Journal of Healthcare Informatics Research (2025) 10:179–208
https://doi.org/10.1007/s41666-025-00220-w

RESEARCH ARTICLE

Unlocking Health Data for Research: Legal, Technical, and Organisational Lessons from a Belgian Interdisciplinary Case Study

Audrey Van Scharen · Karen Cruyt · Jeroen Colon · Selene De Sutter · Johnny Duerinck · Ramses Forsyth, et al. ***[full author details at the end of the article]***

Received: 13 August 2025 / Revised: 9 October 2025 / Accepted: 14 October 2025 /
Published online: 23 October 2025

Abstract

The reuse of clinical health data holds immense promise for advancing medical research, yet remains constrained by complex legal, technical, and organisational barriers. This article examines these challenges through the case study of TumorScope, a Belgian interdisciplinary initiative developing a secure, multimodal data environment for glioblastoma research. Drawing on five years of practical experience integrating imaging, genetic, tissue-based, and clinical datasets, the study identifies key legal, ethical, technical, and operational obstacles to effective data access, linkage, and reuse. Technical issues included fragmented data flows, pseudonymisation complexities, and limited interoperability, while legal and ethical barriers arose from strict interpretations of the General Data Protection Regulation, medical secrecy obligations, and intellectual property constraints. These were compounded by operational challenges such as unclear governance structures, resource limitations, and the limited capacity of Medical Research Ethics Committees to assess data-driven research. The analysis further considers the European Health Data Space Regulation (EHDS) as a potential enabler of responsible secondary data use, while noting uncertainties in its national implementation. Overall, the study demonstrates that meaningful health data reuse requires more than regulatory compliance, it depends on robust governance frameworks, institutional coordination, and sustained investment in infrastructure and expertise. The findings contribute to ongoing debates in healthcare informatics on how to translate the vision of the EHDS into practical, ethically grounded data reuse for patient benefit.

Keywords Secondary use of health data · European Health Data Space (EHDS) · Multimodal data integration · Data governance · Interdisciplinary data sharing · Secure processing environment · National implementation

Audrey Van Scharen and Karen Cruyt contributed equally to this work and share first authorship.

Abbreviations

DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
EHDS	European Health Data Space Regulation
EMR	Electronic Medical Record
GDPR	General Data Protection Regulation
HDAB	Health Data Access Body
MREC	Medical Research Ethics Committee
PACS	Picture Archiving and Communication System
SPE	Secure Processing Environment

1 Background

The reuse of clinical data holds tremendous potential for advancing medical research, particularly in the development of personalised treatments, early diagnosis, and improved understanding of disease mechanisms. In oncology, for example, multimodal analysis combining imaging, genetic, pathological, and clinical data can inform tumour classification, guide treatment decisions, and support predictive modelling through artificial intelligence [1, 2]. However, the process of repeatedly collecting, pseudonymising and integrating such data for research is labour-intensive and fragmented.

Despite growing enthusiasm for digital health platforms designed to enable secure and efficient reuse of real-world data, their implementation seemingly faces persistent technical, legal, and ethical barriers [3]. This article explores those challenges through the case of the interdisciplinary research initiative TumorScope. The TumorScope project aims to create a secure environment for sharing multimodal patient data, including MRI and PET-CT images, tissue samples, genomic profiles, and structured and unstructured information from electronic health records. This environment will facilitate data integration across departments such as radiology, pathology, neurosurgery, and medical oncology, enabling collaborative research on glioblastoma while respecting legal and ethical safeguards. However, in the absence of harmonised infrastructure and oversight, linking multimodal data at the individual patient level remains a complex task. In parallel, concerns about compliance with legal frameworks such as, at the European level, the General Data Protection Regulation (GDPR)[4] and the NIS2 Directive[5], as well as Belgian national legislation such as medical secrecy laws, and contractual intellectual property rights contribute to the difficulty in realising this research environment.

At the time of the project's launch in 2020, the proposed European Health Data Space Regulation (EHDS) [6] [7] offered hope for a more coherent framework for secondary use of health data. Now that the final regulation has been adopted, this article revisits those expectations in light of our practical experience. Using the case study, the article provides a detailed account of the types of data involved and the difficulties in accessing, linking, exploiting and sharing them. It then analyses how current legal and institutional frameworks contribute to these challenges and whether the EHDS will provide meaningful solutions. Particular attention is given to

issues such as the role of medical secrecy, data ownership, trust between stakeholders, and the evolving functions of ethics committees and data access bodies. While grounded in our specific case study, these challenges are broadly representative of those faced in interdisciplinary projects which involve secondary use of health data.

The analysis is based on the first five years of the project, up to October 2024, and draws on extensive conversations with all individuals directly involved. As such, this article reflects those collective experiences and the authors' own perspectives, rather than those of their institutions. The goal is to analyse the hurdles encountered and apply the lessons learnt to this project as well as to future research efforts involving secondary use of health data. By combining technical insights with legal analysis, this article contributes to ongoing discussions about how to operationalise responsible and effective health data reuse in practice. It aims to clarify where current obstacles lie, what risks are realistic, and how emerging European and national frameworks can better support data-intensive research for the public good.

2 Analysis

2.1 Types of Data and Ability to Pseudonymise and Anonymise

A clear understanding of the types of data involved in health research is essential, particularly for legal and policy discussions where technical nuances may be overlooked. Too often, assumptions are made that health data can easily be pseudonymised or anonymised, thereby resolving legal and ethical concerns. However, in practice, the complexity and sensitivity of many medical and biomedical datasets, especially in areas like genomics, medical imaging, and human tissues, make effective de-identification challenging, if not impossible. While some data types, such as consent forms and lab results, can be more easily pseudonymised, more complex data types require careful consideration. Drawing on insights from interdisciplinary research involving secondary use of health data, this section clarifies the main data types commonly encountered and explains why each poses specific challenges for ethical and legal compliant data sharing.

