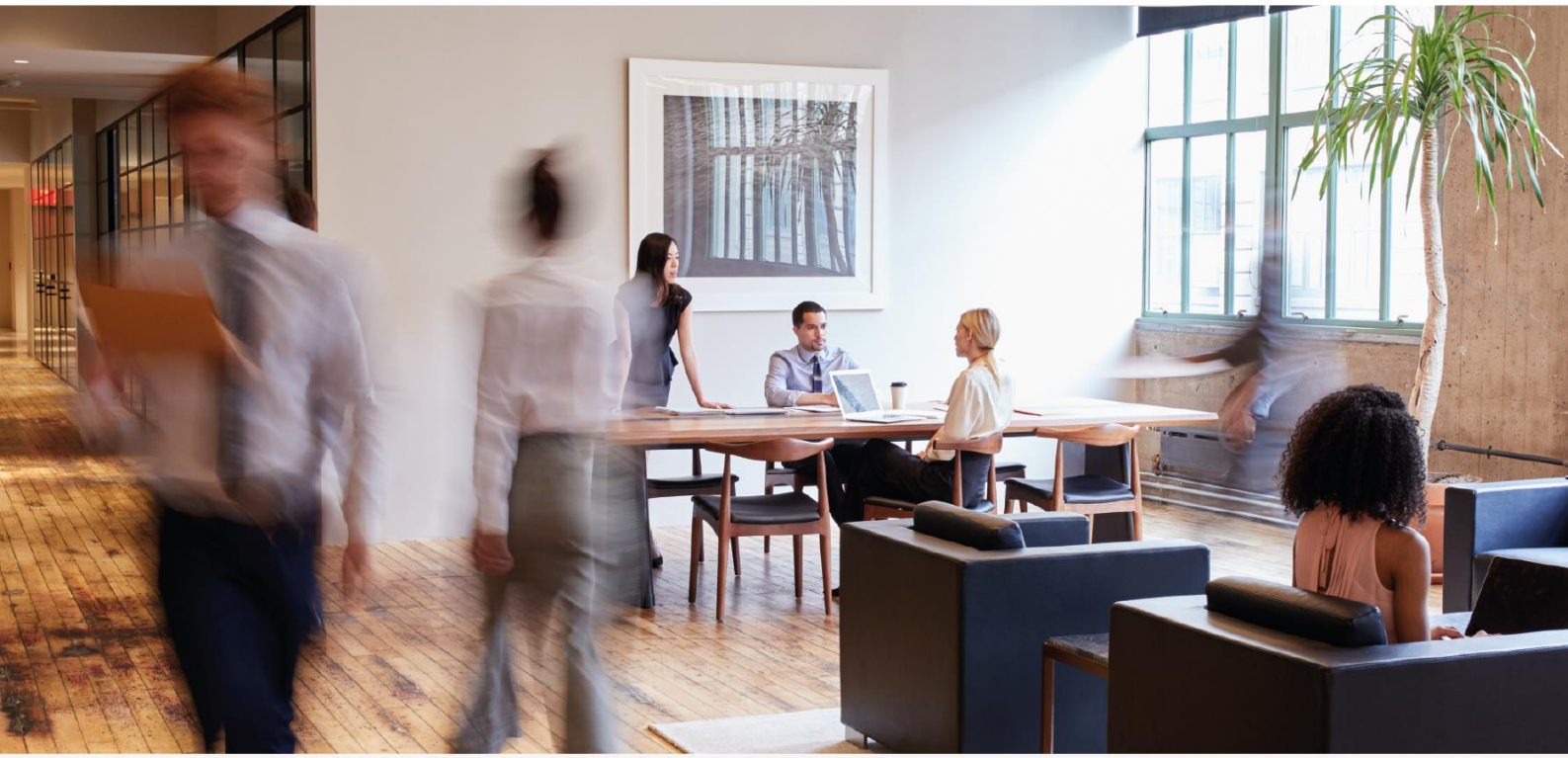
INOVIGATE
achieving outcomes



The Use of Real-World Data for Personalized Medicine

Multi-stakeholder roundtable outcomes on the use and reuse of routine care health data

Report

Authors:
Ingrid Maes, Eline Kok, Geert Dewulf

In support of:

PROJECT ATHENA

June 2022

Colophon

Title:	The Use of Real-World Data for Personalized Medicine *(in the context of the ATHENA project)*
Authors:	Ingrid Maes, Geert Dewulf, Eline Kok Inovigate is a strategy and management consulting company for health industries.
Expert contributors via interviews:	Stefan Gijssels, Diane Kleinermans, Stephanie Rossello, Paul d'Otreppe, Michel Legrand, Erik Ranschaert
Expert contributors via round table discussion participants:	Erik Ranschaert, Michel Legrand, Johan Van Bussel, Luk Bruyneel, Joke Vanlangenaeker, Giovanni Briganti, Yves Moreau, Joris Vermeesh, Karen Crabbé, Hans Constandt, Hanna Maertens, Bart Vannieuwenhuyse, Ingrid Maes, Geert Dewulf, Eline Kok
Commissioned by	Inovigate has performed this research and analysis under complete intellectual independence. The initiative took place in the context of project ATHENA, supported by VLAIO (ref. HBC.2019.2528).
Publication date:	June 2022
How to refer to this document?	Maes I., Kok E., Dewulf G., The Use of Real-World Data for Personalized Medicine, multi-stakeholder roundtable outcomes on the use and reuse of routine care health data *(in the context of project ATHENA*), 2022

Disclaimer	External experts have contributed via in-depth interviews. Input from these interviews was analyzed and discussed in a stakeholder roundtable and resulted in this report. The external experts did not coauthor this report and therefore did not necessarily agree with every element and/or recommendation in this report.

Preface

This report is the second part of a series of two reports linked to the use of real-world data (RWD).

On March 16, 2021, the Belgian Association of Hospital Managers (BVZD/ABDH) organized a seminar on "**Secure reuse of routine care data – Benefits & challenges**". Several aspects of secure and successful reuse of routine care data were presented. Also, several cases that benefit patients, the healthcare system and/or the life-sciences research ecosystem were shared by Ziekenhuis Oost-Limburg (ZOL), Onze Lieve Vrouwziekenhuis (OLV) Aalst, Centre Hospitalier Universitaire (CHU) Liège, Algemeen Ziekenhuis (AZ) Delta, and many others. As a follow-up to the seminar, a roundtable on "**Secondary use for real-world data and development of a data governance framework",** which took place on October 26, 2021, was organized by the BVZD/ABDH to further discuss challenges and opportunities, as well as potential solutions. The outcomes of this roundtable have been outlined in a white paper, "**Recommendations on a real-world data strategy for Belgium, a multi-stakeholder initiative on reuse of routine care health data, 2022".** This is the first report in our series about the use of RWD from the perspective of the BVZD.

To further build on the outcomes in this first report, an outreach to various stakeholders, such as authorities, patient associations, academia, hospitals, etc., was made to incorporate their views in the governance and sustainability of the proposed ATHENA platform during in-depth interviews. This was followed by a multi-stakeholder roundtable to align on a solution **framework for use and reuse of RWD in care and research**, including data process steps and the enabling foundation (like ethical/legal, governance, privacy aspects, etc.).

The **ATHENA project** (Augmenting THerapeutic Effectiveness through Novel Analytics) aims to develop a platform that can access large datasets on the evolution of disease in individual cancer patients. The aim is to search for valuable correlations between data (data mining), while still complying with the highest standards of patient data privacy and security by using a federated model. This approach makes it possible to discover disease mechanisms that can be treated through new personalized therapies. One of the important focus areas within the ATHENA project are the **governance and organizational** aspects of data access and management to ensure sustainability and future value creation. This includes the **legal and ethical** aspects linked to data use.

