 2022 there were no strategic objectives relating to interoperability nor are there budget or human resources allocated. Therefore, in terms of strategic objectives, for healthdata.be, ICT, and the LIMS we consider that Sciensano is at the level of Learning. For data outside of these systems, Sciensano is at the level of Emerging.

1.2. LEADERSHIP AND MANAGEMENT

"Leadership is crucial to data management … and ensuring that interoperability is treated as a strategic objective. In order for organisations to realize their data assets' potential, senior management must be engaged in strategic processes that relate to data interoperability and be aware of, and understand, the value that interoperability brings to their organisations…" (1).

The COVID crisis has risen the awareness among senior managers of the value of interoperability of data, the value of data assets and the need for a data management plan. In terms of leadership and management, Sciensano is in between Emerging and Learning.

1.3. OVERSIGHT AND ACCOUNTABILITY

"Oversight and accountability structures are central tenets of data governance, and to management functions relating to the setting of strategic objectives, leadership and management, and data stewardship. Ensuring that there is effective oversight and accountability built into organisational processes helps to ensure that the processes of making data interoperable — and therefore integratable, sharable, and accessible to others — are properly managed and maintained. There also needs to be oversight of data sets that may contain data that could reveal an individual's identity or other sensitive attributes about individuals or vulnerable groups" (1).

Oversight and accountability structures such as the steering committee, processes of making data interoperable and data stewardship are in place at healthdata.be in order to ensure data governance and oversight of data sets that contain sensitive attributes. For the data at Sciensano, a data inventory is undertaken as part of the data strategy. This will reveal where data is located, who is responsible for the data, who is the data owner (if there is one) or point of contact and whether meta data are available. In different units initiatives are taken to allow oversight and Data Protection Officers (DPOs) and information security coordinators are involved in ensuring that sensitive data is secured.

In terms of oversight and accountability, Sciensano has a level in between Emerging and Learning.

1.4. LEGAL COMPLIANCE

"Legal and regulatory challenges can arise for interoperability when it comes to the sharing and integration of data assets between organisations and across national borders. Laws set enforceable

boundaries of what is acceptable conduct and what is not. In some instances, they govern how data can be shared (for instance, laws that regulate and set standards for data reporting, security, and protection) and in others, they govern what data can, or more often cannot, be shared and integrated (for example, data protection and privacy laws)…" (1).

Article 4, §4 of the Act establishing Sciensano provides the legal basis on which Sciensano is authorised to process personal data. Sciensano must hereby comply with the applicable laws, including the principle of minimum data processing provided for in the GDPR and (Title 4) of the Act of 30 July 2018 on the protection of natural persons with regard to the processing of personal data. Only when it is not possible to achieve the research or statistical purpose with a processing of anonymous data, Sciensano may process pseudonymised data; and only non-pseudonymised data if it is not possible to achieve the research purpose with pseudonymised data. The processing of non-pseudonymised data must therefore always be justifiable.

Article 4, §4/1 of the Act establishing Sciensano (I) specifically concerns the role of Sciensano (and any processing of personal data associated with it) in the management of a public health crisis. The Law of 4 May 2016 on the reuse of government information and the Royal Decree of 02/06/2019 regulate the reuse of government information and explains the federal open data strategy. Sciensano is GDPR compliant and also develops software tools in order to do so.

Sciensano has a legal department that makes sure that Sciensano complies to the laws under which it falls. In terms of legal compliance Sciensano has a level between Learning and Building.

1.5. DATA ETHICS

"Interoperability can give rise to ethical dilemmas. These differ from legal issues in that they are likely to relate to questions of equity or what is considered right or wrong in a particular context, as opposed to lawful or unlawful. An ethical interoperability issue might be whether international data processing firms that obtain access to national demographic microdata as part of a development program should have a right to reuse and profit from that data in the future by integrating it with other data sets. Depending on the national laws of a particular jurisdiction, this might be legal, but it still raises ethical questions about the equitable distribution of the value extracted from national data assets…" (1).

Several activities take place at Sciensano that address ethical issues of (personal) data handling and processing. Within the Joint Action TEHDAS (Towards a European Health Data Space) and the Belspo project AHEAD (Towards the development of a national health data platform), an online consultation was launched to inform citizens about ethical issues related to health data reuse and to give them the opportunity to co-create an ethical and societal framework for health data spaces. Within BY-COVID, a Horizon Europe project, Sciensano is heading the citizen engagement initiative on informing citizens about health data use during the pandemic and health data platforms that are created to improve pandemic preparedness.

The maturity level of Sciensano on data ethics is Learning.

1.6. PROCUREMENT

"The procurement of IT services, data systems, or other solutions that relate to the processing, storage, or sharing of data all raise important interoperability issues for organisations… Vendor lock-in occurs when organisations separately procure different proprietary digital and data solutions as part of different programs of work or donor-sponsored projects without considering existing data architectures or digital infrastructure. This can create siloed information systems that produce systems or data sets that do not interoperate or whose data outputs cannot be integrated with each other" (1).

At Sciensano, the scientific activities are supported by an IT infrastructure with on-premise installations, secured network, high performance computing and well-chosen storage solutions. On top of this

infrastructure, a virtual IT infrastructure (aka Virtual Desktop Infrastructure or VDI) is hosted. The IT department can expand this infrastructure with more memory, processor power, etc... according to the individual units needs and performance requirements without being limited by local hardware requirements. In addition, the VDI platform is secured and accessible for authorized persons only, and it allows intensive and high performant analysis.

Procurement is done for 2 different purposes. The first purpose contains the common IT infrastructure and software needs. Both of them are monitored, and the ICT department foresees yearly growth, evolution and expansions. The second purpose covers the on-demand scientific projects with their own requirements. They are analyzed case by case to select the best and most cost-efficient and performant solution. The mission and vision of a common ICT infrastructure is of course applied, but for example for storage it's important to invest in tailored technology.

The level of maturity on procurement for Sciensano is between Undefined and Emerging.

1.7. LINKS TO BROADER DATA ECOSYSTEM

"Many other government departments, international organisations, private sector actors, and non-governmental organisations — can all be viewed as part of a common data ecosystem, a network of interconnected organisations, data systems, digital infrastructure, and applications. How organisations choose to engage with other entities within their own, or overlapping, data ecosystems is a strategic decision with important interoperability considerations. … The agreements organisations put in place with other parts of the data ecosystem are especially important in ensuring that any planned interoperability is strategically thought through and documented. Data licensing agreements — legal documents that specify what a user is allowed to do and not do with data — are one type of tool for doing this" (1). Another paper (2) highlights the three prerequisites for building a robust health and healthcare data sharing ecosystem: the country's familiarity with managing data, data sharing arrangements and data sharing technology; the prevalence of the use of electronic medical records systems and the presence of a national-level body that is responsible for the development of the health and healthcare data sharing ecosystem.

Universities
In July 2021 Sciensano signed a Memorandum of Understanding with the Walloon universities Catholic university of Louvain, Free University of Brussels, University of Liège, University of Mons, University of Namur and the Flemish universities Catholic University of Leuven, University of Antwerp, University of Ghent, University of Hasselt and free University of Brussels. The MOU stipulates that the common areas of research include, but are not limited to, research on well-being and health. The partners agree on a mutual exchange of data and results with the common objective of scientific advances in the field of in the field of human or animal welfare or health.

Healthdata.be
In Belgium, the health care sector has numerous actors who each apply their own working methods for the collection and registration of health data. This is reflected in a very heterogeneous collection of data, both in quantitative and qualitative terms. The healthdata.be platform makes it possible to standardize and homogenize the registration and preservation of health data while guaranteeing absolute confidentiality in their dissemination for research purposes. By offering a concrete solution to the problem of the fragmentation of data in different health registers, healthdata.be contributes significantly to the development of a system infrastructure for (health) research in Belgium.

eHealth MetaHub
Via its integration with the Collaboratief Zorgplatform (CoZo) Sciensano medical lab results are shared on the eHealth MetaHub.

Health Data Authority

The need for a federal initiative that can organize and support the use and possible linking of health and health care data for scientific and policy-supporting work has been identified several times in the past[3]. In the General Policy Note on Public Health of 2 November 2020, the Minister of Health proposes the development of a HDA (GDA-ADS) to meet this need. The policy unit of the Minister of Public Health subsequently commissioned the FAGG-AFMPS, FODVVL-FPSSPSCAE, KCE, RIZIV-INAMI, and Sciensano to jointly develop a proposal for the healthcare data authority, in consultation with the eHealth platform.

European Health Data Space (EHDS)
Sciensano is actively engaging in the discussion on the EHDS. This is an important link to broader data ecosystems but also allows Sciensano to shape the EHDS. Through the TEHDAS project a better understanding is provided on the governance and data management systems in Belgium and European countries. By engaging in the development of the EHDS, Sciensano is co-creating the standards that will be developed for the Findability, Accessibility, Interoperability and Re-usage of the data. Sciensano will also participate in the development of the technology to be able to operate in a wider data ecosystem complying with Ethical Legal and Social Implications (ELSI) requirements.

In terms of links to the broader data ecosystem Sciensano is Learning.

2. Human interoperability

2.1. DATA STEWARDSHIP

"In the field of data management, data stewardship is the "practice of managing data assets on behalf of others and in the best interests of the organisation" (5). Effective and accountable data stewardship is key to interoperability. It is often data stewards who are charged with implementing procedures and practices that enable greater interoperability. Data stewards are often individual staff members, strategically appointed throughout an organisation, to oversee key data management functions in which they have substantive expertise. For instance, chief data stewards might chair data governance councils or committees, while technical data stewards might be IT professionals who oversee data integration and such in specific departments…" (1). A white paper (2) distinguishes different data roles in the data sharing ecosystem (health data generators, health data regulators and governing entities, health data exchange and IT service providers and health data users). It pleas for renaming data 'owners' data 'stewards'. Data 'owner' becomes less relevant when data is not anymore owned by the data generator but shared between stakeholders that have the responsibility of safeguarding and managing the data.

Healthdata.be plays an important role in making health data accessible in Belgium. Data stewardship can be improved when it comes to ensuring clear, transparent standard procedures, protocols and budgetary requirements in order to get access to the data. There is a need to clearly delineate the role and responsibility of different committees in governing data access between healthdata.be and (the rest of) Sciensano.