The data types originate from multiple samples taken from different parts of aggressive brain tumours (glioblastomas) and the surrounding tissue. As shown in Fig. 1, the process begins with MRI and PET scans of the patient's brain, which are sent to the engineering team. They apply an algorithm to determine areas of distinctive image characteristics within each of the subregions of the tumour. This information is then provided to the surgeon, who selects points of interest that are to be biopsied during the surgery, and enters these locations in the neuronavigation system, a technology that assists in guiding the surgeon during the procedure by using pre-operative images (Fig. 2). Guided by the neuronavigation system, the surgeon selectively collects tissue samples during the operation, while also recording their precise locations with this system. The samples are then sent to specialised labs for detailed analysis of their tissue structure and genomic characteristics. In sum, the data to be shared between the multidisciplinary team of researchers are the

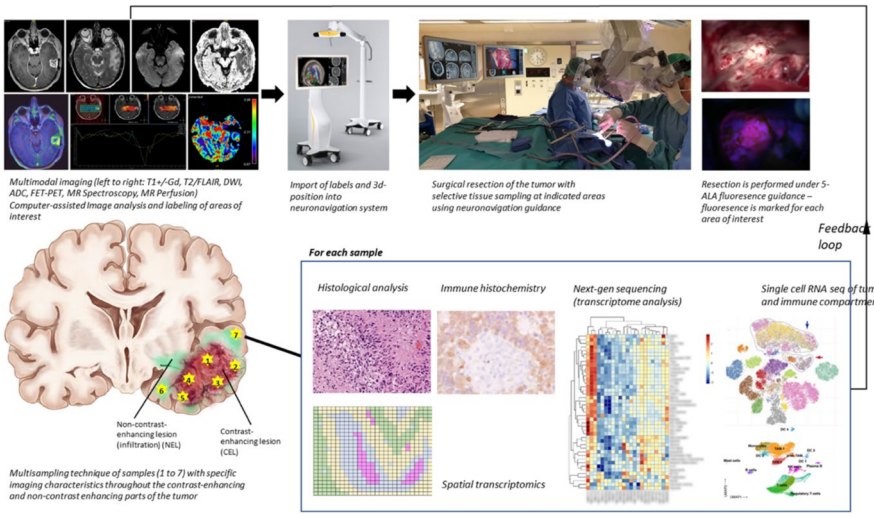

Fig. 1 Flow of steps in surgery for glioblastoma research (copyright: Johnny Duerinck)

intercorrelated imaging, histopathological and molecular data, as well as the clinical data of the patient.

2.1.1 Imaging Data

Building on the surgical workflow outlined above, we now turn to the different types of data generated throughout the clinical process and the steps required to make them accessible and usable for research. As illustrated in Fig. 3, glioblastoma assessment typically involves several imaging modalities. These scan data are retrieved from the hospital's Picture Archiving and Communication System (PACS), which stores various imaging modalities including MRI, CT and PET. A PACS administrator is responsible for extracting the relevant images and applying a pseudonymisation procedure. This process involves removing or encoding identifiable metadata (in the form of DICOM tags), such as patient names, demographic information or hospital IDs, and assigning each image a new study-specific identifier, as well as removing all identifiable data, such as the name of the patient, from the image itself. The key linking these new identifiers to the original patient information remains securely stored within the hospital system, ensuring traceability. The pseudonymised data is then typically securely transferred to researchers in DICOM format, which can then be converted to a format more suitable for processing and analysis, such as NIFTI, retaining only image geometry and voxel data.

Neuroimaging can inadvertently enable facial reconstruction, which could lead to identification [8]. As illustrated in Fig. 4, facial features can be rapidly reconstructed using free and publicly available tools, by excluding the dark background and rendering the volume in 3D. The quality of the reconstruction is strongly related to the resolution of the image in all dimensions and the resulting