In this report, the ATHENA project will be detailed, other initiatives using the same underlying methodology will be briefly explained, and feedback from the multi-stakeholder roundtable on this topic will be incorporated.

Table of contents

Table of figures

Executive Summary

Although our growing scientific understanding and improvements made in medical technology lead to significant advances in human health and wellbeing, many therapy strategies remain insufficient, inappropriate, or heavy on side-effects for patients. It is therefore relevant to seek for the limiting factor of research and improved therapy development. Typically, to obtain novel insights in a disease and its therapeutic possibilities, clinical trials are established, where dedicated participants are followed intensively over a short period of time. However, the small number of participants–and thus patient data–limits the variability in the test population and thus the relevance and applicability of the results to a broad audience. This is even more the case for personalized medicine and for rare diseases where large data sets from clinical trials are not available. A promising strategy to overcome this obstacle is to include all data from patients suffering from a certain condition, including those outside of clinical trials, residing in a standard clinical setting. **Using this real-world data (RWD) allows the discovery of patterns of disease progression and treatment response** through machine learning. Early identification of these patterns in patients can lead to an optimal diagnosis and treatment strategy adapted to the patient's needs, called personalized medicine.

This paper is part of a series of two reports dedicated to the use of real-world data in personalized medicine. The first report called "Recommendations on a real-world data strategy for Belgium, a multi-stakeholder initiative on reuse of routine care health data, 2022", was established after a seminar and roundtable organized by the Belgian Association of Hospital Managers (BVZD/ABDH), including stakeholders from many fields. To further build on the outcomes reported in this first paper, various stakeholders, such as authorities, patient associations, academia, hospitals, etc., were contacted to share their insights and opinion on the matter. Following these interviews, a **multi-stakeholder roundtable** was organized to discuss **the prerequisites of efficient use and reuse of RWD in healthcare and research**. Project ATHENA (Augmenting THerapeutic Effectiveness through Novel Analytics) was used as one of the example initiatives where data collection, management, and analysis are strategically combined with high standards in patient privacy and data security by using a federated model, applied to the specific field of oncology. The views of the expert stakeholders were combined with the ATHENA solution and serve as the basis of this second report.

During the roundtable, all stakeholders were convinced that there is a **need for an intensified use of real-world data**. RWD offers many advantages, of which improvement of the quality of care is the most important. However, stakeholders are aware of the challenges of reusing healthcare data and agree that action needs to be taken at all levels to enable data management and sharing. By setting up strategic collaborations between hospitals in Belgium, we can do better than other European countries and set a standard for RWD use in healthcare. A **top-down and bottom-up approach** are put forward, integrating the common ground solutions for intensified, yet strategic use of RWD. Top-down, the government, potentially through the creation of a Belgian Health Data Authority, is responsible to build and implement a legal and standardized framework for reuse of data and should motivate participation by hospitals. Bottom-up, clinicians and hospitals are responsible for a uniform data collection and organization strategy.

Based on the information obtained from the roundtable discussions, this report specifically lists **conditions for data processing and reuse**. The minimal requirements for RWD reuse are a simple and practical method, responsibility from all stakeholders, an efficient data collection system, funding, and a uniform consent model. Further recommendations are made for enabling **foundation and governance of the data**, in which ownership and controllability of the data are essential and transparency is key. A specific role for the Belgian Health Data Authority is proposed here. Finally, the report also models these parameters to the **privacy-preserving federated learning strategy of project ATHENA** which, for now, is still a research project, but mimics the multi-dimensionality of a similar endeavor at a larger scale, including General Data Protection Regulation (GDPR) and ethical considerations.

All stakeholders agree that Belgium has all the ingredients needed to obtain a prime position in use and reuse of health data for research. First, the **ecosystem needs to be united** over this common goal. By working together, we can advance personalized healthcare and pave the way forward for the country. Secondly, **a federated approach** is to be established, as exemplified in project ATHENA, as it combines

efficient data management with adequate data security. Thirdly, **boosting talent in our region** to catalyze these innovations is crucial. Education and training will prove to be indispensable to reach our goal.

1. Background of the RWD strategy and structure

A structure for the RWD framework was previously created, based on a literature review and the experiences gathered from other European countries.

In this report, only the highlights will be duplicated and the reader is kindly invited to start by reading the first report of the series: "Recommendations on a real-world data strategy for Belgium, a multi-stakeholder initiative on reuse of routine care health data, 2022"

The proposed structure for the RWD framework consists of two circles (figure 1):

- The inner circle, which focuses on the **data processing steps**: data collection, data quality, and data processing;
- The outer circle, which allows us to dig into the **enabling foundation** that is needed to support the data process steps, such as governance, funding, and data privacy/security.

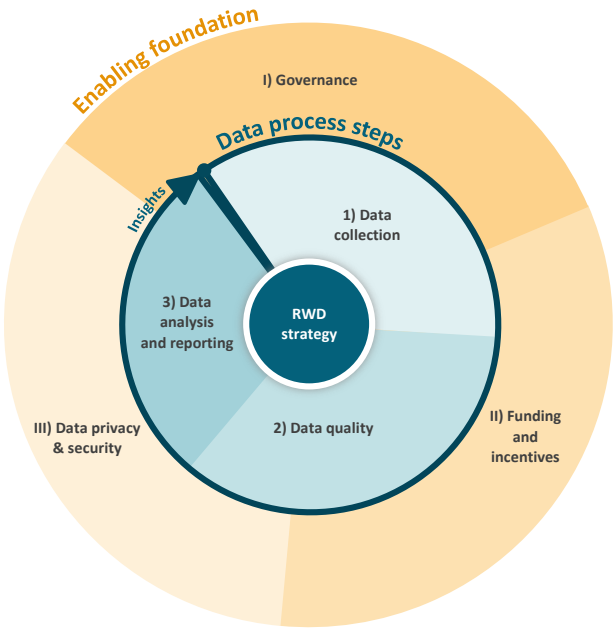


Figure 1: Real-world data solution framework

1.1 Basic principles for routine care data reuse

Principle 1: Key principles agreed upon during the roundtable for data collection are to be pragmatic in defining which data to collect and to collect data only once to avoid duplication of efforts. **One common patient consent model for the whole of Belgium**, based on an opt-out mechanism, is preferred. Data collection should go beyond the hospital and should also include InterMutualistic Agency (IMA) data, containing information from the sick funds, as well as RIZIV data. On top, patient reported outcomes measures (PROMs) should be collected. Hospitals are also accountable for supporting outcomes-based patient or value-based healthcare instead of today's disease-centered care. The key to this is the establishment of a rewards mechanism that supports **output-driven** behavior.

Principle 2: A **minimal standard of exchangeability should be imposed via electronic medical records (EMR) accreditation**, similar to that used for general practitioners' software and in the Obama Care Act, and via hospital and hospital network funding linked to the obligation of data reuse. A common internationally accepted data model between hospitals should be agreed upon. Data should be standardized and put in a common data model.

Principle 3: The **quality and completeness** of data is important. Therefore, funding and compensation for data quality efforts should be made possible.

Principle 4: A **combined private/public funding** is considered a good approach for supporting the interests and capabilities of both.

Principle 5: A **governance model on the hospital network level** must be established.

Principle 6: On a Belgian level, a **multi-stakeholder representation in the governance body** that decides on data access is required. This governance body should include all stakeholders, including pharma.

1.2 The need for a second roundtable

Apart from the technical and clinical objectives of this project, which have been discussed during the first roundtable, it is vital that many other aspects are also accounted for and are further elaborated, including the following:

- Governance and organizational aspects of data access and management to ensure sustainability and future value creation;
- Legal and ethical aspects of the project.

This involves an outreach to various stakeholders, such as authorities, academia, and patient associations to incorporate their views in the governance and sustainability of such a platform.