At Sciensano, heads of service are responsible for the data being used in their service. For some scientific departments and in some services there is one person who is point of contact (focal point) for the data sets being used in that service. For COVID data, separate data stewards were made in charge of the data management but no formal data owners roles were allocated to different data sets (often Sciensano is not the owner but the regions are). In general, the scientific directors are responsible that

---

[3] See for instance "Inventory of databases health care" - KCE report 2006 chapter 5; "Scientific support to the federal health policy" - Rekenhof 2010; "Data for better health – Conclusion note" - Policy Cell Minister of Health 2020; "Call and advice - Better government data for strong policy research and good governance" - SERV - VLIR 2021

the data within their directorates are adhering to the GDPR regulations. Within Sciensano two roles have been formally defined: the data protection officer and the information security coordinator.

In terms of data stewardship, Sciensano is Emerging. Taking into account the expertise of Healthdata.be, overall Sciensano is at the level of Learning.

2.2. PERSONAL DATA PROTECTION AND CONFIDENTIALITY PRESERVATION

"Establishing interoperability between systems requires a careful balance between determining what data can be opened up, shared, or integrated with other data sets and what data should not be. There are numerous reasons why an organisation might not want to make its data interoperable with other data sets, but one key reason is the desire to maintain confidentiality and protect the right to privacy that data subjects hold over their personal and sensitive data…" (1).

Personal data protection and impact assessment
Sciensano has also put in place a specific action plan for monitoring compliance with the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and its transcription in national legislation (law of 30 July 2018). It is important to mention that due to the nature of Sciensano's tasks, the type of data processed (health data) must, according to the regulation, be subject to specific security measures such as data protection impact assessments.

Scientific projects from Sciensano will often require an approval from an ethical committee. Within the study protocols that are assessed by ethical committees, information about data management is included. Sections regarding data management include information on the data flow and personal data protection measures, including data access procedure.

Besides approvals of an ethical committee, the health data processing by Sciensano is often also bound to a) positive deliberations by the Information Security Committee Chamber of Social Security and Health and b) authorizations by the Minister of Internal Affairs concerning the use of National Registry data. Within the authorization requests of these parties, Sciensano employees will have to justify their data needs and elaborate on security measures. Furthermore, the functions and services from Sciensano that will access data have to be specified.

These external ethical and data protections bodies will assess whether the project adheres to the principles of data protection like lawfulness, fairness and transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity and confidentiality. Often a Small Cell Risk Analysis (SCRA) is required by the Information Security Committee. Sciensano has set up a collaboration with different public health administrations that deal with personal health data in order to undertake SCRAs with closed bursaries and as such, standardize the methodologies of the SCRAs and reduced costs.

Healthdata.be specificity
At healthdata.be, the information collected may only be passed onto authorized researchers and supervisory doctors for the purpose of improving the quality and management of the health and healthcare sector. The data can therefore only be passed on for the purposes of scientific research, health monitoring and the promotion of knowledge. The data are never communicated to third parties for a purpose other than research or the protection of public health. For example, they are never communicated to a potential employer within the framework of a recruitment procedure, to a bank in order to obtain a loan or to an insurance company in order to obtain a policy. Moreover, the health data are systematically pseudonimized before they are made available to the researchers. In other words, only the data necessary for the research will be communicated and the information that makes it possible to formally identify the patients will always be coded.

In terms of personal data protection and confidentiality preservation, Sciensano is Building.

2.3. STAFF KNOWLEDGE AND SKILLS

"Staff skills and knowledge are central to human interoperability. For organisations that manage diverse data sets and want to improve their interoperability, it is essential that as they develop data management strategies and set strategic objectives, consider oversight and accountability issues, and consider data stewardship functions, they also consider what skills and knowledge staff will need to effectively perform their duties…" (1).

The Human Resources department of Sciensano and the different scientific departments are very aware of the need for training personnel and recruiting new people with the skills to improve data management and data analytics skills. Data managers and coordinators are scattered at Sciensano; the requirement of knowledge and skills to join up data is not reflected in the job descriptions and often remains project-based. Several initiatives to address these skills needs are under way such as seminars with external speakers, new training packages and the recruitment of people with IT infrastructure and data science skills. Additionally, there are also plans to train Information Security Coaches. There are also skills needs on knowledge for the implementation of the FAIR principles, meta-data management and federated analysis.

In terms of staff knowledge and skills, Sciensano is in between Emerging and Learning.

2.4. INTERNAL AND EXTERNAL COMMUNICATION

"Another tenet of human interoperability is ensuring that there is effective internal and external communication around key data management issues that affect users' ability to use and integrate data. Internally in an organisation, formal communications channels need to be established between individuals, teams, and departments when setting strategic objectives in data management strategies, leadership and management functions, oversight and accountability, and data stewardship functions. Externally, it is essential to establish user feedback loops that enable communication of data quality, gaps, usability, and other dimensions of human interoperability to be raised and resolved. Similarly, for non-technical audiences, it is important that organisations (…) can effectively communicate with policymaking audiences who are the ultimate users of their data…" (1).

Internally, more formal communication channels should be put in place to allow communication between individuals and departments, data focal points, DPOs and information security coordinator and healthdata.be.

Externally, Sciensano communicates with its stakeholders on data management issues via direct contact with data providers or contacts with the Information Security Committee via Single Points Of Contact (DPO's).

Sciensano has developed expertise in key health data management issues at EU level through the EU Health Information System Unit. Sciensano is recognized as a solid player in the European community. The Unit's activities has boosted Sciensano's presence at European level and contributed to Sciensano being a renowned institutions for its leadership in European projects, knowledge and expertise in the field of health information, vision for developments in health information and reuse of data, and as a trusted international collaborator with quality outputs.

In terms of internal and external communication, Sciensano is Learning.

2.5. ADAPTABILITY

"An organisation's ability to ensure that its internal data management processes are adaptable is an important dimension of human interoperability. As digital and data innovations flourish, organisations

and their staff need to ensure that their data management strategies keep apace of developments. Adaptability requires communication and empowerment of team members, as well as flexibility to change course at any point in order to meet evolving stakeholders' needs...A good example of the value that being adaptive can add relates to the need for individual and organisational awareness around the emergence of new data infrastructure components. For instance, organisations seeking to improve the interoperability, accessibility, and use of their data assets should be aware of trends pertaining to the accessibility and application of digital infrastructure, and be able to adapt to them or adapt them to their needs…" (1).

In order to address the data and information needs of policy makers, researchers and users, data management systems need to be standardised and harmonised. It requires allocated responsibilities, roles and specific departments in charge to be adaptable and to adjust to new information needs. Currently the data management processes at Sciensano are insufficiently streamlined and too complex because different departments have different data management systems in place. A very diverse data management landscape decreases adaptability.

Sciensano exists since 2018. During the COVID crisis, Sciensano has, for certain activities, shown great adaptability in terms of delivering information to tackle the COVID crisis by providing real time policy advice. Therefore, Sciensano is still Learning but has also shown levels of Building.

3. Data interoperability

3.1. DATA AND META DATA MODELLING CAPACITY

"Data and metadata modelling are the foundations of data interoperability. They are part of the design process in which the internal structure and interrelations across different data sets are defined, optimized, and described to capture all relevant business metrics and dimensions (6). The process of modelling data and metadata includes a number of steps, ranging from the discovery, analysis, and scoping of data requirements through to communicating them in the form of a model (5)…" (1).

Each department at Sciensano, most probably, consumes and generates data. These data are useful, necessary for the acquisition of knowledge, for decision making, for the development of public health plans... Data is at the heart of Sciensano's activities and scientific work. Sciensano's file server hosts more than half a petabyte of data.

Currently, data outside of healthdata.be is stored in the datacentres of Sciensano. The ICT department distinguishes storage for active, passive and archived data and provides the technical solutions for these 3 types of data. Metadata catalogs are important in the context of implementing a data strategy for Sciensano as it addresses the notion of sharing and embracing the FAIR principles: Findable, Accessible, Interoperable, Reusable. It is about the life cycle of data and its value-valuation: with a simple motto: "create once, use many times". There are however, existing beautiful examples of how meta data can be described.

What is the need to maintain a metadata catalogue? How it relates to the data strategy?
The process of moving from a closed approach in siloes to open, distributed and federated infrastructures aims to add value to data. We need to improve our working methods and create new approaches to data exploration and analysis. The data strategy provides an operational framework for data governance according to the concept of the three axes Technology, Organisation, Process (TOP): The technology axe, "T", relates to all the technologies aimed at facilitating the management and exploitation of data according to the rules of the business. The metadata catalogue is an essential element to guarantee the discovery and use of data. When "P" relates to the application of the rules aiming at enhancing the value of data (FAIRification of data) and "O" is about settling a data governance authoritative team responsible for the implementation of the data strategy. The FAIR guiding principles

facilitate data valuation by making data findable, accessible, interoperable and reusable. The application of these principles is the implementation of a FAIR metadata catalogue. This is to help "human" users and machines in the discovery, access, integration and analysis of data.

In terms of data and meta data modelling capacity, Sciensano *overall* is at the level of Emerging. Taking into account healthdata.be the maturity level is Building.

3.2. DATA ORGANISATION AND CLASSIFICATION CAPACITY

"Coordination within and across organisations, and whole national (statistical) systems or data ecosystems for that matter, to decide on common classifications and identifiers is key to enabling data interoperability. Standard vocabularies, classifications, and unique identifiers are part of the basic data infrastructure of a country or organisation. They help improve consistency and avoid ambiguity in the description of a data set, while enabling users to more easily locate and link together related data elements…. Standard classifications enable the integration of multiple data sets, as well as their consistent analysis and interpretation…" (1).

Except for healthdata.be where common classification systems are in place, there are different practices in terms of classification and identifiers across the different scientific directorates, services and units at Sciensano. In terms of data organisation and classification capacity Sciensano is Learning.

3.3. DATA ACCESS, OPENNESS AND SHARING

"…there are also technical considerations at the data interoperability layer. At the data interoperability layer, considerations around how to strategically plan for the publication of open data also touch upon how data sets are modelled and data is classified. Where possible, open-source software and data standards should be used for open data and metadata publication because these will allow data users to more easily integrate new data sets into their systems. There is a growing interest in tools and technologies that allow for the publication of data in such a way that machines can easily identify and integrate related information sources directly through the world wide web. This "linked data" approach facilitates interoperability and data integration by connecting both structured and unstructured data sets with special metadata elements designed to be referenced over the web…" (1). To build up the capabilities that are required to create a data sharing ecosystems, a data sharing framework needs to be put in place that involves multiple stakeholders comprising several building blocks: data sharing strategy, technological capabilities, regulatory and legal capacity and an approach to operationalise data sharing (2). This paper also distinguishes different sensitivity tiers (open access, controlled access-restricted and controlled access-closed) which could inform the optimal data sharing strategy.