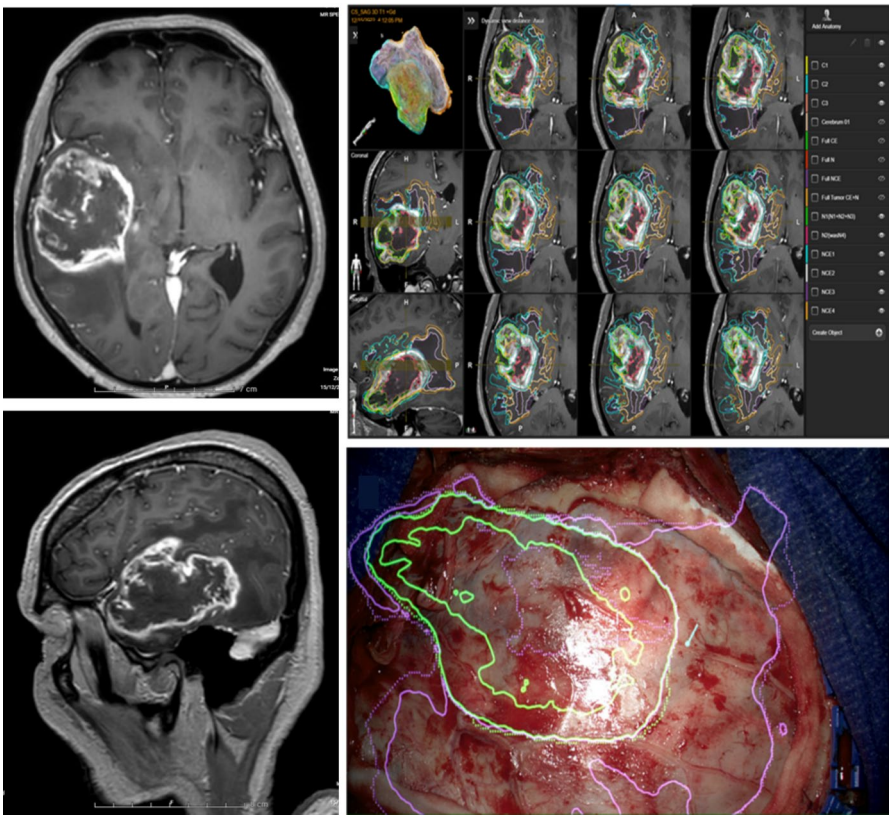

Fig. 2 MRI images (T1 contrast-enhanced sequences) of glioblastoma brain tumour (left), separate regions to be selectively biopsied put into the neuronavigation system (right, top) and overlaid in the microscope view as an augmented reality overlay (right, bottom) (copyright: Johnny Duerinck)

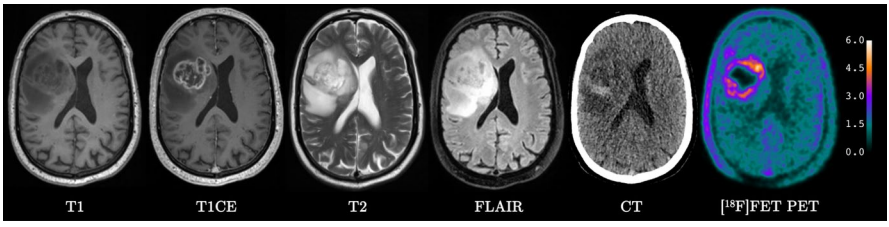

Fig. 3 Imaging modalities for glioblastoma assessment, with the lesion located in the right hemisphere (left side of the image). Modalities include T1-weighted (T1), contrast-enhanced T1-weighted (T1CE), T2-weighted (T2), T2-fluid-attenuated inversion recovery (FLAIR), computed tomography (CT), and O-(2-[18F]fluoroethyl)-L-tyrosine ([18F]FET) positron emission tomography (PET)

amount of slices in the image. In the example shown, the T2 scan, with only 30 slices, produces a coarse reconstruction and does not cover the full face, which hampers identification. To further mitigate the risk of re-identification, common

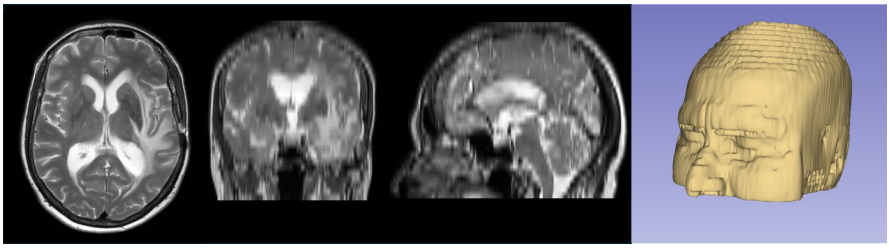

Fig. 4 T2-weighted magnetic resonance image of patient with high resolution (0.5 mm) in axial plane and lower resolution (5 mm) in coronal and sagittal plane (left) and resulting facial reconstruction (right)

preprocessing steps known as facial feature removal or even skull-stripping is applied, the latter removing everything that is not the brain from the image that is used for research [9].

Finally, medical imaging data can contain patient identifiers within the pixel data itself, so-called burned-in text content. Patient names, weight, length or BMI, along with hospital identifiers may appear in reports of CT imaging or captured views from ultrasound acquisitions. Deidentification requires identification of the affected DICOM files using computer vision techniques, and their exclusion or redaction by masking out text areas.

2.1.2 Genomic Data

The biological samples, taken from the tumour biopsy, sent to specialised labs for detailed analysis of their tissue structure and genomic characteristics are the second type of data which bring specific challenges. Genomic data refers to the DNA sequence of an individual, offering insights into biological function and disease. In tumour research, a genomic variant, a change in DNA sequence, enables identification of cancer-specific mutations and supports the development of targeted therapies. Research has shown that as few as 25 genomic variants can be sufficient to identify an individual [10, 11]. As a result, there is broad consensus that genomic data can never be anonymised [12]. The information on the *specific sample* may be pseudonymised, meaning that the patient's identity cannot be derived directly from these data, without a linking code. In research projects, the required data can range from one variant to a targeted subset of genes, or even the whole genome of a person. These genetic data include raw sequencing files (FASTQ format), alignment files (BAM format), and variant files (VCF format). At our institution, these files are stored locally within a secure infrastructure and can only be accessed by users who have received explicit authorisation. Prior to obtaining access, authorised researchers must complete a training session with the bioinformatics team to ensure appropriate data handling. When data are requested for research purposes, pseudonymisation is performed manually: sample identifiers, which are originally based on lab numbers and patient initials, are replaced with neutral, non-informative codes.

2.1.3 Tumour Tissue Data

A third important data source in our research involves human tissues, namely tumour tissue, and their associated microscopic images. All human tissue samples are labelled with pseudonyms to identify the patient they came from. Microscopic images of these tissues are accompanied by a picture of a label, as shown in Fig. 5. Figure 6 further illustrates how pseudonymised identifiers are presented on these labels. Metadata accompanying the image (as shown on Fig. 7) includes a system-generated filename that incorporates both the pseudonym and the precise timestamp of when the image was made, accurate to the second. Interestingly, even the amount of pixels of an image is unique, meaning that an image can always be linked back to the patient with access to the database of all images.