Taken together, it is the aim to create a sustainable project environment where results can be exploited in future research and development projects in other disease areas and participating medical centers. The roundtables with all the stakeholders are a key element to move forward in the realization of this project, as the outcome exposes valuable insights into the needs for real-world data and the conditions for the reuse of real-world data.

2. Project ATHENA

As an tentative answer for the underlying RWD reuse platform, project ATHENA was proposed to multi-stakeholders.

The path forward in personalized medicine is about connecting...
connecting data hubs, connecting institutes and connecting partners.

ATHENA (Augmenting THerapeutic Effectiveness through Novel Analytics) is a collaborative network that brings together a unique, multidisciplinary, and complementary partnership of academia, hospitals and industry to explore and use the concept machine learning for the realization of predictive analytics in oncology. By creating a **federated and standardized analytics platform**, it will be possible to combine different data types into one predictive model. It will allow access for partner hospitals and companies to detect and validate new therapies while fully preserving the privacy of patients.

ATHENA brings together experts in the field of data integration and analysis, allowing participating medical centers to utilize their patient data to optimize care.

Treatment of cancer is currently insufficient, inappropriate and heavy on side-effects for many patients. So, what is hampering research in finding novel and more adequate therapy strategies for everyone who needs them? A typical methodology for obtaining novel insights in the diagnosis, progression, or treatment of disease is to set up defined clinical studies. However, the small number of participants–and thus patient data–limits the variability in the test population and thus the relevance and applicability of the results to a broad audience.

Clinical studies only cover a selection of the population, leading to biased care recommendations that do not apply to everyone.

A promising novel strategy for overcoming this obstacle is **to include all data** from patients suffering from a certain condition. This so-called real-world data (data obtained from patients in a standard clinical setting of care) allows the discovery of different patterns of disease progression and drug response through machine learning. Early identification of these patterns in patients will lead to an optimal diagnosis and treatment strategy adapted to the patient's needs and aiming at the maximization of survival rates.

Machine learning allows us to recognize patient markers invisible to the naked eye and link them to diagnosis and treatment outcomes.

Changing the landscape of medical care, step by step.

The only way to benefit from the full potential of real-world data is to transition from individual hospital silos to **networks of excellence and expertise**. It is also necessary to minimize hospital boundaries so that all data from patients suffering from the same condition can be included and the insights are not limited to the smaller number of patients treated at a specific hospital or site.

The patients' involvement and consent are key at every stage in the proposed medical care landscape; hence, conserving and respecting all **privacy and security** aspects of the data is of utmost importance throughout the process.

For each site in the network, the medical data is locally integrated, (pre-)processed, and further enriched with omics data in a local repository in a common data model. On each site, a biomarker and knowledge discovery process will be performed on the local data by means of machine learning algorithms, which have been trained by aggregating the model parameters (not the data) over the different sites.

This knowledge will be used to **optimize the individual patient care pathways**, and furthermore, it will allow us to gain a **deeper insight** for better drug target discovery by combining the pseudonymized insights from each site and performing a more holistic analysis (figure 2).

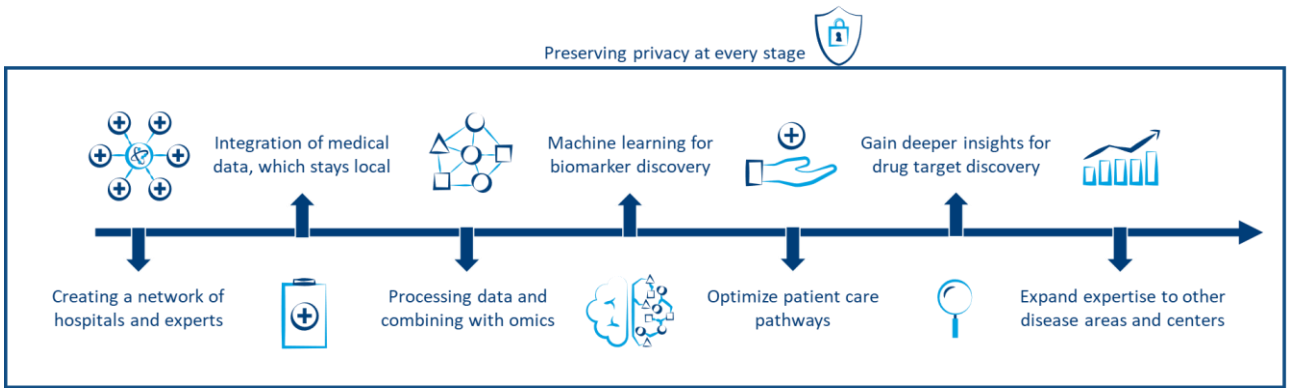


Figure 2: The improved landscape of medical care presented in ATHENA

With benefits for everyone involved

All parties involved will benefit from this federated learning approach:

- The patients will benefit from significantly **improved treatment**. Inclusion offers a personal involvement in care optimization;
- Healthcare professionals will be able to provide optimized and personalized care to patients. Early diagnosis and improved treatment will lead to **better disease outcomes**;
- Life science companies and institutes can **accelerate research** on drug targets through real-world data and biomarker discovery.

The objectives of ATHENA

Project ATHENA aims to:

Facilitate the investigation and creation of a **federated** and **privacy-preserving machine learning** platform within hospitals.

Combine **clinical data** (medical history, treatment, imaging) with **omics data** (genomic mutations) to build enriched data sets.

Help healthcare practitioners (HCPs) to deliver better targeted treatments more effectively by exploring the potential of **care pathway automation.**

Explore the concept of machine learning to establish **predictive analytics** in oncology as a novel approach towards personalized medicine.

Gain deeper **insights** in two important **cancer types** with high medical and societal need (bladder cancer and multiple myeloma).

Build a foundation for **future expansion** into other diseases and institutions.

Led by a multi-disciplinary partnership

The consortium of project ATHENA consists of several top-notch partners, each contributing great knowledge and expertise to the project.

Inovigate manages the project flow in ATHENA and jointly coordinates data gathering and storage.

Janssen coordinates project ATHENA and contributes in particular to the setup of a standardized data collection system, based on patient records from participating hospitals.

KU Leuven facilitates the collection of clinical/laboratory data and analysis using state-of-the-art techniques, with a focus on bladder cancer.

UGent facilitates the collection of clinical/laboratory data and analysis using state-of-the-art techniques, with a focus on multiple myeloma.

Imec jointly governs the omics data platform and guarantees a fully secured data integration, processing and analysis pipeline using specialized algorithms.

Illumina jointly manages the omics data platform.

Robovision handles the processing and analysis of visual patient data, such as MRI, CT scans, etc.

AZ Groeninge is data partner of project ATHENA, specifically for bladder cancer.

CHU Liège is data partner of project ATHENA, specifically for multiple myeloma.

OLV Aalst is data partner of project ATHENA, specifically for bladder cancer.

A source of inspiration for the ATHENA project can be found in the work done by Prof. Yves Moreau, KU Leuven, and his team in project MELLODDY, short for Machine Learning Ledger Orchestration For Drug Discovery. This project is an Innovative Medicines Initiative (IMI) 2 joint undertaking supported by Horizon 2020. Keeping this work in mind, ATHENA wants to analyze omics data together with clinical data residing in different hospital locations while preserving patient privacy.

3. Other initiatives based on a federated or alternative innovative data models

We have highlighted some interesting alternative initiatives below as a potential source for inspiration for Belgium.

German Cancer Consortium's Joint Imaging Platform

The strategic initiative "Joint Imaging Platform" establishes a distributed IT infrastructure for image analysis and machine learning at all *Deutsches Konsortium für Translationale Krebsforschung* (DKTK) sites. It will facilitate pooling of analysis methods that can be applied in an automated and standardized manner to the patient data of different centers. The underlying infrastructure facilitates applications such as federated learning across multiple clinical centers.