Sciensano makes its microdata available for (statistical) research. Sciensano has set up microdata services for external users through healthdata.be. This service allows to link with the national register number[4]. Data from f.e. the Health Information survey and the Food Consumption Survey are also available via different services with the procedures described on the website pages. These data are particularly sensitive to privacy. For this reason, microdata research is subject to strict conditions and security requirements. External research can again be cross-thematic research, which requires the microdata of Sciensano to be combined and linked up with the researchers' own data. During the COVID crisis there was a data request group that was set up to deal with the different data requests from

---

[4] The key is the National Register Number– pseudonomised by an algorithm by a Trusted Third Party. Additionally, any Geographical Information System related data can be linked (e.g. via Statbel) to the National Register Number based microdata. If no National register Number is available in the microdata, information (e.g. poverty, air pollution, …) can be linked based on alternative information like a statistical sector, municipality etc… as long as that information is defined on the level of statistical sector, municipality or in another way can be extrapolated (statistical) to these levels.

universities. These take place every two weeks, are attended by the DPO and legal department. Under positive advice, a Data Transfer Agreement is made between Sciensano and the university.

Open science, including open data, is an important driver of change in science. A new way of conducting research increasingly demands the sharing and opening up of data and the proper handling of research data. According to the European Commission, open science should become the norm for publicly funded research[5].

The procedure to share data processed by Sciensano with third parties is described in a Standard Operating Procedure. Open data is available on the following link: https://www.healthybelgium.be/en/health-status/about-the-health-status-report/interactive-databases-sciensano. Sciensano makes its data available through platforms
- https://data.gov.be/en at https://epistat.wiv-isp.be/covid/ and
- https://www.healthinformationportal.eu/ and
- https://fair.healthdata.be/search/?theme=COVID-19

The Belgian Official Journal published an open access provision in Belgian legislation on 5 September 2018. This is article XI.196, paragraph 2/1 in the Code of Economic (Belgian) Law which reads as follows: "*The author of a scientific article which is the result of research financed at least halfway by public funds shall retain (...) the right to transfer his rights to a publisher of a journal or to grant an exclusive or ordinary licence, to make the manuscript available to the public free of charge in open access in a journal after a period of 12 months for humanities and social sciences and 6 months for other sciences after its first publication, provided the source of the first publication is mentioned*".

The content of this article is mandatory and applies regardless of the law chosen by the parties as soon as a connecting factor is located in Belgium. This right only concerns scientific journal articles, not books (or contributions to books).This right is retroactive and therefore also applies to older publications (before 2018). It is therefore perfectly possible to make earlier publications available in open access via this article. Sciensano adheres to a policy of open data and open science respecting the GDPR framework. For open science Sciensano published on its website its scientific papers.

Healthdata.be
The technical platform healthdata.be, founded in 2015, aims to provide policymakers with up-to-date, reliable and valid scientific reports in a fast and autonomous way so that they can develop and (re)direct their policies. In order to compile these reports, researchers need (technical & administrative) tools that they can use quickly, easily and autonomously to collect and process the desired data in a reliable and effective manner. The collection of these data may not generate any administrative burden or technical burden for the healthcare providers and their service providers involved, and must guarantee maximum privacy and data protection for the healthcare providers and patients.

In 2021, the SARS-COVID 2 cooperation between the healthdata.be platform and regional health authorities resulted in the development of a real-time, high-performance and scalable technical platform for data exchange that can also support future cooperation projects.

The healthdata.be platform has become a reference for national and international ambitions in a relatively short period of time. For example, in 2021 the platform was consulted to provide its services

---

[5] This "should" should be clearly balanced because health data on raw microdata level will never be "open" in the broadest interpretation of open data.

for initiatives such as the Belgian HDA, the European Health Data Space (EHDS), and the European Health Emergency Preparedness and Response Authority (EU HERA).

An important change concerns the governance of the healthdata.be platform: in addition to the already operational Steering Committee of the healthdata.be platform, a new Guidance Committee has been installed. This Guidance Committee, composed of representatives from the federal and federated health administrations, has been charged with the planning and prioritisation of healthdata.be as well as the budgeting of resources.

Open data and COVID

The COVID crisis has acutely exposed the importance of open and FAIR data for the visibility and public image of public health institutes. External researchers and data journalists have discovered Sciensano and developed a strong interest in valorising the COVID-19 data generated, collected, and managed by Sciensano. Although these requests have caused additional administrative burden, the reuse of data has led to a win-win for the institute – external researchers validated insights or developed new ones, data journalists helped spread scientific messages, and international organisations developed country comparisons. In other words, the reuse of data has led to a more widespread and more efficient dissemination of the outputs of Sciensano, and has allowed to generate insights and products for which Sciensano would not have the time or resources. Public health institutes with effective open data platforms have been praised, while others have been reprimanded.

Given the current attention given to open and FAIR data, Sciensano should pro-actively distribute relevant outputs as open and FAIR data. This will allow fostering a positive image of openness, transparency and willingness to collaborate. A future FAIR data platform therefore needs to be highly visible, e.g., via a subdomain.

The FAIR data portal will furthermore serve as an overview of data activities, thereby increasing visibility of the different Sciensano projects. Finally, the portal will also be an asset for project proposals, as Sciensano researchers will be able to document that their research outputs will be published in the data portal, which by the way is increasingly a requirement of European and Belgian funding bodies.

In addition to a FAIR data platform, Sciensano needs to provide user-friendly visualisation tools to explore the generated outputs in an interactive way. The COVID dashboard has shown the great interest and value of such investments, but other examples also exist:

- Belgian National Burden of Disease Study (BeBOD), non-fatal burden of cancer
- Belgian National Burden of Disease Study (BeBOD), mortality and years of life lost:
- Standardised Procedures for Mortality Analysis (SPMA)
- Health Expectancies
- Belgian Health Interview Survey Interactive Analysis (HISIA)
- Epidemiology of Infectious Diseases (Epistat)

In terms of data access, openness and sharing, Sciensano is Learning with elements of Building.

3.4. DATA ANALYTICS AND AUTOMATION

"As digital and data revolutions have exploded around the world, so too has the ability of a broader range of entities and individuals to collect, organize, structure, and analyse more diverse types of data. As more data, and types of data, are joined together, interoperability becomes more important to ensuring that they are modelled in ways that enable processing and analysis, including by automated means…" (1).

A first requirement for data analytics is having an overview of all the data that Sciensano (outside of healthdata.be) has in-house and their meta data in order to be able to identify the potential for data valorisation. Sciensano has been taken steps in data analytics by linking survey and administrative data

sources and undertaking research using environmental data and geo spatial information for mapping. However, the current infrastructure does not allow undertaking this research in an automated way and without copying and transporting the data between institutions.

In terms of data analytics and automation, Sciensano is Emerging.

3.5. DATA PROTECTION AND STORAGE

"At its most basic level, data protection is the regulation of how access and use of data stored on computers, digital devices, and paper records is controlled. Data protection can be viewed through three lenses. First, it is a series of legal obligations in jurisdictions where data protection laws have been passed that establish rules around how data should be protected. The European Union's General Data Protection Regulation (EU 2016) is a good example of this. Second, data protection can relate to the protection of individuals' personal data and is closely related to privacy and confidentiality preservation. Finally, it can relate to a mix of data-related and technical considerations. Technically, it is related to cybersecurity and incident response, while at the data layer it can relate to data security, …" (1). A white paper (2) distinguishes several ways of storing data (data warehouse, data vaults, data lakes and data fabric). It is observed that many stakeholders holding a large amount of data would have a data warehouse storing data safely and efficiently. However, when security and protection requirements get layered over time and organisations get reluctant to share data as they are considered sensitive, these data warehouses may turn into data vaults. The transition should therefore be from data warehouse to data lake or fabric which are more progressive data sharing infrastructures.

Data protection

In order to improve the security of its data, Sciensano is developing an information security management system (ISMS) according to the standard ISO 27001. This holistic approach allows for better security management by providing decision makers with data to document their decisions. Sciensano uses information security risk analyses to develop a plan to address the identified risks.

Data storage

All data is stored on the central storage space provided at Sciensano. Disk space is provided for both scientific and administrative data. This storage space is automatically made available to all Sciensano employees while using Sciensano PCs (or specific volumes for a large amount of data). Storing data on any external supports (ex: hard disks, USB key) is not permitted. Data stored locally on PCs can only be of a temporary nature, which is allowed in the following cases:

- Offline use of centralized data during surveys, research on missions, interventions and actions with partners without network connection
- Offline use during travels, relocations

The standard principle within Sciensano is that the organisation's data is not stored in an external cloud data location. All data is stored, preserved and managed internally at Sciensano. Data processes are described at the project request stage (pre-cycle PMO). This process includes the nature of the data to be produced / collected, the export and import of data and the final storage of the data or results.

In terms of data protection and storage, Sciensano is in between Learning and Building.

4. Technological interoperability

4.1. DATA INFRASTRUCTURE

"Although it might be self-evident, it is worth explicitly stating that all entities and organisations hoping to engage with the broader data ecosystem and working towards improving the accessibility and use of their data assets need to have basic digital infrastructure, such as laptops, database capacity (server or cloud), and Internet connectivity. However, the concept of what constitutes digital infrastructure is vastly

broader than what it was even just a few years ago. There are now numerous categories of infrastructure that all rely, and impact, upon interoperability. A few examples of interconnected infrastructural components could include the following:

- Data storage infrastructure (whether local servers or cloud-based).
- Data management infrastructure, such as bespoke information management systems.
- Data processing infrastructure (increasingly cloud-based for large data sets).
- Data dissemination infrastructure composed of data platforms, community hubs, etc." (1).

Both Sciensano and healthdata.be provide separately a virtual environment for analysis of their data in a secured environment with different workflows. Currently, the supply of microdata for health related purposes is limited to microdata that are located at healthdata.be. This workflow includes approvals from healthdata.be and permissions of the Information Safety Committee or Ethical committee. The analysis is done with a default set of software and policies offered by the IT infrastructure.

The data infrastructure for the data not managed by healthdata.be is localized in file server folders with controlled and restricted access via the identity management application Sciensorg. Delegated rights, tracing and automation are combined to enhance a secure solution for all data access. Anonymous access to Sciensano data is done via the public website or Sciensano project sites. All data hosted in Sciensano (non-healthdata.be) is only accessible with a personal Sciensano account and is limited to the roles the accountholder is member of (Sciensorg). The DPOs analyse the demand for data access and give an advice or approval.

In terms of data infrastructure Sciensano is Learning.

4.2. CYBER SECURITY AND INCIDENT RESPONSE

"As more key data management functions move online and remote work arrangements become the rule rather than the exception, cybersecurity and protocols for incident response become more important. Take the example of cloud computing. Due to its reliance on hardware independent virtualization technology, cloud computing enables organisations to quickly back up data, applications, and even operating systems to a remote data centre, and to deploy them to multiple users in many different locations. However, all this data transmission over the Internet exposes it to cybersecurity threats…" (1).