2.1.4 Clinical Patient Data

The fourth and last data type are patient data that are extracted from the electronic medical record (EMR). These records contain a broad range of data, including medication lists, diagnoses, and vital signs. A portion of this structured data

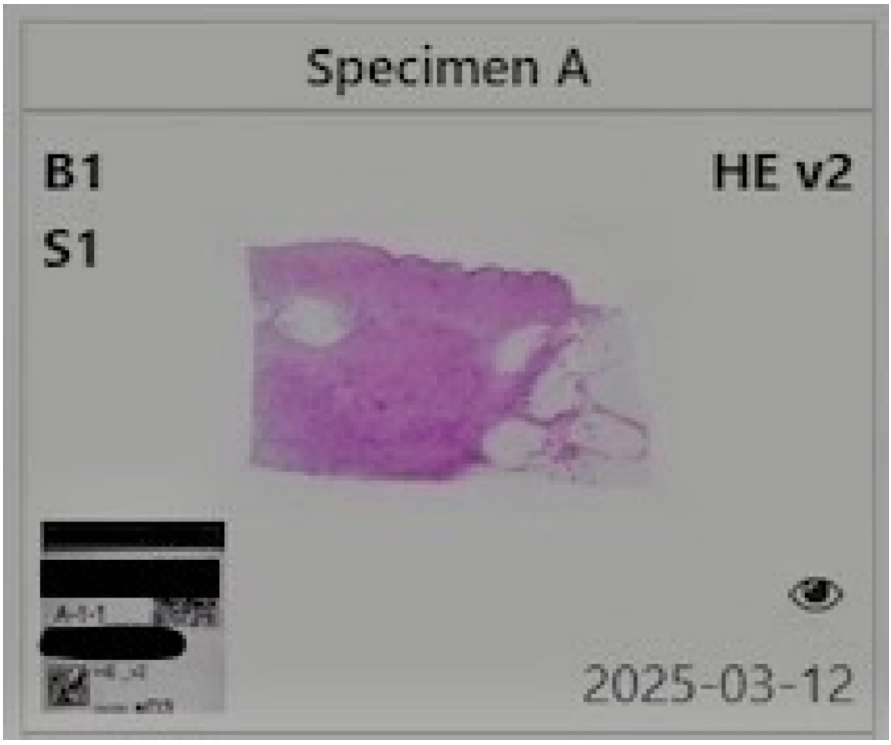

Fig. 5 Researcher view of a digital microscopic image of human tissue, including the label of the specimen with pseudonymised data (blacked out)

Fig. 6 Detail of the label of the specimen with pseudonymised data (blacked out)

is standardised using for example the SNOMED CT classification system, which enhances consistency and interoperability across systems. In addition to structured data, EMRs also contain unstructured information or free text, such as clinical notes, which presents added challenges for extraction and analysis due to variability in language and format. EMR data are typically stored within the hospital information system, a secure platform which is generally managed by the IT department.

This overview shows each type of data follows its own flows, formats, and storage systems, and is managed by different actors responsible for de-identification. This

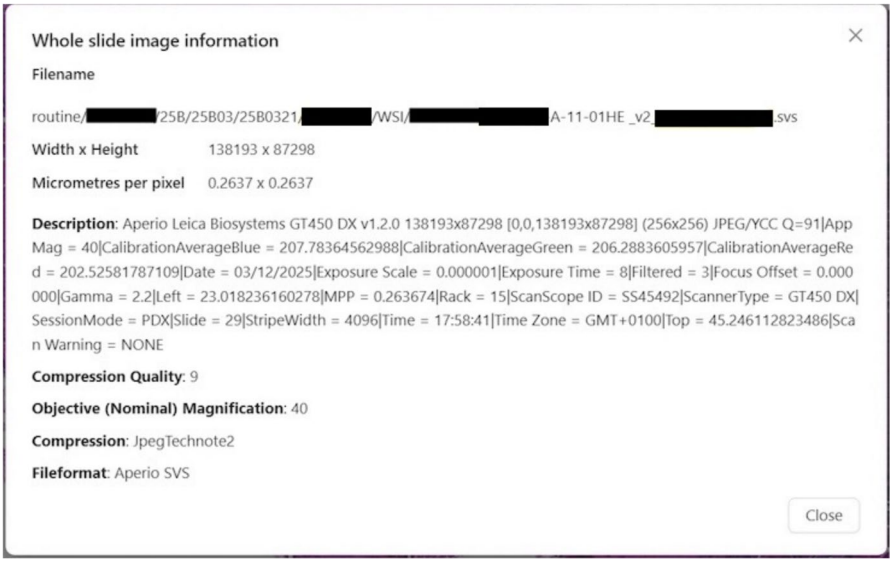

Fig. 7 Metadata linked to the microscopic image, with pseudonymised data in the automatic file name (blacked out)

fragmentation presents a significant challenge when data need to be linked at the individual patient level across modalities. In multimodal research, this is particularly problematic, as such research relies on the combination of genetic data, tumour tissue information, MRI and PET-CT images, and clinical data from the hospital information system.

The following two sections outline the main practical challenges encountered during the early years of the project. The first section focuses on technical obstacles, including issues related to data access, the processing of large multimodal datasets, and the necessary infrastructure. The second section addresses the legal and regulatory constraints faced by the project team.

2.2 Navigating Technical Barriers to Data Sharing

2.2.1 Data Silos

Obtaining real-world health data for research is complex, partially due to the disconnect between clinical and technical perspectives on data availability, location or format. Relevant patient information is typically spread across multiple, separately governed systems. These are for example the EMR for clinical records, PACS for medical imaging, a pathology laboratory information system for microscopic tissue imaging, genetic databases, and investigator-initiated registries. While it is technically possible to link datasets originating from these different silos, this requires both access and coordination between multiple departments. Data engineers depend

on other stakeholders, such as the custodians of radiology and genetic databases, to provide access to the raw data, which must then be manually linked. However, since databases are managed by different departments, responsibilities are fragmented, slowing down data access and integration.