MELLODDY

MELLODDY aims to leverage the world's largest collection of small molecules with known biochemical or cellular activity to enable more accurate predictive models and to increase efficiencies in drug discovery. This project aims to enhance predictive machine learning models on decentralized data of ten pharmaceutical companies, without exposing proprietary information.

EHDEN European Health Data and Evidence Network

EHDEN aims to develop a federated and equitable ecosystem of institutions generating clinical data, with researchers across academia and industry. It is supported by certified and qualified SMEs, harmonizing clinical data and creating a network technology for real-world research. It will incorporate appropriate policy and regulatory requirements, such as the GDPR and ethical research guidelines, via privacy by design in its sociotechnical infrastructure.

HealthChain

This platform is built to address data interoperability challenges between the payer and provider ecosystem by bringing data from multiple providers all to one place. Since blockchain acts as a distributive ledger that keeps track of all transactions, it ensures the security, transparency, and mobility of data. In addition, patients can grant users access to their electronic health records (EHR) and revoke access as needed. This enables patients to communicate directly with their physicians and share their health records for online consultations.

DARWIN EU

The European Medicines Agency (EMA) is establishing a coordination center to provide timely and reliable evidence on the use, safety, and effectiveness of medicines for human use, including vaccines, from real-world healthcare databases across the European Union (EU). This capability is called the Data Analysis and Real-World Interrogation Network (DARWIN EU®). DARWIN EU will deliver real-world evidence from across Europe on diseases, populations, and the uses and performance of medicines. This will enable EMA and competent national authorities in the European medicine regulatory

network to use these data whenever needed throughout the lifecycle of a medicinal product. DARWIN and EHDEN are closely linked together.

4. A proposal for a Belgian RWD framework based on multi-stakeholder engagement

To **facilitate the ATHENA project** and beyond (continuous sharing of data as part of a continuous learning environment), the consortium brought stakeholders together at a **roundtable** to **align on a solution framework** for use and reuse of RWD in care and research, including data sharing models, ethical/legal aspects, consent, data donation, and incentive mechanisms.

Stakeholder interviews and ATHENA roundtable to define common ground

To gain deeper insights, several representatives of key Belgian healthcare system stakeholders were interviewed. The interviewees selected are considered experts in the use of health data, each of them active in a very specific domain (figure 3).

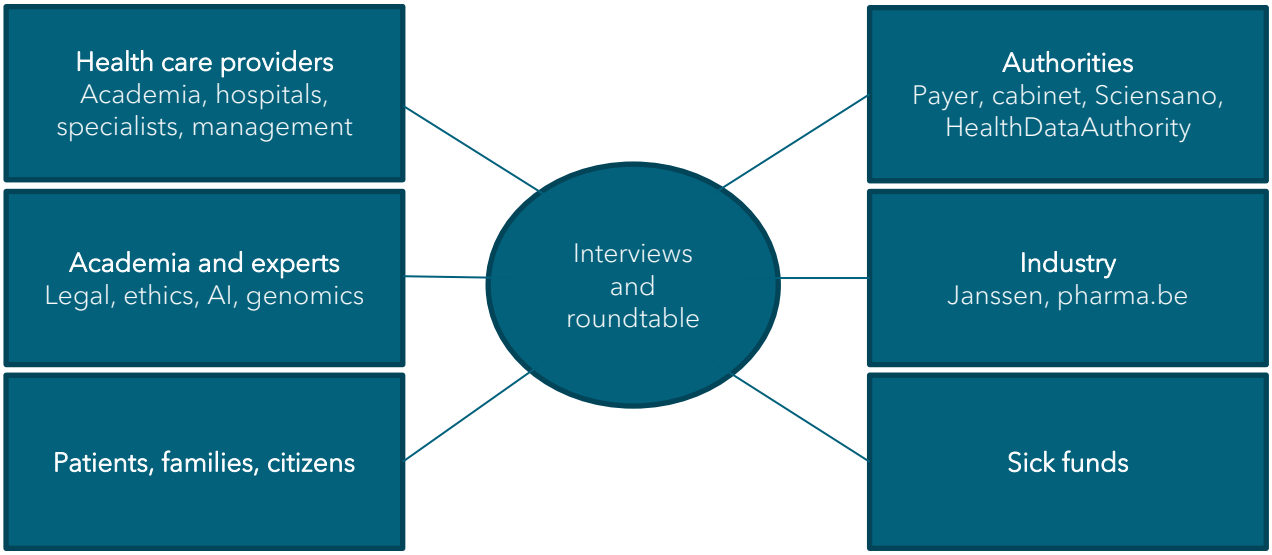


Figure 3. Interviewees and participants of roundtable

For these **interviews**, a questionnaire was developed to discuss the following topics:

1. Use and reuse of real-world data in general;
2. Decision framework to create a supporting base and solution for use and reuse of RWD;
3. Use of a federated data model.

Insights from the interviews were compiled and led to common ground solutions per building block of the RWD framework. These common ground solutions have been proposed to the **roundtable** participants and have been subject to debate with the aim of creating a support base for the use and reuse of health data. Fifteen participants were gathered on February 24, 2022, representing a cross-section of stakeholders from both the northern and southern parts of Belgium. The roundtable discussion brought many new insights, and a consensus emerged on many topics. During the roundtable, it was confirmed that everyone is convinced that there is a **need for an intensified use of real-world data** (figure 4), and that action needs to be taken at all levels to make this possible (table 1).

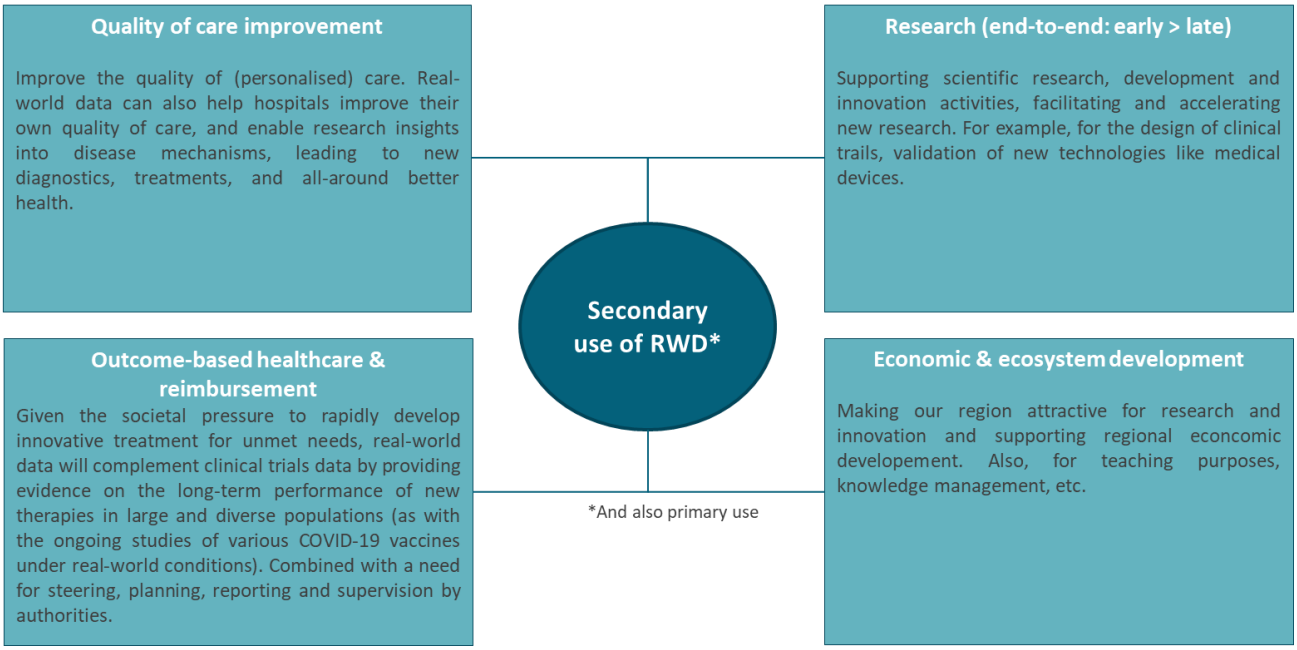


Figure 4: Needs for the secondary use of real-world data

Table 1: The agreed common ground, discussed during the round table, on theactions