The ICT infrastructure of Sciensano is protected by several layers of security, with network classifications, double authentication, custom solutions, identity management, anti-virus, anti-malware, anti-spam and intelligent data monitoring. Patching and maintenance of an IT infrastructure is crucial as a first defence against cyber incidents. Therefore, daily monitoring and a maintenance plan are mandatory in Sciensano. Incident registration and workflows are hosted on a separate secure platform, and internal QSE ticketing allows to register infringements on systems. Incident response and communication will be enhanced and optimized to raise the awareness of employees, partners and managers. Security training, principles and best practices are implemented in all ICT services (development, support centre etc) and renewed or optimized when needed. ISO27001 requirements are being implemented.

In terms of cyber security and incident response Sciensano is Emerging.

FUTURE

—

1. Organisational interoperability

1.1. STRATEGIC OBJECTIVES

Sciensano will have a data strategy as part of its management plan by end 2022. Sciensano is a data provider of healthdata.be but also a customer with whom healthdata.be has service level agreements. The data strategy will state clearly determine the position and the structure of Healtdata.be inside of Sciensano for the data strategy. Depending on the gap analysis, the implementation of the data strategy is expected to take at least 5 years. The data strategy will enable interoperability with other data producers and users to create the added value for health policy and research of Sciensano's data and will be coordinated with the initiatives of the Health Data Authority.

In terms of strategic objectives Sciensano wants to reach the maturity level of Building and Consolidating within the next five years.

As part of the implementation plan Sciensano sets two actions:
1. The data strategy is part of the management plan as a fifth mission for Sciensano
2. Determine the position and the structure of Healtdata.be inside of Sciensano for the data strategy

1.2. LEADERSHIP AND MANAGEMENT

Sciensano will install a data governance committee with a mandate to lead on all four dimensions of interoperability (data, human, technological and strategic). The leadership will support new roles in data management. In the job descriptions of data analysists and scientists, interoperability is identified as a requirement. Data management requirements and rules could become part of employment regulations.

In terms of leadership and management Sciensano wants to reach the level of Consolidating in the next five years.

As part of the implementation plan Sciensano sets two actions:
1. A data governance committee is set up within Sciensano
2. The four dimensions of interoperability are recognised as a necessity by directors (allocation of roles, job descriptions, employment regulations)

1.3. OVERSIGHT AND ACCOUNTABILITY

Sciensano wants to ensure oversight and accountability by creating a new central role for a data governance officer, allocate resources to identify data stewards for different scientific directorates, data coordinators for different services and data contact points for different data sets. In terms of data access and collection, Sciensano recognizes the value of a separate unit for data entry, data validation and data quality assurance. This would also increase transparency, professionalize the quality checking and validation of data and the recognition of Sciensano as a trusted third party through the separation of 3 roles: data quality assurance/validation, data pseudonymisation and data analysis.

Data collection and management are costly activities. Therefore it is also important that sufficient resources are allocated to these activities in the applications for research funding.

Sciensano wants to reach the maturity level of Building in the next five years.

As part of the implementation plan Sciensano sets 2 actions:

1. Oversight structures are in place for Sciensano data that are not in healthdata.be with a clear strategy for personal and non-personal data. A separate unit/group is in charge of data entry, collection, quality assessment and validation allowing separation of the 3 roles: data entry and validation; data pseudonymisation; research.
2. Attention is paid in the applications for research funding that sufficient resources are allocated to the tasks of data collection and management

1.4. LEGAL COMPLIANCE

Sciensano wants to reinforce and offer its legal expertise with regards to data protection, the Data Governance Act and GDPR. Sciensano plays a key role in providing feedback on the reformulation and updating of the data laws.

In terms of legal compliance Sciensano wants to mature to the level of Consolidating within the next five years.

As part of the implementation plan Sciensano sets 2 actions:

1. In house legal knowledge and expertise on data access, processing and sharing is used for Sciensano to play an active role in the updating current data laws.
2. FAIR data principles are made explicit in the contracts and agreements with data providers and partners

1.5. DATA ETHICS

Sciensano has fragmented expertise in data ethics. In view of the importance of data ethics, one could argue for a role of data ethicus within Sciensano. Existing personnel is eligible for this role.
In terms of data ethics Sciensano wants to reach the maturity level of Building within the next five years.

As part of the implementation plan Sciensano sets 1 action:
Create a data ethics committee and a role for a data ethicus at Sciensano.

1.6. PROCUREMENT

Sciensano aims to coordinate ICT solutions in the interest of the entire organisation. In terms of procurement Sciensano wants to reach the maturity level of Building within the next five years.

As part of the implementation plan Sciensano sets 1 action:
Coordinate all procurement of ICT and technical solutions across the organisation.

1.7. LINKS TO BROADER DATA ECOSYSTEM

In terms of its interactions with a broader data ecosystem, Sciensano aims to establish formal networks with national universities via Memoranda of Understanding and strategic agreements, with other administrative bodies via the HDA and EU projects such as PHIRI, TEHDAS and the European Health Data Space.

In terms of the broader data ecosystem Sciensano wants to reach the maturity level between Building and Consolidating within the next five years.

As part of the implementation plan Sciensano sets 2 actions:

1. Formalise interactions within the data ecosystem via MOUs and strategic agreements with universities and federal/regional (knowledge) institutions on data exchanges and collaborations.
2. Sciensano is a key player in the creation of the HDA and the EHDS

2. Human interoperability

2.1. DATA STEWARDSHIP

Within Sciensano two roles with regards to data security have been formally defined: the data protection officer and the information security coordinator. Additionally, there are also plans to train Information Security Coaches. New roles will be allocated for data management. Data champions have been nominated who will fulfil a more informal role as reference points with expertise on data issues (data visibility, meta data, fair principles, legal issues, DMP online).

For each of the scientific directorates there will be clearly defined data roles for data stewards with data coordinators per service and data contact points for the different data bases being handled in that service or unit. The data coordinator and contact points have to make sure that the data is findable and accessible, that there is a logbook of the variables in the data sets, that meta data are available and that other departments can make use of the data. Clear rules will be set on data access, use and collaboration rules for scientific output.

For data stewardship Sciensano wants to reach the maturity level of Building and Consolidating within the next five years.

As part of the implementation plan Sciensano sets 1 action:
Define roles for data leadership functions to allow FAIR data management: appoint a chief data officer and appoint data stewards per scientific directorate, data coordinators per service and data contact points for different data sets.

2.2. PERSONAL DATA PROTECTION AND CONFIDENTIALITY PRESERVATION

Sciensano wants to reinforce its expertise and knowledge on privacy and confidentiality preservation in order to address issues such as data minimisation and proportionality (data controller should limit the collection of personal information to what is directly relevant and necessary to accomplish a specified purpose). In order to increase transparency and trustworthiness, it is important that data functions and roles (data entry and validation, quality assessment/Trusted Third Party and pseudonymisation of data/research and analysis) in an organisation are separated. Sciensano has some of these separate units in place (data protection office (DPO) and information security); this unit could be further reinforced with methodologists having expertise in statistical disclosure control, privacy preserving techniques. The separation of functions will also be further developed. A separate (more technical) unit where the data enters Sciensano with a validation and quality control function and staff with methodological expertise in data quality assessment and validation, allows professionalising data management at entry point.

Because of the two formal roles that exist with regards to privacy and confidentiality, Sciensano has a higher level of maturity for this indicator. Sciensano wants to reach the maturity level of Consolidating within the next five years.

As part of the implementation plan Sciensano sets 1 action:
Strengthen personal data protection and information security with methodological expertise on statistical disclosure control, privacy preserving techniques and risk assessment.

2.3. STAFF KNOWLEDGE AND SKILLS

Sciensano is planning several training initiatives to sensitize researchers for FAIR principles, to increase knowledge on data management and protection and to understand the implications of laws and regulations with regards to data for its researchers. Sciensano researchers will be trained to increase awareness and foresee data protection issues. For innovation and data science to be further developed, the recruitment of data scientists is seen as a priority.

For staff knowledge and skills Sciensano wants to reach the maturity level of Building within the next five years.

As part of the implementation plan Sciensano sets 1 action:
Set up Sciensano Academy to strengthen training on
- Data starters kit for researchers on data management and risk assessment of privacy
- Legal awareness of data issues
- Methodological training (Statistical Disclosure Control, Data quality and validation)
- Data science

2.4. INTERNAL AND EXTERNAL COMMUNICATION

For internal and external communication Sciensano wants to reach the maturity level between Building and Consolidating within the next five years.

Sciensano plans to increase internal communication on agreed data principles and issues and provide a chance to share best practices and examples of value generation via modern communication channels and a data science community. There are also the EPI Tuesdays and SEP seminars that increase knowledge and exchange of information with regards to FAIR principles and data maturity.

Via the EU projects such as PHIRI, TEHDAS and the EHDS Sciensano communicates and shares its experiences and examples with others in the national and international context through engagement with stakeholders. Through these EU projects, dedicated training activities are organized on e.g. FAIR principles, meta-data catalogues, data governance, GDPR compliance etc. These events are open to all Sciensano staff and are internally communicated. These are ideal settings to exchange good practice within Sciensano but also to attract expertise outside Sciensano through the EU HIS unit international network.

As part of the implementation plan Sciensano sets 1 action:
Develop an internal and external communication plan on the data strategy for Sciensano via seminars, a Data Science Community and national and EU projects.

2.5. ADAPTABILITY

Agile is an iterative and incremental approach to project management. It offers flexibility, transparency, quality, and continuous improvement. It increases focus on the specific needs and reduces waste through minimizing resources. The increased flexibility enables teams to easily adapt to change and it offers better control of projects. Sciensano considers adopting an agile approach via a multidisciplinary approach with a good overview in the central data collection team. A separate data entry/collection unit with professional and specialist expertise and methodologists to consult on generic methods also increases the adaptability of the organisation through harmonisation in case of a new crisis.

In view of the experiences during the COVID crisis, Sciensano wants to reach the maturity level between Building and Consolidating for adaptability within the next five years.

As part of the implementation plan Sciensano sets 1 action:
Build and further develop on evidence and results and innovate using an Agile approach.

3. Data interoperability

3.1. DATA AND META DATA MODELLING CAPACITY

For the data that are not in healthdata.be, there is a need for an exhaustive data inventory that includes the data that Sciensano (outside of healthdata.be) possesses and generates, whether it is personal data or not, where it is saved, if meta data are available and who is the contact person for the data set. For

data virtualization the ICT department is ISO270001 certified. Programming languages such as SAS, R and Python for data analysis and use of open data sources should be encouraged to further improve scientific output and services.