2.2.2 Need for a Shared Pseudonymisation Key or a Harmonised Pseudonymisation Process and Additional Computation

For the intended research of our case study it was necessary to integrate multimodal datasets, including MRI, microscopic images and genetic profiles at the patient level to develop and train a predictive model. This requires that the different types of data can be linked to one another. Such linkage is only feasible if a shared pseudonymisation key or a harmonised pseudonymisation process is applied across all datasets, which implies that the relevant departments collaborate closely. Ideally, this pseudonymisation key can be assigned through a reproducible process. Alternatively, a designated data manager could take responsibility for coordinating and combining the datasets. In that case, this manager would have the authorisation to query all data, and would extract and link the identifiable data and assign consistent pseudonyms. Moreover, the secure environments currently used for storing and operational use of the individual data types either lack sufficient computing capacity for high-performance computing research purposes or do not permit the necessary data manipulations. As a result, all data variables must be combined in a single, local environment that allows for the required computational processing.

However, a challenge for the acquisition and use of patient MRI and genetic data for the development of machine learning tools was the lack of a centralised registry that could support this integration. Since the genetic and MRI databases are managed separately, following a silo approach, it was decided to split the project into two independent, parallel initiatives – one researcher focusing on the genetic mutation prediction task, and a second researcher on the imaging analysis—with the prospect of future harmonisation. The genetic research relied on publicly available databases to obtain information on driver mutations (which contribute to cancer progression) and passenger mutations (which are byproduct of tumour growth) for training a predictive model (D2Deep). Although the model showed strong predictive performance on independent cancer test datasets[13], its accuracy may have benefited from incorporating the in-house data, particularly for passenger mutations, which are more variable and patient-specific – unlike driver mutations, which are often recurrent and shared across tumours from different patients[14]. As such, the TumorScope prototype could serve as a valuable large-scale repository of passenger mutations, enhancing the model's ability to generalise to new cases and support clinical decision-making. Another limitation of relying on publicly available datasets was that they typically included only a single biopsy per patient, linked to an image. This does not accurately reflect the heterogeneity of the whole tumour in gliomas, which can include different molecular and genetic profiles within one tumour. Therefore, linking the genetic profile to an image without localising where the genetic information originated from within the image had limited use. In contrast, access to hospital

data would have made it possible to work with multiple biopsies per tumour, offering a more representative and spatially resolved genetic profile.

2.2.3 Access to Data for Different Purposes and Data Control

An added layer of complexity arises as clinicians requesting data often do so with the front end interface in mind, typically the hospital information system user interface that they work with in daily practice (Fig. 8), while data engineers must work from the back end, where the data are actually stored in various databases.

This mismatch in knowledge creates practical difficulties, especially when requests are ambiguously formulated. For example, if a researcher asks for information on a particular medication, it can be unclear whether this refers to every individual administration, only administrations after a certain date, or a record of dosage patterns. Such ambiguity complicates the task of translating clinical research questions into database queries. Compounding these knowledge hurdles is the technical issue that the requested data are not always available or extractable. Researchers themselves are often unaware of which data exists, where specific data are kept, or whether data exist in a structured format. Free-text fields, for instance, pose serious challenges for data engineers.

Moreover, even when data access is technically feasible, concerns about data control and trust can create further barriers. Data custodians expressed hesitation about transferring data outside their own environments to another platform, which offers sufficient computing power to train models, due to concerns about losing oversight and the potential risk of data misuse. While maintaining in-house control over data may seem preferable, doing so across departments and research contexts presents its own significant technical and organisational challenges. As such, data requests frequently underestimate the complexity of both access and preparation.

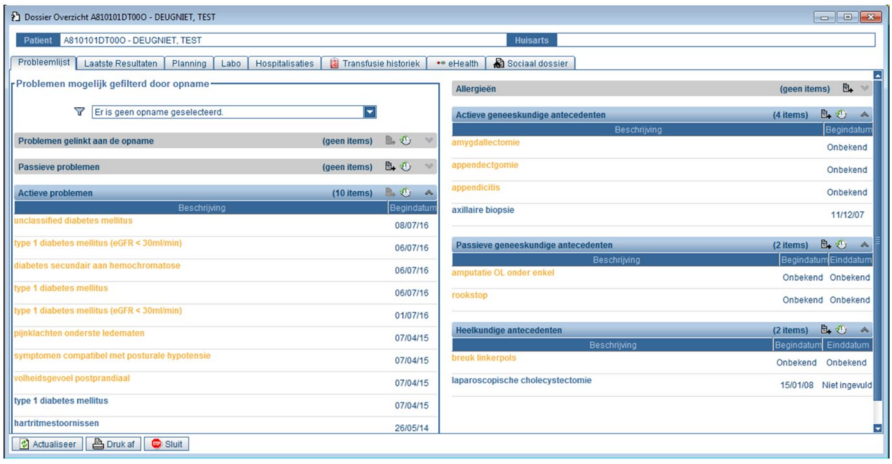

Fig. 8 Image from the user interface of the EMR

This hesitation reflects a deeper tension. Hospitals are rightfully cautious when handling sensitive patient data, particularly considering their legal and ethical obligations. Yet this culture of heightened awareness, which can differ from one department to another, is often inconsistent with the academic expectation of open data for research. Researchers advocate for trust and autonomy but may not fully appreciate the compliance requirements to which clinical institutions are bound. Traditional Medical Research Ethics Committees (MRECs) reviewing such big data research may also lack insight into the technical feasibility of data access and safety measures, sometimes approving projects that cannot realistically be implemented. Adding to this challenge is the significant financial strain faced by hospitals in Belgium: the infrastructure, expertise, and dedicated data teams required to implement secure and compliant data-sharing practices are resource-intensive and costly. These tensions are further reinforced by the legal frameworks that govern data access, which often act as both a necessary safeguard and a practical barrier to data sharing, as discussed in the next section.