RWD offers advantages	• Belgium could position as a leader in reuse of RWD by connecting all data from primary care, secondary care and all other available databases. Only a few countries are doing this today; • When RWD is used for benchmarkdiseases useful insights could be obtained. The doctors, who are not always aware of real-world data today, will start to see the benefits, which improves the quality of data collection.
We need to look at others and do better	• We need to learn from other countries and apply best practices or even go beyond that. • We should build a system together with physicians and nurses., based on good use cases, so they see the benefit and remain motivated and incentivized.
Based on (international) standards	• Europe is positioning the patient as the owner of the data. We need to explain to patients the benefits of this principle, but also the dangers; • There is a great need for a foundational layer with core data available to all (like: treatment, gender, age, etc.). Depending on the use case (for example for reimbursement), additional data layers can be added; • All participants agreed that the basic data layer should be accessible to all. If an extra layer of data is needed, the requester should pay for it, on a fee for service basis.

Despite the challenges, there are major motivations for routine care data reuse. **Improvement of the quality of care** is the most important reason why we should collect data and reuse it. Not sharing data has been considered unethical by the participants in the roundtable, because access to any available data on patients is required to give the best possible care. In the digital world, patients expect digital services and the use of data to support data-driven medicine. This is a duty. Reuse of clinical data is essential to fulfill the promises for high quality healthcare, improved healthcare management, and reduced healthcare costs.

Moreover, quality measurement and learning healthcare systems, as well as effective clinical research, are major objectives.

To achieve this, **a data culture must be established in which outcomes are compared and benchmarking between hospitals is performed**. Hospitals should work together and take action to collaborate at all levels. We must be more ambitious and set ambitious goals. The Scandinavian use of registries is an excellent inspiration, but we must do better. Excellent, complete, and well-validated data allows improvement of care.

This can be achieved through a combination of **top-down** and **bottom-up** approaches.

Top-down approach

In the top-down approach, the **government needs to ensure that the legal framework** is in place to facilitate the safe and ethical reuse of health data for research with a clear GDPR interpretation. Patients also need to be made aware of the importance of sharing data, as it helps improve healthcare for all. As with organ donation, we need an 'opt-out' system for data donation rather than an active 'opt-in' method. This is already being done in Scandinavian countries.

The government will have to act to implement the framework and motivate hospitals and all owners of health data. The government should **impose data standards**, including interoperability standards, for hospital EMR and harmonize data infrastructures across diseases (via a broadened cancer registry) and international initiatives (via European Reference Networks).

The creation of the Belgian Health Data Authority or even a broader Belgian Health Data Institute could bring more value to the broader stakeholder community. It would enlarge the scope of health data beyond the five federal organizations by also including data from the IMA/AIM database (containing data from sick funds), e-Health data, and health data from primary and secondary care.

Bottom-up approach

In the bottom-up approach, clinicians are the experts in defining the core data set. They should define, together with the patients, the frequency and duration of data collection and optimize clinical routines based on data collection needs, adapting the clinical guidelines.

Hospitals must develop a better strategy to organize their data internally and make it accessible. Some hospitals in Belgium have dozens of different internal systems and databases with patient data. The data landscape is extremely fragmented within the healthcare system. There are plenty of technical solutions to data problems, but leaders need to agree on which to implement.

In this approach, hospitals will take action to develop the framework. They must **take the initiative to securely share data between them using a common model to enrich data.** Investment in a good EMR is essential because it will solve many of the challenges related to data reuse, as data will be collected in a structured way. There are no legal or other hurdles to start with data curation. This could be done immediately. Hospitals should demonstrate proactivity by setting up their data strategies and starting initiatives to share data among them. They should also use the common data model to enrich data.

During the roundtable, three main topics were discussed:

- **Minimum conditions** for data processing steps;
- Requirements for enabling **foundation and governance**;
- Models of **privacy-preserving federated learning** and whether ATHENA is the right model for Belgium.

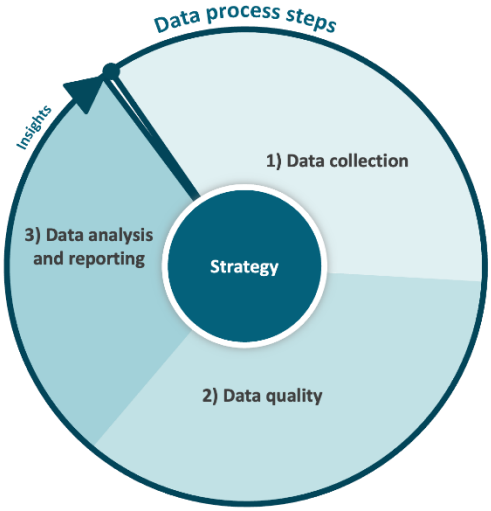


Figure 5. Data processing strategy

Conditions for data processing and reuse

Before anything else, the **data strategy must be clarified** (figure 5). This means agreement on the purpose of gathering data is required, including defining the "use case". As there are many "use cases" possible, very different approaches and methods may be necessary. In the case of personalized medicine, access to raw data may be necessary, whereas in a population management case, aggregated anonymous data would do. Once a definition is available, one can decide what data is needed. Sufficient stakeholders should be involved in deciding on this. Many people may make proposals but, in the end, clinical experts need to decide.

The data collected may consist of many different layers, starting with a foundation layer of data (treatment, gender, age, etc.). This is the core dataset. A core dataset needs to be set by clinicians, not by researchers. What researchers want is different from what is found in a clinical setting. Depending on the use case, you can add additional layers of data and go deeper. For market access use, you will need data other than routine care. Research purposes will require an augmented data layer of real-world data.

Building a high-quality data set will require **efforts from hospitals**. Ten years ago, every hospital had to look for an EHR, not at all structured, and with a lot of free text. Each organized the data their own way, as this EHR was only meant for internal purposes, with a clear uni-directional data-flow in mind. In addition, several hospital departments started to collect data in individual data formats and stored in many separate silos. This led to a fragmented and scattered situation, with data that were not immediately usable. More recently, different applications linked to specific diseases or tasks were introduced and produced an even more unstructured dataset. The introduction of a common coding system or ontology for all health related data is key. This coding needs to start at the hospital level, but a governmental authority such as the Belgian Health Data Authority should impose the technical standards and codes to be used. But we should also consider quick wins. Why don't we start using claims data, as they contain a great deal of interesting information? Or consider putting IMA data on the Observational Medical Outcomes Partnership (OMOP) common data model platform?

Obviously **working with real-world data is complex**. In an EHR, the most interesting information can be found in the free text; hence, the free text needs to be coded. Imposing coding means that you are impacting the clinical workflow. Most of the hospitals have limited knowledge on coding (in Snomed Clinical Terms); they are missing skilled people and need researchers or data managers to interpret the codes. Moreover, you add an administrative burden to the physicians, as we need them to validate the coded information. It all starts with a **vision of a hospital** convincing its health care providers of the benefits of using health data and a structured EHR. To get these on board, you need to show them the advantages of such a system. AZ Sint-Lucas, Ghent, had a hard time selling data structuring to physicians. They built a patient-centric data warehouse allowing benchmarking of certain diseases (MS, IBD, etc.). This registry started to benchmark the diseases and brought novel useful insights. As the physicians started to see the benefits, the quality of the data set improved dramatically. Some hospitals, such as CHU Erasme, started using Snomed CT to obtain structured data throughout the hospital under the clear guidance of the chief medical officer.