What is Data virtualisation?
Data virtualisation consists of providing a service interface that hides the technical details related to the data, such as its location in the database, its storage structure, the access APIs (end-points), the query language, etc. The virtualisation interface addresses the core needs of the users streamlining the "access to data" in one web based platform, making it efficient and fully operational. All independent services are integrated in one user experience platform.

A user experience platform (UXP) is an integrated set of technologies used to provide interaction between a user and a set of applications, processes, content, services or other users. A UXP has several components, including portals, content management, search, rich Internet application (RIA) tools, analytics, collaboration, social and mobile tools. It may be delivered as a suite of products or as a single product. The diagram presented from the Central Bureau of Statistics Data Strategy gives a good overview of the data virtualisation principle.
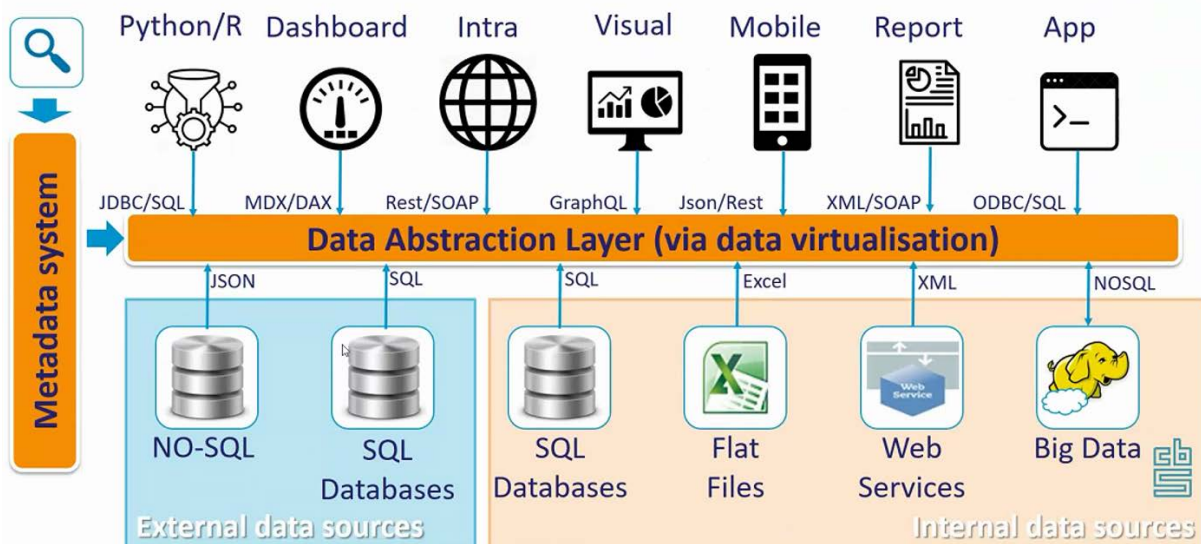
*On top are exposed all key end-services offered to the users.

How to conceive the data virtualisation layer, a "Sciensano data store"?
The Sciensano data store can be designed as a Digital Experience Platform (DXP) where data, information, code, data use and user support… are all organised, managed and optimised in order to generate a "successful and satisfactory" user experience. A data virtualisation service enables one integrated gateway, a single entry point, a "Sciensano data store": a central place where data and metadata can be found, an open science, open data and collaboration space. It is a one-stop-shop that compiles services in order to facilitate access to data through efficient Natural Language Processing and an augmented access to the data displaying some metrics e.g.: mean values, standard deviations, number of samples, ... but also preview options such as displaying the data for the certain regions and the possibility to download specific cohorts.

**Discoverability** (on the diagram: "metadata system"): a FAIR DATA POINT (FDP)
"FDP is a metadata repository that provides access to metadata in a manner that follows the FAIR (Findable, Accessible, Interoperable, and Reusable) Principles for data/metadata publishing. FDP is a

software that, from one side, allows digital objects owners/publishers to expose the metadata of their digital objects in a FAIR manner and, from the other side, allows digital objects consumers to discover information (metadata) about the offered digital objects. Commonly, the FAIR Data Point is used to expose metadata of datasets, but metadata of other types of digital objects can also be exposed such as ontologies, repositories, analysis algorithms, websites, etc."

Processing (on the diagram: "Python/R" / also "dashboard" or "report"): <u>JUPYTER HUB</u>
User centric approach (dxp): users want to use the tools they are most comfortable with.
"JupyterHub brings the power of notebooks to groups of users. It gives users access to computational environments and resources without burdening the users with installation and maintenance tasks. Users - including students, researchers and data scientists - can get their work done in their own workspaces on shared resources which can be managed efficiently by system administrators."
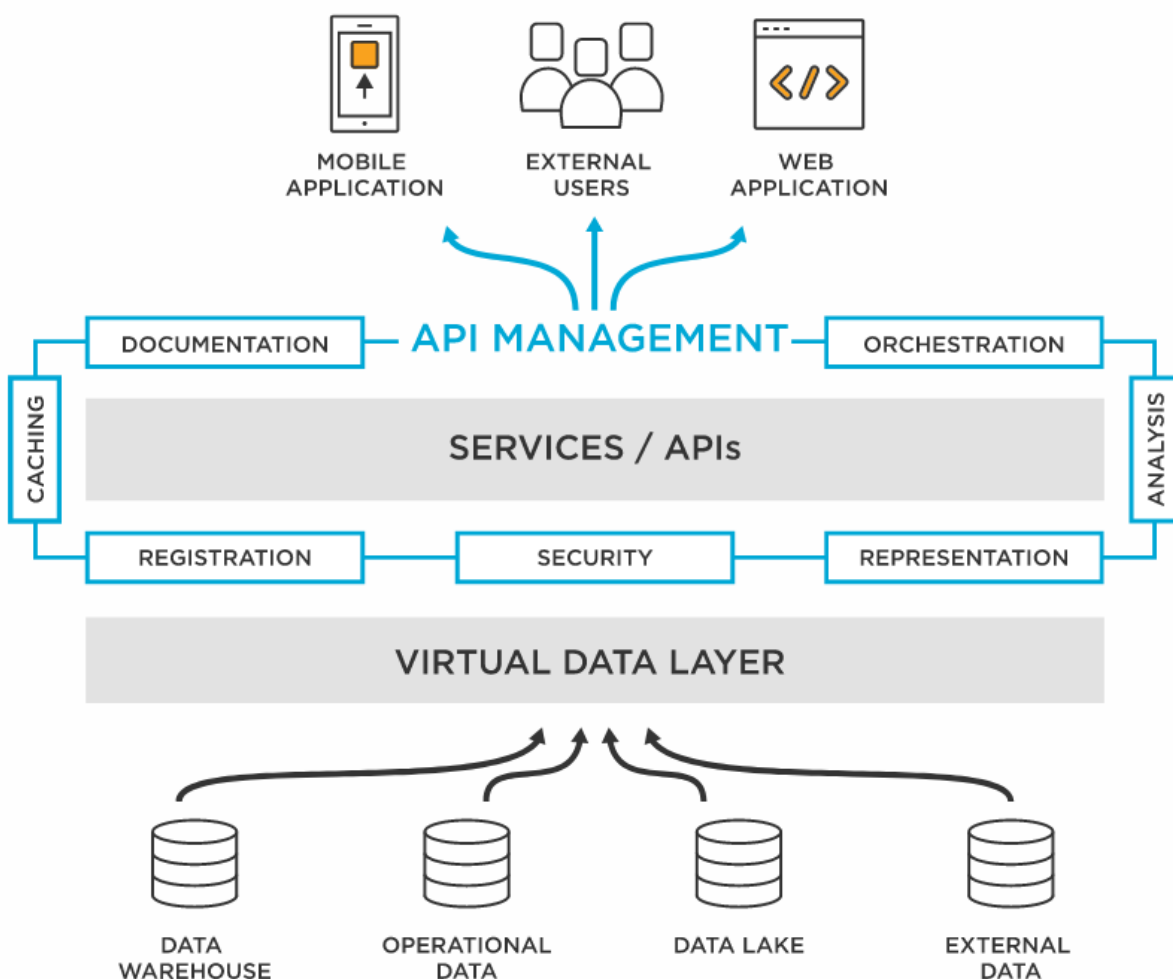

The following use cases 1 and 2 serve as examples providing information on the role of a data virtualisation team at Sciensano. The first use case comes from the aerobiology and mycology service. The diagram use case 1 reflects the data flow for a "pollen" autonomous recording station (longitude/latitude). It shows the bulk of technologies involved in the data lifecycle from the data curation to its processing and production of indicators. The data is made accessible through INSPIRE geospatial web services allowing the interoperability with other geospatial infrastructures and access to environmental data.

Use-case 1: "Pollen recording in-situ station"
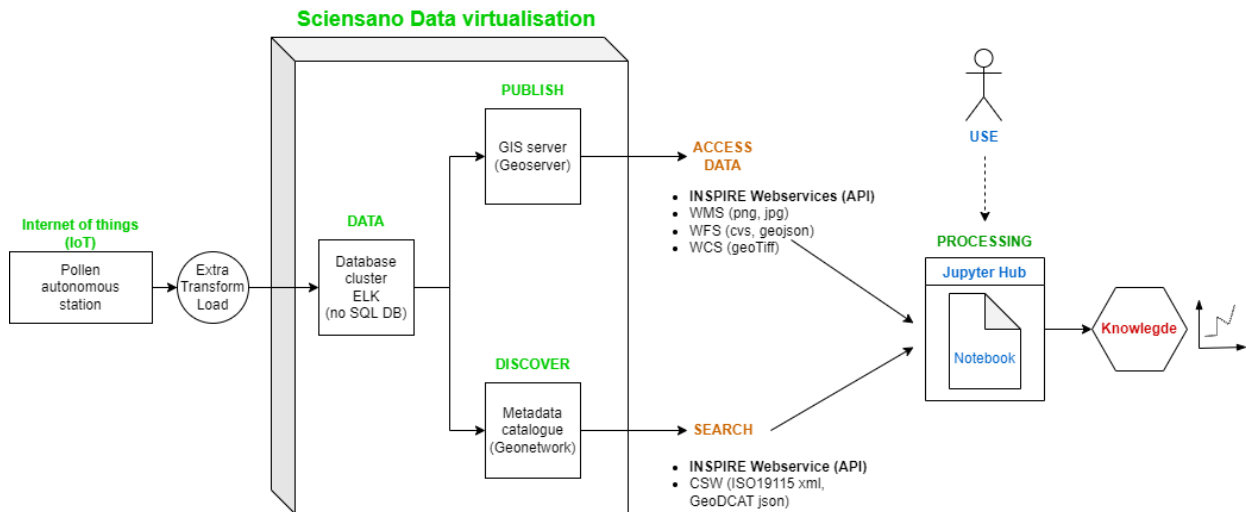
The second use case presents the principle of a "knowledge hub" supporting access to sensitive health data. It allows a user to search and find data through a linked-data metadata catalogue (i.e.: SPARQL query). It provides a computational environment (i.e.: Jupyter hub) and access to training samples (i.e.: open, anonymised or synthetic data) to program data analyses in the form of computational notebooks. A technical interoperability framework is set-up in order to allow a data analysis (i.e.: a notebook) to be computationally reproduced in an isolated and secured processing environment (i.e.: isolated Jupyter hub environment) by accessing sensitive data.