2.3 Legal Frameworks in Practice – Guardrails or Roadblocks?

2.3.1 Fear of Non-Compliance, Legal Conservatism and the Duty to Protect Privacy

Many challenges are rooted in a cautious and often risk-averse interpretation of legal frameworks, such as the GDPR. In interdisciplinary research teams, legal interpretation is often undertaken by members with technical, medical, or other non-legal backgrounds, as such teams do not always include legal experts or involve institutional legal departments. As a result, certain legal requirements may be interpreted more conservatively or restrictively than necessary. Moreover, clinical institutions understandably operate under strict obligations to protect patient privacy, and these obligations manifest in layered restrictions on access, storage, and processing of health data. These legal safeguards are essential, yet in practice they can act as significant barriers to research, particularly when interpreted in ways that aim to eliminate all residual risk rather than manage it proportionately. As discussed in the second section of this article, each type of data presents its own specific challenges for de-identification and access, often involving different processes, formats, and responsible actors. Concerns over re-identification[15], particularly in relation to genetic data, which is widely accepted as never truly anonymous, and imaging data like MRI scans, which can potentially be used for facial reconstruction (as shown in Fig. 4), prompt extensive debates over what constitutes effective pseudonymisation. As GDPR Recital 26 points out, identifiability must be assessed by considering all means reasonably likely to be used, considering cost, time, and technological capabilities. Yet in practice, the fear of non-compliance often leads to overly conservative solutions that hinder rather than enable secure data use. Instead of starting with a minimum viable product and incrementally adding features and adjusting safeguards appropriately, an idealised level of risk elimination is sought. One resulting interpretation could be, for instance, that it should be technically impossible to download a dataset or capture it in any form, including through screenshots, even for

authorised users, in-house researchers who are bound by confidentiality obligations and adhere to standards of scientific integrity.

Legal conservatism is also reflected in the approach taken to drafting Data Protection Impact Assessments (DPIAs). This assessment, made legally mandatory by the GDPR in cases of high-risk data processing involving new technologies, should be treated as a living document. In practice, there is often a tendency to resolve every risk before proceeding with data sharing. However, the core principle of the DPIA is not only to mitigate as many risks as can be foreseen, but also to assess whether the remaining risks, those that cannot yet be technically mitigated, are acceptable in light of the intended purpose of the data processing and the reasonable expectations of the data subjects involved.

This legal conservatism is further reinforced by the systems designed to enforce it. Clinical data are typically stored in tightly controlled hospital IT systems governed by strict access protocols, with legal access requiring a therapeutic relationship or, in its absence, the use of pseudonymised data. As discussed in the remainder of this section, these requirements can become particularly complex in research contexts. This tension is echoed in the EHDS, which calls on Health Data Access Bodies to evaluate data access requests. However, the legal, technical and financial resources to interpret and apply regulations such as the GDPR in a timely, research-enabling manner when it comes to secondary use are often lacking. Balancing the public interest in research with the duty to protect privacy remains a complex task in practice [16]. Mechanisms for secure processing, like trusted research environments or secure processing environments (EHDS), aim to provide secure solutions, but they also impose additional constraints on data processing and accessibility – for example, due to their cost of development and implementation. In this landscape, legal requirements, while necessary, can inadvertently delay, reshape, or even derail promising research efforts [12].

2.3.2 Legal Frameworks to Enable Data Sharing, the Issue of Effective Pseudonymisation

Interestingly, despite the seemingly strict and sometimes contradictory interpretations of data protection rules, there is a growing number of legislative initiatives that actively promote the sharing of health data, such as the GDPR, the European Health Data Space Regulation, and the Data Act. Recent case law on key issues such as the distinction between anonymised and pseudonymised data suggests a shift toward interpreting the law in line with this underlying purpose [17, 18]. This trend reinforces the case for a balanced, contextual approach: one that upholds privacy while enabling the kind of research that ultimately serves public health. On the other hand, the European Data Protection Board (EDPB) issued guidelines on pseudonymisation in January 2025 that adopt a more conservative stance [19]. As illustrated by the case Single Resolution Board v European Data Protection Supervisor (SRB v EDPS; General Court)[18] and the case Gesamtverband Autoteile-Handel e.V. v Scania CV AB (Court of Justice)[17], this divergence highlights that even at the highest levels, there remains uncertainty over what constitutes effective pseudonymisation. The courts, supported by the European Commission in the SRB v EDPS case, appear to

favour an interpretation aligned with the spirit of the GDPR, which was conceived not only to protect individuals' rights and privacy, but also to enable the free and lawful movement of personal data across Europe. By contrast, the EDPB and EDPS tend to follow a stricter reading of the letter of the law. As things stand, researchers and institutions must navigate this ambiguity while awaiting further case law and official guidance, bearing in mind that while guidelines of the EDPB and EDPS are influential, only the regulation itself and case law are binding.

Furthermore, it is important to recognise that while pseudonymisation is a valuable tool for mitigating privacy risks, it is not mandatory under the GDPR, nor is it the only acceptable safeguard. The GDPR refers to pseudonymisation as one possible technical and organisational measure among others, depending on the context and nature of the processing. For instance, Article 25(1) on data protection by design and by default, and Article 32(1)(a) on security of processing, both list pseudonymisation as an example of an appropriate safeguard, to be considered in light of factors such as cost, technological capacity, and risk. Similarly, Article 89(1), which governs processing for scientific research purposes, refers to pseudonymisation as an optional safeguard, to be used where compatible with the research objectives. Importantly, Article 6(4)(e) notes that the compatibility of further data processing depends in part on the existence of appropriate safeguards, which may include pseudonymisation. These provisions, taken together, confirm that while pseudonymisation is strongly encouraged, it is not required in all circumstances. Rather, the GDPR supports a risk-based approach that allows for a range of suitable measures, chosen and implemented in proportion to the specific context, purpose, and risk level of the data processing activity. Such measures can include encryption, access control mechanisms, data logging and training for all people who are working with the data.