It was obvious to all roundtable participants that the **training and education** of all stakeholders is crucial.

- The physicians need to be convinced of the advantages and learn to **record data** in a structured way;
- The IT staff needs to get the necessary **skills to code** (in Snomed CT);
- Patients, once convinced of the advantages of sharing data, need to be **aware of possibilities** such as a health wallet.

Further recommendations by roundtable participants are given in table 2.

Table 2: Summary of recommendations and observations from the roundtable on the minimum conditions required for real-world data processing and reuse, as well as governance

We need **a simple and practical method**	• Each hospital uses many apps and has many systems to collect data in different data types, all of this very fragmented and not usable. A solution must be found for this; • The clinicians indicate that the government should determine a data collection method (software), only then will everyone do it with the same methodology. Today, there is hesitation in choosing because there is a fear that the government will impose another one and the investment will be for nothing; • A condition has been stated, namely, that it must be possible to follow a patient when he goes to different hospitals, which is often necessary for rare diseases.
We need **responsibility**	• Deciding which data to collect should be a multi-stakeholder decision with the expert clinicians as the final decision-makers.
We need **efficient data collection**	• Before discussing the core data set, we need to agree on "the purpose". This starts with the definition of "use case". Once a good definition is available, we can decide on what data we need; • The first step is a coding system at the hospital level; this needs to be organized and imposed. The most interesting things are filled out in free text, so the free text needs to be coded; • Imposing coding means that you are impacting the clinical workflow. Most of the hospitals have limited knowledge on coding (in Snomed CT); they are missing skilled people for this coding and need researchers to interpret the code; • You also add an administrative burden as we need validation by an MD. A solution must be provided for this; • There was no consensus around pseudonymized data, as in many cases it is still necessary to go back to the raw data; • On synthetic data, participants agreed that this is a good start, but not the right methodology to make real decisions.
We need **funding**	• We need to incentivize the hospitals to make the effort, and they need a great deal of training to obtain the necessary skills; • Training is not only necessary for physicians but also for IT experts within hospitals;

	• Incentives must be handled with care. We must learn from the incident in the UK (referring to a massive opt-out by patients after a negative press campaign).
We need **a uniform consent model**	• On one hand, if asking among patient associations and patient groups, most see no problems with sharing data. However, Sciensano received daily complaints about the use of personal data during the Covid-19 crisis; • All stakeholders agree that transparency is key. When patients sign today, they do not know what they sign. If you talk to the patient and explain the need for data, tell and explain the patient, you will create trust; • The legal team at Sciensano is too small and needs help.

Conditions for enabling foundation and governance

Europe is positioning the **patient as the owner** of the data. We need to explain everything to the patients. Incentives are very tricky (see the UK example referring to massive opt-out by patients after negative press campaign). We need to explain to the population that data reuse helps and brings many advantages. On one hand, if asking among patient associations and patient groups, most see no problems with sharing data.

GDPR should not be a barrier if the patient is properly informed, and consent is well organized. According to the European Patient Forum, 80% of all patients are willing to share data if there is a benefit for the population. Patient organizations can help in communicating this information to patients. **Transparency** is key. When patients sign a consent form, they do not always know what they sign. If you properly explain the process to patients, you create trust. In a local experiment, only 4% of patients did not want to share their data. These are arguments for an opt-out system. During the Covid-19 crisis, Sciensano received daily complaints on the use of personal data. To answer these questions, the legal team of Sciensano needed help from external parties. We were in a pandemic and built a system with a legal basis to create a safe environment to collect data that people can trust.

Legal gaps might be an important hurdle

Next to the technical engineering needed at all levels to make health data usable, a great deal of **legal engineering** will be needed to facilitate these technical aspects. This legal engineering needs to be smart and keep all options open. Again, the methods should be selected based on the use case. However, there is a high level of urgency to get this done before we lose speed. The Belgian Health Data Authority picked up this urgency and started with a work package on the legal aspects, together with an external party. They want to involve all stakeholders with previous experience in this legal framework and come up with a proposal by the summer of 2022.

The description "**data ownership**" needs to be used carefully. Many parties pretend to be owners of health data: patients, physicians, hospitals, payers, etc. We should consider the physician the creator of the data file rather than the owner. A better description could be "**data control**" or "**data-controller**". A data control framework can be a good framework as it covers all aspects.

Patients also become important contributors to their own health data files

The use of technology in combination with a shift toward patient-centricity in healthcare has resulted in an opportunity to create **patient-generated health data**. These can be created from active and passive data sources. Active sources provide direct information obtained, for example, through questionnaires. Passive sources use information obtained via smart devices or health apps. Patient organizations report that most patients are eager to share these data if they help to improve care for themselves as well as for the population.

Next to technical challenges, such as new analytical tools needed to successfully use these data and adjustments to be made to existing workflows and compliance rules, extra care will be needed for data

security and privacy compliance when collecting and combining such data from different levels and sources. Why not gather all these files in a patient's personal health wallet on their phone, including a widget where patients can accept to participate in research projects?

The role of the Belgian Health Data Authority

The roundtable participants agree that the Belgian Health Data Authority comes at the right moment and is instrumental in positioning Belgium as a prime RWD site. Their approach allows access to data for secondary use from different governmental sources to gradually expand to other organizations, using a "federated approach" as they will focus on metadata and help others find what they need. Moreover, they will sit in the driver's seat to set up a governance model with all stakeholders. The Belgian Health Data Authority (BHDA) could become the higher-level authority in Belgium deciding who can use data and under what circumstances. Next to the BHDA, we might need a trusted third party (TTP) that could independently decide on secondary data use, work closely together with the privacy commission and be in charge of anonymizing and pseudonymizing of health data.

Today, a **legal framework** on how to reuse data for secondary use is still missing in most hospitals. This becomes even more apparent considering the missing trust among some HCP's if an industrial company is involved in the reuse of these data, although others see no difference whether a hospital or an industrial company is analyzing, for example, genomics. The roundtable participants applauded the BHDA for taking the lead in setting up a legal framework in the short term and hoped it would use the knowledge gathered by the many parties already working on such a framework.

The legal framework to be used has a lot to do with **the way you need access to the data**. As mentioned earlier, this very much depends on the use case. For population management, aggregated anonymous data will do perfectly. When using data for personalized medicine, however, you obviously need to be able to return to the original raw data. An interesting example was given based on the data warehouse (DWH) used for disease benchmarking. By analyzing their DWH, they observed that one treatment was noticeable, doing worse than another. Upon such findings on disease and/or treatment, you need to be able to go back. Using pseudonymized data where the hospital (or a trusted third party) has the key to de-pseudonymize the record can solve this issue.

Rare diseases can be tricky as a patient can pop up in different databases in several hospitals and be counted several times. In those cases, you need to be able to follow a patient through time and geography. You may consider using an extra layer with a **unique identifier** to identify doubles, but this may need serious data exchange and data cleaning efforts. Data duplication and replication should be limited to a strict necessity.

5. Is Athena, using a privacy-preserving federated approach, the right model for Belgium?

Healthcare institutions produce huge amounts of data linked to the health of their patients. Having access to all this data could improve the quality of diagnostics and therapy. However, there is only limited access to this data outside the hospital where the data has been generated. Reasons are multiple: unclear ownership of data, lack of consent to use the data, GDPR, different interpretations of privacy laws, etc. This has led to the compartmentalization of health data with silos of different architecture using different data formats, standards, and security protocols.