Use-case 2:
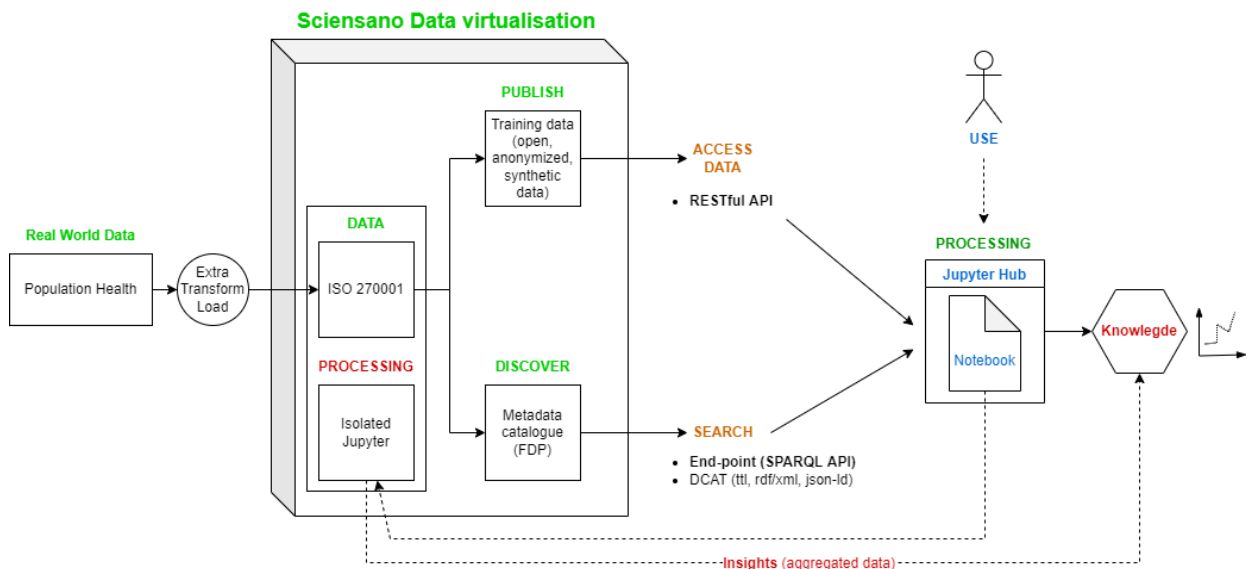
In terms of data and meta data modelling capacity Sciensano wants to reach the maturity level between Building and Consolidating within the next five years.

As part of the implementation plan Sciensano sets 5 actions for a Data store DXP:
1. Make a data inventory
2. Meta data are available for all Sciensano data outside of healthdata.be
3. Data virtualization as central infrastructure to allow data linkage with data at healthdata.be and external partners

4. Encourage programming languages such as SAS, R and Python
5. Allow the access to complex data sets

3.2. DATA ORGANISATION AND CLASSIFICATION CAPACITY

At the moment there are no common existing classification systems/rules/procedures. In terms of data organisation and classification capacity Sciensano wants to reach the maturity level between Building and Consolidating within the next five years.

As part of the implementation plan Sciensano sets 1 action:
Ensure common classification systems of all Sciensano data.

3.3. DATA ACCESS, OPENNESS AND SHARING

During the COVID crisis Sciensano played an important role in making certain data available real time as well as allowing linking data of different institutions. It also appeared that Sciensano was very dependent on other organisations for data access. In the context of the HDA Sciensano wants to continue playing an important role in making real time data available as well playing a catalysing role in the linkage of different dataset between different institutions. In view of the increase of available open data, one could consider the appointment of a data scout who pro-actively lobbies for new datasets that are free or affordable. At the same time legal actions should be taken to change the law to facilitate access to data for research and output of Sciensano, as they exist for some other health institutions.

In terms of data access, openness and sharing Sciensano wants to reach the maturity level between Building and Consolidating within the next five years.

As part of the implementation plan Sciensano sets 5 actions:
1. Open data via 1 fair.data platform for Sciensano
2. Standard process for Data Transfer Agreements
3. Standard Activity Based Costing for data requests, storage, archiving, back-up for private and public
4. Appoint a data scout
5. A procedure is started to change the law to facilitate access to data for research and output of Sciensano

3.4. DATA ANALYTICS AND AUTOMATION

It is certainly Sciensano's ambition to increase the use of data science tools and methods (webscraping, text mining, data mining, Machine Learning and Deep Learning, Artificial Intelligence, privacy preserving data science techniques) and exploit existing relevant (open) data to fulfil Sciensano's mission even better by making policy information and research more timely, more detailed in order to address complex policy questions.

Sciensano has access to large registers, surveys and other big (visual, sample, genome, chemical) data which could be used to develop models and calculate propensities (to develop a disease for example), to develop early warning models or to link large data sets via network analysis and platforms for the early detection of health threats.
New data (sources) are generated every day and are often open (for example environmental data). Hospitals are getting better at registering their data (with ambitions to develop syndromic surveillance) due to the requirements for electronic patient records and new systems have been set up during the COVID crisis that collect data from pharmacies and GPs. These systems will allow real time data to become available which offers new potential for data linkage and more in depth and complex public health studies.

In order to develop innovation, Sciensano will set up the Data Science Community in 2022. Through this community, Sciensano will bring together epidemiological and (bio)statistical experts with data science expertise in order to develop and promote a hybrid methodological approach through use cases; the data science community will also facilitate capacity building for the research community within Sciensano. Additionally, the data science community could help to identify (open) data sources (data

scouting) and skills among our regional (Flemish AI community) and national partners to achieve a holistic One Health approach in research. Recruitment of people with the right skills and current innovation research funding at EU and national level should make this innovation possible.

Towards a Health population digital twin

A digital twin is, in essence, a digital infrastructure that uses real world data to create simulations that can predict how a product, a process or a system will perform. This infrastructure can integrate the internet of things (semantic Web), artificial intelligence and software analytics to enhance the output.

With the advancement of machine learning and factors such as big data technologies, these virtual models have become a staple in modern analytics to drive innovation and improve knowledge. Creating a health digital twin can allow the enhancement of research by using advanced analytical, monitoring and predictive capabilities, test processes and services.

In terms of data analytics and automation, Sciensano wants to reach the maturity level of Building within the next five years.

As part of the implementation plan Sciensano sets 2 actions:
1. Set up the Data Science Community
2. Include data science and data engineering skills in vacancy texts

3.5. DATA PROTECTION AND STORAGE

The ICT department supports various technologies for storage and archiving. Active data is hosted on the most performant disk technology (flash memory principle), and is being used for the most performant activities. Passive data is hosted on slower economical storage. Archiving, which equals offline data, is placed on tapes or on disks with low power usage. Data protection is ensured via identity management processes, however the data itself can be protected with other security measures such as encryption, password protected files, integrated application roles and quarantined computers or virtual PC's. The next level of data protection that needs to be achieved is separately classified networks where data cannot be transferred.

Sciensano could foresee a (cloud) service provider certified ISO270001 for securing and managing personal health data that address researchers needs. Sciensano could also foresee a geospatial server for storage and managing all GIS (non personal) health data. In terms of data protection and storage, Sciensano wants to reach the maturity level of Consolidating within the next five years.

As part of the implementation plan Sciensano sets 2 actions:
1. Increase level of data protection with separate classified networks.
2. Consider the possibility of a cloud server provider for securing personal data.

4. Technological interoperability

4.1. DATA INFRASTRUCTURE

The ICT department will evolve to the next level of Infrastructure as a Service, Software as a Service and Security as a Service for the on-premise environment. The ICT department will support the missions and visions of Sciensano with its legal, accepted and approved data provider position.

In terms of data infrastructure, Sciensano wants to reach the maturity level of Building within the next five years.

As part of the implementation plan Sciensano sets 1 action:

Develop ICT to the next level of Infrastructure as a service, Software as a service and Security as a service for the on-premise environment.

4.2. CYBER SECURITY AND INCIDENT RESPONSE

In terms of cyber security and incident response, Sciensano wants to reach the maturity level of Building within the next five years.

As part of the implementation plan Sciensano sets 1 action:
Develop Sciensano to the next level of cybersecurity and incident response management.

5. Data driven public health institutes

5.1. THE ROLE OF PUBLIC HEALTH INSTITUTES IN SOCIETY

From the International Association of National Public Health Institutes (IANPHI) website, we find the following statement around the role of PHIs and the benefit of consolidating these functions in a National PHI (NPHI). "To do the job of public health – detect, measure, and tackle health challenges through population-based interventions – every country must carry out a set of functions that are the cornerstone of strong public health systems. Key among these are:

- Population health assessment (assessing the health status of the population)
- Health protection (surveillance and response)
- Research (evidence to inform policies and programs)"

Many countries consolidate these functions organisationally in a "national public health institute" – a science-based organisation (or network of organisations) that provides leadership and coordination for public health at the national level. In most cases, NPHIs are part of the government (usually under the Ministry of Health) or closely attached to it.

The world's NPHIs vary greatly. Yet, despite their differences in history, scope, and resources, NPHIs all provide core public health functions that improve their countries' efforts to address public health challenges both within and beyond their borders. Consolidating these functions — and the associated skills, disciplines, experience, and expertise — in an NPHI provides many benefits:

- Improved delivery of public health **services**
- More **efficient** use of funds
- Ability to generate and **share knowledge**, data, and evidence to inform public health decisions and policies
- Increased capacity to mount a **quick, decisive, and coordinated response** during a public health emergency
- Visible **national leadership for public health** issues
- Ability to develop **public health policy agendas** and resource allocation in line with the country's own priorities
- Consistent **policies** and **harmonized procedures**
- **Linkages** among all those working to improve public health in the country

NPHIs also create a central focus for human resources in public health and provide a career path and nucleus of public health professionals to carry out core public health functions."