2.3.3 National Legislation: Medical Secrecy in Belgium and Consent

While data protection legislation such as the GDPR often dominates discussions on health data use, it is not the only relevant legal framework. Particularly in the context of pseudonymisation, it is important to note that in Belgium, the reuse of clinical data is also governed by medical secrecy obligations established under criminal law. This means that only individuals legally bound by medical secrecy, such as healthcare professionals, are allowed access to identifiable patient information. However, in the context of scientific research, the disclosure of identifiable (i.e. not pseudonymised) patients' health data by healthcare professionals is permissible under certain conditions, notwithstanding medical secrecy obligations. In prospective studies, explicit written consent is usually required from participants under applicable laws, such as the Clinical Trials Regulation[20], which may release the healthcare provider from confidentiality duties. Furthermore, the GDPR explicitly recognises consent as a lawful basis for processing personal data, including health data – but it is important to note that this GDPR consent is legally distinct from study participation consent [21, 22]. This aligns with the broader notion of informational self-determination in healthcare, which empowers individuals to decide how their health data may be processed [22–24]. Under the GDPR, valid consent, defined as a freely given, specific, informed,

and unambiguous indication of the data subject's wishes, given through a clear affirmative action, may justify the processing of personal data and can also constitute a lawful ground for breaching medical secrecy [22]. While not legally required, it is advisable to document consent in writing for evidentiary purposes. In retrospective research, no new data are collected, and patients are no longer actively involved, making re-consent impractical, the GDPR may offer a sufficient legal basis for data processing. In that case, when working with large datasets, gathering consent from each patient might be impossible, too time-consuming, or burdensome to the patient. In such case, the GDPR may offer a legal basis other than consent for data processing. This reflects the Regulation's recognition of the substantial societal value of scientific research, including large-scale data initiatives [22, 25]. Scientific research is broadly interpreted under the GDPR and may also rely on alternative legal grounds depending on the specific data processing activity [26]. Regardless of the study type, all GDPR requirements must still be met, though certain obligations, such as some participant information rights, may be temporarily restricted to protect research integrity. Nonetheless, strict requirements, including data security measures, remain in force.

This interpretation is however not universally accepted and remains subject to debate. Experts with a stronger background in medical secrecy than in data protection law often argue that the GDPR and medical secrecy exist in parallel as distinct legal frameworks, and that one cannot simply override the other. From this perspective, the existence of a valid legal basis under the GDPR does not automatically lift the obligation of medical secrecy [27]. This view is also sometimes adopted by medical professionals themselves, who invoke medical secrecy as a reason not to share patient data, even when data protection laws would otherwise permit it. This interpretation also implies that individuals who are not bound by medical secrecy, such as data engineers without a therapeutic relationship with patients, should not have access to identifiable patient records for the purpose of preparing data for research. Yet this presents a practical and legal dilemma. Data engineers have become essential profiles in modern health research: they are typically trained in data science or informatics, and are specifically recruited for their ability to manage, extract, clean, pseudonymise, and curate complex health data in a secure manner. However, they are not covered by medical secrecy by default, nor do they have any therapeutic relationship with patients. This raises the question of how to integrate these new but necessary professional roles within the existing legal and ethical frameworks.

This caution is understandable, as the legal consequences of violating medical secrecy are severe. Under Belgian criminal law, unauthorised disclosure of confidential patient information can be punished with imprisonment of between one and three years, a fine ranging from one hundred to one thousand euros, or both (Article 458 of the Criminal Code). This divergence in interpretation only adds to the legal and operational complexity surrounding health data sharing, creating uncertainty for institutions, researchers and medical professionals trying to navigate overlapping obligations.

2.3.4 Protection of Intellectual Property and Sponsorship Obligations

Another legal challenge that can emerge is that some of the data stored in electronic health records originate from earlier research projects, some sponsored by pharmaceutical companies, and may be governed by contractual clauses that explicitly limit their reuse. These agreements, often designed to protect intellectual property or fulfil sponsor obligations, may conflict with efforts to repurpose data for broader research objectives. As such, medical secrecy and pre-existing contractual obligations add yet another layer of complexity to the already challenging landscape of health data sharing.

2.3.5 Regulatory Activities

In addition to existing data protection, secrecy and confidentiality laws, new regulatory developments are contributing to an atmosphere of uncertainty and caution around data sharing. One example is the NIS2 Directive, which expands the scope of cybersecurity legislation to include research institutions and universities as critical entities, placing them alongside hospitals, which are designated as highly critical. Compared to its predecessor, NIS2 introduces fewer administrative hurdles, streamlining compliance processes in some respects. However, it simultaneously increases institutional responsibility by holding management personally accountable for breaches or non-compliance, and adds additional responsibility and pressure for IT teams [28]. This shift in liability heightens the perceived risks of data sharing, particularly in interdisciplinary research projects that involve multiple partners and sensitive data.

2.3.6 Ethical Requirements

MRECs are formally mandated to assess the ethical acceptability and legal compliance of research protocols. Their role is primarily focused on ethical considerations and does not extend to practical aspects such as data availability or technical feasibility. Nevertheless, in practice, MREC review has become the de facto final approval step for scientific research in many institutions, as it is often the only legal requirement before research can begin. For retrospective studies using existing datasets, this legal requirement does not exist in Belgium. However, even in such cases, ethics approval by an MREC prior to data access is widely accepted and often required by academic journals.