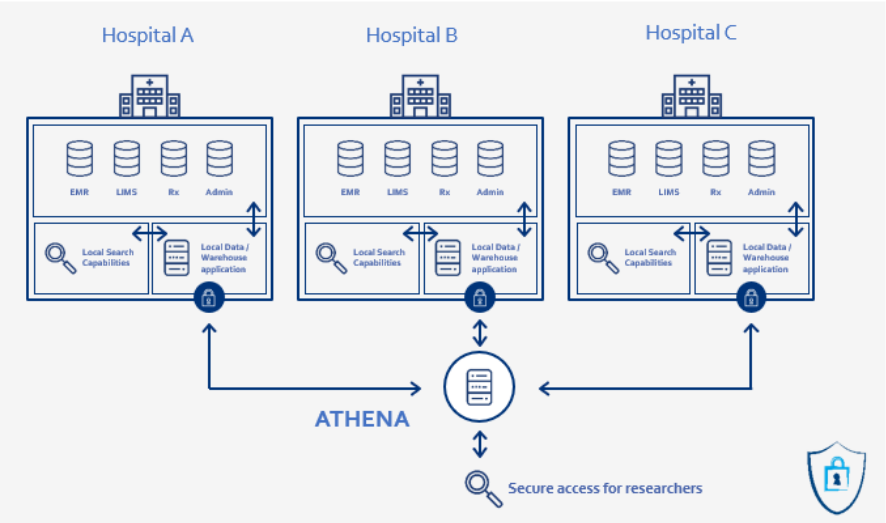


Figure 5. Federated and privacy-preserving architecture

In some situations, data are pooled into centralized health databases with growing concerns about necessary storage capacity and the need to duplicate the data. As an alternative, **federated data systems** have been proposed to address current barriers to access to siloed data. Using this approach, we do not share, transfer, or pool data but rather allow querying or visiting data while they remain in their original place.

ATHENA is using this federated approach as a research project. The roundtable participants agreed on its many advantages, but questioned how quickly it will be fully ready for routine use. There is an economic reality behind every project, and much needs to be done before going beyond research. It may well take multiple years before we can scale what ATHENA is proposing to 100 hospitals, and 300 different projects. It will take time to have it function 24/7 throughout Belgium, keeping all aspects, such as performance, robustness, reliability, security, privacy, etc. in mind. The immediate importance of ATHENA is its **multi-dimensionality** with different actors, purposes, and use cases. As everyone wants to solve many different clinical problems, interesting and to-the-point use cases need to be defined in order to get quick results and insights in the potential of a platform such as ATHENA. If the use case is successful, it will help convince everyone that this approach can be successful and beneficial for everyone.

An additional application for a federated model is **combining and integrating clinical data with data from Sciensano and IMA**. The use of a unique patient identifier could be helpful in achieving this. It is obvious that one of the reasons to use a federated approach is linked with the privacy of data and securing GDPR compliance. Researchers at the KU Leuven Center for IT and IP Law have been working on the MUSKETEER project, where they have been providing legal and ethical guidance on the development and use of machine learning algorithms to augment shared knowledge in federated privacy-preserving scenarios.

The major benefits they see of federated learning, are the GDPR benefits:

- Compliance with the principle of data minimization, as no transfer of data is needed;
- (Indirect) compliance with the principle of purpose limitation, as data are not on central server;
- Less vulnerability to certain privacy attacks.

Some GDPR challenges remain to be addressed:

- Which data qualify as personal data?
- Who is responsible for GDPR compliance (controller/processor)?
- How can accurate predictions be ensured?

The roundtable also discussed an often proposed alternative: the use of synthetic data. Synthetic data can help start a scientific project but cannot be used for clinical decision-making or scientific publications. However, it takes a lot of work to make them, and you have many of your research questions answered with the same amount of effort. IKNL, the Integral Netherlands Comprehensive Cancer Institute, has large synthetic datasets, but they cannot be used for medical/clinical conclusions. IKNL is proposing this dataset for researchers who want to decide on what data they need to answer their research questions or to develop software and analytical methods with realistic results. They clearly mention that the dataset cannot be used for clinical decision-making or for scientific publications on cancer.

Conclusions

Belgium is home to one of the world's strongest health clusters and has long been ranked as the best country in the EU for clinical trials. However, if Belgium is to maintain this attractivity–and its quality of health–the country needs to **improve its position regarding the use and reuse of health data for research**. We have all the right ingredients in place in Belgium. In three years, we could be one of the top countries in Europe for real-world data. But we must act now and think in a multidisciplinary way.

Uniting the ecosystem for better data science in Belgium

Extensive multi-stakeholder engagement has resulted in this report. It provides input for further dialogues with stakeholders responsible for the RWD framework implementation: hospitals, care providers, governments, sick funds, payers, but also industry. Multi-**stakeholder engagement and dialogue** will be key to further detailing the plan and the priority actions and building consensus to put them in practice. Data use is integral to the future of healthcare; therefore, stakeholders from across the ecosystem need to come together to usher in a change in Belgium. By joining forces, we will be able to make a real difference to people's health, while also supporting the best interests of our different stakeholder groups. By working together, we can advance personalized healthcare and pave the way forward for the country.

Toward a federated approach in the future

Machine learning, and particularly deep learning, has led to a wide range of innovations in the area of digital healthcare and has unlocked doors to gaining insights that were previously impossible. The Federated Learning (FL) model that is proposed in ATHENA is a promising approach to obtaining powerful, accurate, safe, robust, and unbiased models from a decentralized setup. By enabling individual healthcare sites to **train collaboratively without the need to exchange or centralize data** sets, FL neatly addresses issues related to the egress of sensitive medical data. Consequently, it may unlock novel research and has the potential to dramatically improve cancer patient care globally. Important to note here is, that Europe is also looking at federated approaches to enable RWD for secondary use and hence our suggested approach will be compatible with the EU approach.

Today, we can see how FL has an impact on nearly all stakeholders and the entire treatment cycle, ranging from improving medical image analysis, providing clinicians with better diagnostic tools, and allowing true precision in medicine by helping to find similar patients to fostering collaborative and accelerated drug discovery and decreasing costs and time-to-market for pharma companies. However, not all technical problems have been addressed with a proper solution yet, and FL will certainly be an **active research area** throughout the next decade. Also, legislation will need to be adapted, and appointments of trusted third parties made to steer, guide, and eventually manage FL initiatives. Despite these inherent shortcomings today, we truly believe that its potential impact on precision medicine, by transforming real-world data into actual real-world insights backed-up by real-world evidence, will be to ultimately improving patient care.

Train our future

With the booming of technology and novel techniques and methodologies linked to data privacy, data security, and data science, it remains extremely important to **start training in the use of these approaches in data science at medical schools,** specifically on the more novel, innovative topics such as data privacy linked to data science.

Abbreviations list

ABDH-BVZD	Association Belge des Directeurs d'Hopitaux- Belgische Vereniging van Ziekenhuis Directeurs
ATHENA	Augmenting THerapeutic Effectiveness through Novel Analytics
DWH	Data warehouse
EHR	Electronic health record
EMA	European Medicines Agency
EMR	Electronic medical records
EPR	Electronic patient record
FAIR	Findable, accessible, interoperable, re-usable
FDA	Food and Drug Administration, USA
FL	Federated learning
GDPR	General Data Protection Regulation
HCP	Healthcare professional
HTA	Health technology assessment
IMA-AIM	Intermutualistisch Agentschap- l'Agence Intermutualiste
IMI	Innovative Medicines Initiative
MD	Medical doctor
MELLODDY	Machine Learning Ledger Orchestration For Drug Discovery
NIHDI	National Institute for Health and Disability Insurance (= RIZIV/INAMI)
OMOP	Observational Medical Outcomes Partnership (OMOP) Common Data Model
PROM	Patient Reported Outcomes Measures
RIZIV-INAMI	Rijksinstituut voor Ziekte- en InvaliditeitsVerzekering – Institut National d'Assurance Maladie-Invalidité
RWE	Real-world evidence
RWD	Real-world data
SMART	Specific, measurable, achievable, relevant, timely
WHO	World Health Organization

References

Annemans L. (n.d.). The use of real world data throughout an innovative medicine's lifecycle. *Fgov*. Retrieved from https://www.inami.fgov.be

Bégaud B., Polton D., von Lennep F. (2017). Les données de vie réelle, un enjeu majeur pour la qualité des soins et la régulation du système de santé. *L'Exemple du médicament*.