Data have become the biggest asset for many organisations. As most institutions possess or process data, and PHIs have access to the most sensitive data available (as they concern one's health), their expertise in terms of data protection, data security, data processing and analysis, turning data into (public health) information, communicating to policy makers and the public, ICT and data analytics, rapid intervention and (data) monitoring systems, is unique. Their roles could be fulfilled even better by becoming data driven research institutes, introducing modern tools to increase the data quality control,

taking the role as a Trusted Third Party, using new data sources such as open (big) data, stimulating the use of methods and techniques in data science and analytics (machine learning, artificial intelligence) and undertaking research addressing complex health research question that could not be addressed previously.

The functions identified by IANPHI are still relevant but there is scope to reinforce them over the next five years. In view of the current developments in the EU data landscape these could be: pre-empting and identifying health policy questions through a Health Research System, undertaking research proactively and stimulating data driven research and innovation. NPHIs could further benefit from aligning experiences, attracting new skills and expertise and adopting data strategies in order to create a larger impact in society.

Also at the national level, PHIs could play a role in national health data management by having a data strategy and setting an example for other (health) organisations. The Belgian government has started an initiative to set up the Health Data Authority where all health organisations commit to new regulations and procedures around linking and sharing health data. NPHI should play a proactive role in supporting these kind of initiatives.

5.2. EU AND BEAUTIFUL EXAMPLES

There are some beautiful examples in other EU countries of HDAs; hereby a couple of inspiring examples.

Findata is the Health and Social Data Permit Authority in Finland and is the main authority permitting the secondary use of health and social data. Findata provides centralized management for access processes and discoverability of data collections and datasets from several different data sources through a publicly available metadata catalogue.

The Danish Health Data Authority is part of the Danish Ministry of Health. It is responsible for the maintenance of the national health registries, of which there are 40. They contain data related to the health of the Danish population and healthcare services. The Danish Health Data Authority has set up libraries describing the data collections that they manage.

The French Health Data Hub is a public structure which aims to manage access to non-nominative data that is hosted on a secure platform. The platform includes copies of already existing databases and are pseudonymised or anonymised. The data will be accessible to project coordinators contributing to the public interest conditioned to an approval process.

CONCLUSION

—

Sciensano is very conscious of the value of public health data and the protection thereof. Sciensano has managed to publish data on the COVID epidemic and was, still is, the point of reference and information for policy makers and the public during the entire crisis. It is in the aftermath of this crisis and the start of the HDA initiative of the health minister Frank Vandenbroucke that Sciensano has developed its data strategy by undertaking a data maturity assessment of the current situation, by formulating its future data maturity ambition and by developing a corresponding implementation plan. Sciensano wants to set an example for other health organisations in Belgium by having a professional data management system and strategy where all components of data, human, technological and organisational interoperability are dealt with.

The data maturity assessment clearly showed that Sciensano is still at Emerging, Learning levels of data maturity and interoperability except for certain indicators related to data protection and legal aspects where maturity is higher. It is important to make a distinction between Sciensano in general on the one hand and Healthdata.be on the other hand. Due to its role, data maturity of healthdata.be is generally higher. In the next five years, Sciensano wants to mature to levels of Building and Consolidating which means that for all indicators of interoperability the entire organisation will be organized in such a way that Sciensano can adhere to FAIR data principles and that Sciensano can reuse, valorize existing and new data sources in order to provide health information and fulfil its mission by reflecting trustworthiness, transparency, quality and security in dealing with (personal) health data.

The implementation plan sketches the first steps to achieve higher data maturity levels. Important implementation actions are the allocation of data management roles in the organisation (data governance officer and data stewards), to make an exhaustive data inventory, to ensure the description of meta data of all Sciensano data, to increase ICT capacity for data virtualization and to make the data strategy part of the management plan as a fifth strategic mission.

Since its creation in 2018 and its (pro)active role during the Covid crisis, Sciensano has received a lot of attention and questions on its role in society. Sciensano is a public health institute that not only provides advice and services for the government and the public but also undertakes health research, forming the base of its other activities. Through these activities, Sciensano has an extensive network of collaborations with universities, knowledge institutes and other health institutions through which data is exchanged and shared. It is in this role that Sciensano wants to take a more proactive innovative approach to become a data driven organisation that, through its expertise of data storage, handling, processing, protection and analytics to monitor health and provide health information, wants to set an example in the Belgian data landscape and increase its impact in society.

REFERENCES

—

(1) https://www.data4sdgs.org/sites/default/files/file_uploads/Joined_Up_Data_Maturity_Assessment_draft5.pdf

(2) Saw Swee Hock School of Public Health (2022). White Paper Responsible Data Sharing in Health and Healthcare. Asian Society.

(3) European Commission (2022). Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/strategy-data

(4) Sciensano (2022). https://covid-19.sciensano.be/fr/missions-de-sciensano

(5) DAMA International. (2017). Data Management Body of Knowledge (2nd Edition). New Jersey: Technics Publications

(6) Chang, R. 2018. A Beginner's Guide to Data Engineering — Part II. medium.com/@rchang/a-beginners-guideto-data-engineering-part-ii-47c4e7cbda71

ANNEXES

—

Annex 1 • Data maturity assessment of Sciensano 2022

Operability criteria	Data maturity Sciensano	Explanation
ORGANISATIONAL INTEROPERABILITY		
Strategic objectives	Emerging/Learning	The ability to join up data is recognized but it is not explicitly identified as a strategic objective. Interoperability is identified as a strategic objective in an organisation's technical units, but not outside of them.
Leadership and management	Emerging/Learning	Ad hoc leadership on interoperability issues emerges organically but is not coordinated. Leadership around interoperability emerges across various technical units but remains fragmented Silos persist.
Oversight and accountability	Emerging/Learning	An organisation is aware of the need to create accountability chains to oversee how data is being joined up, but is not yet taking steps to create them. Oversight structures emerge across different technical units but are not coordinated or aligned. Accountability over how interoperability efforts are undertaken is fragmented and unclear.
Legal compliance	Learning/Building	Active steps are taken to better understand legal compliance requirements around data retention, transmission, sharing, and use, and make them available to data users. Compliance with applicable laws on data sharing, transmission, and use is embedded in oversight and accountability functions, and is reflected in an organisation's data strategy, which is published online.
Data ethics	Learning	Active steps are taken to better understand the ethical impacts that joining up data might have and to understand how they might unintentionally cause harm. Rudimentary ethical impact assessments are undertaken on an ad hoc basis.
Procurement	Undefined/emerging	An organisation is not aware of the impact of the procurement of technical and data solutions on interoperability. There is an emerging understanding of the need to join up data across procurement activities, but there is no coherent approach. There is a heavy reliance on outside contractors to fill gaps but no coordination between units on how this is done, often resulting in incompatible data solutions being procured.
Links to broader data ecosystem	Learning	An organisation starts to attach pro forma licensing terms with provisions on data integration to data that

		it releases, transmits, or shares but does not monitor or engage with data users. An organisation starts to document the data that it receives from other organisations. An organisation starts to document and coordinate its engagements with other parts of the data ecosystem.
HUMAN INTEROPERABILITY		
Data Stewardship	Emerging/Learning	Joining up data forms part of certain staff members' function but it is not reflected in their job descriptions and is ad hoc. Staff in different units have recognized functions relating to data interoperability, but there is little or no central coordination.
Personal data protection and confidentiality preservation	Building	There is routine consideration of applicable law, principles, best practices, and guidance. An organisation undertakes privacy impact assessments before and during data-related projects and those assessments include considerations pertaining to the risks associated with data.
Staff knowledge and skills	Emerging/Learning	Some staff have the knowledge and skills to join up data, but this is not reflected in their job descriptions and is tangential to their main functions. Knowledge and skills relevant to interoperability start to be recognized as part of job descriptions in some units, but the approach is fragmented.
Internal and external communication	Learning	Staff and units start to share examples of good practice with each other, but this is not coordinated. The value of interoperability starts to be understood by non-technical staff but is not yet communicated externally.
Adaptability	Learning/Building	Disparate units across an organisation start to formally recognize the need to ensure that staff's functions and oversight of data systems are adaptable so as to ensure that value continues to be generated from their data assets. The value of empowering staff to be adaptable in how they use data, including in how they join it up with other data, is recognized by an organisation and is reflected in its data strategy. Staff have the authority to adapt their working processes and oversight of organisational data assets in ways that enhance its value, including by joining them up.
DATA INTEROPERABILITY		
Data and metadata modelling capacity	Emerging/Building	There is an emerging understanding of the value that data and metadata modelling can confer to data assets, but data modelling is not a priority for technical units. Technical units coordinate their approach to both data and metadata modelling and align efforts to consistently model data based on their organisational needs. Internal needs are still prioritized over external groups, but data is modelled consistently.

Data organisation and classification capacity	Learning	There are informal attempts between units to use common classifications, but these are not formalized or coordinated across all relevant units. There is some, but not consistent, use of common classifications across the organisation.
Data access, openness, and sharing	Learning/Building	There are coordinated efforts in technical units to ensure that data is accessible and shared responsibly, including relevant licensing permissions or limitations for future data integration and use. Some data is made open on an organisational platform, but data sets are incomplete, not timely, or have not been quality assured. Data is shared responsibly in ways that protect any rights that third parties may have over it. Data that is published openly is done so in machine readable formats under a clear open data license with terms of use, and has been stripped of attributes that may result in the re-identification of individuals or vulnerable groups. Open data portals are accompanied by relevant contextual information and are visualized in ways that promote use by numerous audiences.
Data analytics and automation	Emerging	There is disparate understanding of the role of interoperability in undertaking automated data analytics across organisational units. There is limited understanding of how interoperable data should be used to train algorithm.
Data protection and storage	Learning/Building	Disparate units routinely apply appropriate data protection techniques to their data sets before data integration, but there is little to no consistency in how those techniques are applied. There is some, but limited, understanding of the risks of re-identification inherent to interoperable data. Personal, sensitive, and sensitive group data is subject to appropriate protections before being integrated, shared, or processed through automated analytics tools. Risks of re-identification inherent to interoperable data are understood and are applied, but not routinely.
TECHNOLOGICAL INTEROPERABILITY		
Data infrastructure	Learning	All appropriate staff members have access to adequate hardware and software tools, as well as network connectivity. There are secure servers and data repositories, but they are used inconsistently by staff and organisational units; there is little oversight of digital infrastructure.
Cybersecurity and incident response	Emerging	Disparate staff and units across an organisation have awareness or show concern about the risks posed to their reusable data by a cyberattack or other data breach. Champions emerge who push for a data breach protocol or policy.

Annex 2 • Data maturity of Sciensano: maturity levels for the future

Operability criteria	Data maturity Sciensano	Explanation
ORGANISATIONAL INTEROPERABILITY		
Strategic objectives	Building / Consolidating	The need to join up data across systems is recognized as a strategic objective in an organisational data strategy. The value of standards and robust data governance is recognized. The strategic value that joined-up data can bring to decision-making is recognized in organisational strategies. Interoperability forms part of an organisation's external engagement strategy with other data producers and users. *Implementation actions*: 1. A data strategy is part of the management plan as a fifth mission for Sciensano 2. Determine the position and the structure of Healtdata.be inside of Sciensano for the data strategy
Leadership and management	Consolidating	There is a data governance committee or council and it has an explicit mandate to lead on interoperability issues. The value of joined-up data is understood by organisational leaders and managers, and is clearly identified as a function in relevant job descriptions. *Implementation actions:* 1. A data governance committee is set up within Sciensano 2. The four dimensions of interoperability are recognised as a necessity by directors (allocation of roles, job descriptions, employment regulations)
Oversight and accountability	Building	Oversight and accountability functions are embedded in an organisation's strategy and reflected in the leadership structure. Fragmented chains start to join up and common standard operating procedures emerge. *Implementation actions:* 1. Oversight structures are in place for Sciensano data (that are not in healthdata.be) with a clear strategy for personal and non-personal data: a separate unit/group is in charge of data collection, quality assessment and validation allowing separation of 3 roles with regards to data collection and validation; data pseudonymisation; research. 2. Attention is paid in the applications for research funding that sufficient resources are allocated to the tasks of data collection and management
Legal compliance	Consolidating	An organisation's data transmission, sharing, and use activities fully comply with applicable laws and sometimes exceed legal standards. *Implementation actions*: 1. In house legal knowledge and expertise on data access, processing and sharing is used for Sciensano to play an active role in the updating current data laws. 2. FAIR data principles are made explicit in the contracts and agreements with data providers and partners
Data ethics	Building	The types of ethical questions that joined-up data might give rise to are understood and appropriately categorized. Appropriate steps are taken to minimize harm caused by a breach of ethical standards. Ethical impact assessments are routinely undertaken.

		Implementation action: Set up a data ethics committee and a role of data ethicus for Sciensano
Procurement	Building	The procurement of compatible and interoperable data systems across an organisation is formalized and coordinated. There is a common procurement policy across the organisation that requires staff to consider interoperability issues when procuring new systems. Reliance on external contractors is strategic and coordinated. *Implementation action:* Coordinate all procurement of ICT and technical solutions across the organisation
Links to broader data ecosystem	Building / Consolidating	An organisation effectively categorizes its data and licenses it for use appropriately. An organisation documents all data that is shared with it and has a general understanding of what it can and cannot do with it. An organisation engages with other parts of the data ecosystem in a coordinated way, pursuant to its data strategy. There is a well-established and bespoke set of licenses that set out clear parameters for use, including integration depending on the category of data involved. An organisation documents all data that is shared with it and has clear guidance and procedures in place that govern whether and how that data can be joined up with other data sets in its control. An organisation makes engagement with other parts of the data ecosystem a strategic priority and has a well-coordinated approach with clear processes for joining up its data with external data. *Implementation actions:* 1. Formalise interactions within the data ecosystem via MOUs and strategic agreements with universities and federal/regional (knowledge) institutions on data exchanges and collaborations. 2. Sciensano is a key player in the creation of the HDA and the EHDS
HUMAN INTEROPERABILITY		
Data Stewardship	Building / Consolidating	There is a coordinated staffing plan that reflects the various dimensions and roles relating to data stewardship, including interoperability, across the organisation. Units are coordinated and communicate with each other about what data they are joining up, how, and why. A strategically thought-through plan for data management is overseen by an organisation's data governance council or committee and includes a clear plan for stewardship of data, including data interoperability functions. Staff across the organisation are aware of how data is used, joined up, and shared with other entities *Implementation action:* Define roles for data leadership functions to allow FAIR data management: appoint a central data governance officer and appoint data stewards per scientific directorate, data coordinators per service and data contact points of different data sets
Personal data protection and confidentiality preservation	Consolidating	The preservation of individual privacy and data confidentiality form part of an organisation's legal and ethical review and are integrated across the data life cycle. An organisation adheres to the highest applicable standards of privacy and confidentiality preservation. An organisation integrates privacy and confidentiality preservation as part of its data strategy and explicitly provides guidance surrounding the risks of interoperable data, such as the mosaic effect. An organisation is forward-looking and cognizant of the potential privacy risks inherent to interoperable data posed by emerging technologies such as the Internet of Things, biometric ID verification, or general automated processes. *Implementation action:*

		Strengthen personal data protection and information security with methodological expertise on statistical disclosure control, privacy preserving techniques, SCRA and risk assessment.
Staff knowledge and skills	Building	There is a coordinated approach to knowledge and skill strengthening across an organisation that explicitly recognizes and addresses interoperability needs. *Implementation actions:* Set up Sciensano Academy to offer training on: 1. Data starters kit for researchers on data management and risk assessment privacy 2. Legal awareness of data issues 3. Methodological training (Statistical Disclosure Control, Data quality and validation) 4. Data science
Internal and external communication	Building / Consolidating	Mechanisms to facilitate internal communication and sharing of best practices around interoperability form part of an organisation's data strategy. Cross-unit communication helps to translate best practices and examples of value generation for external audiences. An organisation has a variety of coordinated internal communication channels open between units and staff, enabling the sharing of best practices and examples of value generation. An organisation is a champion of the value of joined-up data to data ecosystems and actively communicates its experiences and examples with others in compelling and effective ways, including through engagement with data journalists and storytellers *Implementation action:* Develop an internal and external communication plan on the data strategy for Sciensano via a Data Science Community and national and EU projects.
Adaptability	Building / Consolidating	The value of empowering staff to be adaptable in how they use data, including in how they join it up with other data, is recognized by an organisation and is reflected in its data strategy. Staff have the authority to adapt their working processes and oversight of organisational data assets in ways that enhance its value, including by joining them up. An organisation becomes a leader in adaptive management, and staff feel empowered and are confident in their ability to adapt their oversight of data systems as needed, including how they join up data, to maximize value. *Implementation action* Build and further develop on evidence and results and innovate using an Agile approach.
DATA INTEROPERABILITY		
Data and metadata modelling capacity	Building / Consolidating	Technical units coordinate their approach to both data and metadata modelling and align efforts to consistently model data based on their organisational needs. Internal needs are still prioritized over external groups, but data is modelled consistently. An organisation routinely utilizes canonical data and metadata models that follow standardized patterns, making them reusable and conducive to data sharing. The selection and application of canonical models is done

		through careful planning, including through engagement with data users and other entities in the data ecosystem. *Implementation actions for a Data store DXP* 1. Make a data inventory 2. Meta data are available for all Sciensano data outside of healthdata.be 3. Data virtualization as central infrastructure to allow data linkage with data at healthdata.be and external partners 4. Encourage programming languages such as R and Python 5. Allow the access to complex data sets
Data organisation and classification capacity	Building / Consolidating	There is a coordinated approach to the use of data classifications across the organisation. Units work together to identify the most appropriate classifications for their data and ensure that the data under their control is appropriately classified. The organisation not only routinely and appropriately uses data classifications but also produces its own classifications to fill gaps and ensure consistency. The organisation engages actively with other entities in the data ecosystem to improve commonly used classification systems and establish new ones as needed. The organisation effectively communicates the value of consistent data classification for interoperability. *Implementation action:* Ensure common classification systems of all Sciensano data
Data access, openness, and sharing	Building / Consolidating	Data is shared responsibly in ways that protect any rights that third parties may have over it. Data that is published openly is done so in machine readable formats under a clear open data license with terms of use, and has been stripped of attributes that may result in the re-identification of individuals or vulnerable groups. Open data portals are accompanied by relevant contextual information and are visualized in ways that promote use by numerous audiences. An organisation operates an effective data sharing policy that provides guidance on the various ways in which data sharing should take place, from publication under an open license, through to the use of data sharing or processing agreements. Legal advice is available to staff wanting to share data that will be integrated with other data sets by third parties. Open data is not just published in machine and human readable formats but is also made available as linked data through the semantic web. There are feedback loops with key audience groups and the organisation is responsive to user needs. *Implementation actions:* 1. Open data via portals and platforms via 1 fair.data platform for Sciensano 2. Standard processes for Data Transfer Agreements 3. Standard Activity Based Costing for data requests, storage, archiving, back-up for private and public 4. Appoint a data scout 5. A procedure is started to change the law to facilitate access to data for research and output of Sciensano
Data analytics and automation	Building	Data analytics and machine learning functions are reflected in an organisation's data strategy. The relative benefits and risks of running automated analytics over interoperable data, or using it to train algorithms, are generally understood but there is not yet a consistent approach across an organisation.

		Implementation actions: 1. Set up the Data Science Community 2. Include data science and data engineering skills in vacancy texts
Data protection and storage	Consolidating	All data is protected using the appropriate techniques and either responsibly archived or permanently deleted at the end of its intended life cycle. Access to sensitive data sets is monitored and documented to ensure accountability over data protection. Prior to integration, sharing, or processing through automated analytics, all data is assessed for risks of re-identification or other harms and is only used when there is a high degree of certainty that the data will remain safe following reuse. An organisation helps to set standards for data protection within the broader data ecosystem and champions responsible data use. *Implementation actions:* 1. Increase level of data protection with separate classified networks where data cannot be transferred. 2. Address the possibility to have a cloud server provider for securing personal data.

TECHNOLOGICAL INTEROPERABILITY

Data infrastructure	Building	All appropriate staff members are aware of, and trained in, how to use an organisation's data management and processing systems. Secure servers and data repositories are routinely used by staff members and oversight of digital infrastructure is part of an organisation's data strategy. *Implementation action:* Develop ICT to the next level of Infrastructure as A service, Software as a service and Security as a service for the on-premise environment.
Cybersecurity and incident response	Building	A clear data breach policy setting out sequential steps and responsibilities is established. Staff receive training on what they should do in the event of a data breach and are taught about the risks associated with the reuse of interoperable stolen data. *Implementation action*: Develop Sciensano to the next level of cybersecurity and incident response management

CONTACT
Sofie De Broe • T+32 2 478 843194 • Sofie.debroe@sciensano.be

MORE INFORMATION

–

Visit our website
www.sciensano.be or contact
us at info@sciensano.be