Belgische Kamer van volksvertegenwoordigers. (2019). Voorstel van resolutie betreffende het gebruik van big data in de gezondheidszorg. *De Kamer*. Retrieved from https://www.dekamer.be/FLWB

Biovox. (2021, 10 20). Belgium has all the ingredients to be number one for real-world data in healthcare. *Biovox*. Retrieved from https://biovox.eu/belgium-has-all-the-ingredients-to-be-number-one-for-real-world-data-in-healthcare/

Boyd M., Zimeta M., Tennison J., Alassow M. (2021). Secondary use of health data in Europe. *Open Data Institute, Roche*. Retrieved from https://theodi.org

Cave A., Kurz X., Arlett P. (2019). Real-World Data for Regulatory Decision Making: Challenges and Possible Solutions for Europe. *Clinical Pharmacology & Therapeutics*.

Consent working group (WG) of the National Steering Committee of the BMBF's MII. (2020). Guide to the Use of Nationally Harmonised Patient Information and Consent Documents for the Secondary Use of Patient Data. *MII*. Retrieved from https://www.medizininformatik-initiative.de

Counsel O. O. (2010). Compilation of patient protection and affordable care act. *U.S. House of representatives*.

De Cock J., & Kurz X. (2021). Co-Creating Real-World Evidence Excellence for Decision-Making: Meeting Regulatory and HTA/ Payer needs. *RWE4Decisions*.

EIT Health. (2020). Implementing Value-Based Health Care in Europe. *Handbook for Pioneers*

Facey K. M., Rannanheimo P., Batchelor L., Borchardt M., de Cock J. (2020). Real world evidence to support payer/ HTA decisions about highly innovative technologies in the EU-actions for stakeholders. *International Journal of Technology Assessment in Health Care*

FDA. (2018). Framework for FDA's: Real-World evidence program. *FDA*. Retrieved from https://www.fda.gov

FDA. (2021). Real-World Evidence. *FDA*. Retrieved from https://www.fda.gov

Fgov. (n.d.). De elektronische gegevensuitwisseling, het Hub-Metahub-systeem, de geïnformeerde toestemming. *Fgov*. Retrieved from https://www.ehealth.fgov.be

Hallock H., Marshall S.E., 't Hoen P.A.C., *et al*. (2021) Federated Networks for Distributed Analysis of Health Data. *Front Public Health*

Houts C. R., Patrick-Lake B., Clay I., Wirth R. J. (2020). The Path Forward for Digital Measures: Suppressing the Desire to Compare Apple and Pineapples. *Basel : S. Karger AG*.

IKNL. (n.d.). Synthetische dataset. *IKNL*. Retrieved from https://iknl.nl

IMI GetReal. (2013). Advancing Evidence Generation for New Drugs: IMI GetReal's Recommendations on Real-World Evidence. *IMI Europe*. Retrieved from https://www.getreal-institute.org

Kalliola, M. (n.d.). Ideas for: Joint Action for the European Health Data Space. *SITRA*.

KPMG. (2017). A blueprint for success in healthcare data and analytics (D&A). *KPMG*. Retrieved from https://assets.kpmg

Krebsregister Baden-Württemberg. (n.d.). Schulung für niedergelassene ärzte und Praxismitarbeiter: "Melderportal des Krebsregisters Baden-Württemberg". *Krebsregister.* Retrieved from https://www.krebsregister-bw.de

Lablans M., Schmidt E. (2020). Clinical Communication Platform (CCP-IT): Datenschutzkonzept. *Heidelberg: DKTK.*

Martinez J., & Tonon C. (2021). La gouvernance des données de santé: leçons de la crise du Covid-19 en Europe, en Chine et aux Etats-Unis. *Ifri: Etudes de l'Ifri.*

McGraw D., Mandl K. D. (2021). Privacy protections to encourage use of health-relevant digital data in a learning health system. *npj Digital Medicine*.

Miani C., *et al*. (2014). Health and Healthcare: Assessing the Real-World Data Policy Landscape in Europe. *RAND HEALTH Q*

Ministry of Social affairs and health Finland. (n.d.). Act on the Secondary Use of Health and Social Data. *Government Finland*. Retrieved from https://stm.fi

Mol P., Stausberg J., Ullmann A., Lange S., Burkhardt B., Braune S., Staeck F. (n.d.). Interdisziplinäre plattform zur nutzenbewertung: Welchen (Zusatz-) Nutzen haben Registerdaten? *Berlin: Springer Medizin verslag.*

Musketeer. (2021). Benefits and challenges of federated learning under the GDPR. *Musketeer*. Retrieved from https://h2020musketeer.medium.com

Nafsika H. K. (2019). Overview of the RWD/RWE landscape. *F. Hoffmann La-Roche.* Retrieved from https://cddf.org

Novation E. *et al*. (2012). ICT-standaarden in de zorg: Een praktisch overzicht. *Nictiz Den Haag.*

OECD. (2015). Health Data Governance: Privacy, Monitoring and Research, OECD Health Policy Studies. *Paris: OECD Publishing.* Retrieved from https://www.oecd.org

Privor-Dumm L., Vasudevan P., Kobayashi K., & Gupta J. (2020). Archetype analysis of older adult immunization decision-making and implementation in 34 countries. *Vaccine*

Raeymaekers P. (2020). Zorgvuldige technologie: Teckno 2030 De rol van technologie voor een betere gezondheidsgerelateerde levenskwaliteit van personen in hun dagelijks leven. *Brussel: Koning Boudewijnstichting.*

Raeymaekers P., Balthazar T., Denier Y. (2020). Big data in de gezondheidszorg. Technische, juridische, ethische en privacy-gerelateerde randvoorwaarden voor (her)gebruik van gezondheidsgegevens voor onderzoek. *Brussel: Zorgnet-Icuro*.

Rossello S., Díaz Morales R., Muñoz-González L.. (2021). Data Protection by design in AI? The case of federated learning. *Computerrecht: tijdschrift voor informatica en recht*

Safran C. (2007). Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *J Am med inform assoc.*

Verdonck P., & Van Hulle M. (2017). Datawetenschappen en gezondheidszorg. Standpunt Koninklijke Vlaamse Academie van België voor Wetenschappen en kunst. *KVAB*. Retrieved from https://kvab.be

Von Grätz P., Moosmann A. (n.d.). Translating Cancer Discoveries into Clinical Practice. *Heidelberg: German Cancer Consortium, German Cancer Research Center.*

Wang S. V., Schneeweiss S. (2021). Assessing and Interpreting Real-World Evidence Studies: Introductory Points for New Reviewers. *Clinical Pharmacology & Therapeutics.*

WHO Regional office for Europe. (2020). The protection of personal data in health information systems - principles and processess for public health. *WHO*. Retrieved from https://apps.who.int

Willems A.F. (2021). Vrij verkeer van data. *Leuven : KU Leuven*

World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. *WHO*. https://apps.who.int

You S. C. (n.d.). Common Data Model Of Everthing in Medicine: Journey for integration of Environmental, Genomic data, Radiology, and Patient-Generated Health Data with clinical data in OMOP-CDM. *OHDSI*.