The limitations of this ethics review process become particularly apparent in emerging fields such as real-world health data research. MRECs frequently face structural challenges, including a lack of in-house technical expertise, time constraints, and fragmented or inconsistent interpretations of key legal concepts [29–31]. As a result, projects may obtain ethics approval while critical issues such as data protection, interoperability, or operational feasibility remain insufficiently addressed. These challenges underscore the need for more coordinated and multidisciplinary review processes that integrate ethical, legal, and technical expertise from the earliest stages of research development.

The technical and legal hurdles outlined above underscore the complexity of responsibly reusing clinical data for research. With the recent introduction of the EHDS, which seeks to enable both primary and secondary use of health data and support research and innovation, a key question arises: to what extent does this new regulation address the legal ambiguities and technical barriers that projects like ours have encountered?

2.4 The European Health Data Space Regulation

2.4.1 A New Framework for Primary and Secondary use of Health Data

The EHDS, published in the Official Journal of the European Union on 5 March 2025, was conceived with two main goals in mind. The first goal is to ensure patients have access to and control over their health data, by enabling access in an electronic format that can be used across the European Union. That way, a Belgian citizen who has an accident during their ski trip in Austria can immediately give the Austrian doctor access to their electronic file, so they have all the information to treat them, and in return, their Belgian doctor will have access to the newly created health data related to the accident. In serious cases, this cross-border access can be a matter of life or death. This is what is called the primary use of electronic health data in the regulation. These personal electronic health data include data relating to the health of a person and genetic data, all processed in an electronic form, and the primary use includes using these data for providing medical care, as well as for handling related social, administrative or payment services. Under the EHDS, electronic health data even includes "data determinants of health"[6], such as social or educational factors that can contribute to a person's health status.

The second goal relates to what is called secondary use of electronic health data in this regulation. This is any processing of these data for a purpose which differs from the initial purpose for which they were collected or produced. This can be allowed, but only for those purposes which are mentioned in article 53 of the EHDS, one of which is "scientific research related to health or care sectors that contributes to public health or health technology assessments, or ensures high levels of quality and safety of healthcare, of medicinal products or of medical devices, with the aim of benefiting end-users, such as patients, health professionals and health administrators" (art. 53.1(e) EHDS). As stated in recital 61 of this regulation, scientific research is to be interpreted in a broad manner and includes the training of AI algorithms for the permitted purposes of article 53 EHDS. To access a dataset for one of these permitted purposes, a data user applies for a data permit with the newly established national Health Data Access Body (HDAB), who curates a catalogue of available datasets. The creation of this catalogue depends on the active and mandatory participation of health data holders, including hospitals, research institutions, and digital health service providers. Under the EHDS, these data holders are legally required to make specific categories of electronic health data available for secondary use. To that end, they must provide standardised descriptions of their datasets, including metadata on structure, content, and access conditions. If the permit

is granted, a data user will be able to process anonymised or pseudonymised, as needed, data in a secure processing environment.

2.4.2 Limitations on the use of Health Data in the EHDS: Intellectual Property and Trade Secrets

There are, of course, limitations to the use of these electronic health data. One of them is that a data permit will only be granted if the intended use for the data is for one of the permitted purposes in the regulation. Another limitation relates to intellectual property rights and trade secrets. Article 52 of the EHDS states that health data protected by intellectual property rights, trade secrets, or regulatory data protection must still be made available for secondary use, but health data holders are required to inform the HDAB of such protections, specify which parts of the dataset are affected, and justify the need for protection. The HDAB is then responsible for determining what protective measures are necessary. It can also impose conditions on access, including contractual arrangements between data holders and data users. Access can be denied based on the protection of intellectual property and trade secrets, posing a serious risk to data availability. If a data holder or applicant disagrees with a decision to refuse access, a complaint procedure is available to review the decision. However, the regulation does not appear to provide a similar complaint mechanism for data holders who object to access being granted despite intellectual property or trade secret concerns.

Although the regulation opens the door for reuse of electronic health data by having Health Data Access Bodies decide on the necessity and conditions of their protection, and stating that access will only be refused if it would pose a serious risk for the protection of the intellectual property and trade secrets of the data holder, it does not inspire much confidence that these data will actually be reusable by other data users. The procedural requirements, combined with the possibility of restrictive contractual conditions, suggest that access could remain limited in practice, undermining the EHDS's goal of facilitating secondary data use. Only time will tell how Health Data Access Bodies interpret this article, and especially the notion of "serious risks" that can be the basis for refusal of access.

2.4.3 Medical Secrecy in the EHDS

A second legal obstacle to accessing health data, as previously discussed, is the principle of medical secrecy. While the EHDS does not explicitly include the notion of medical secrecy, recital 24 of the regulation acknowledges the importance of the principle of health professional-patient confidentiality, emphasizing that electronic health data should be handled in a way that respects this ethical principle. Additionally, recital 55 emphasises the need for all health data holders to contribute to making electronic health data available for secondary use, if this occurs through secure processes and with due respect for professional duties, including confidentiality obligations. Article 54(e) of the EHDS adds that secondary use of health data must not involve activities that conflict with ethical provisions laid down in national law. This raises the question of whether medical secrecy could be considered such an ethical

provision. Nevertheless, as discussed above, under the GDPR, valid legal bases can already justify limited and well-regulated breaches of medical secrecy. The EHDS builds upon this by creating an additional framework for lawful secondary use, and it does so within the boundaries of the GDPR, which remains fully applicable and continues to provide safeguards for individual rights and data security. The EHDS could be interpreted as imposing a new, overriding obligation to share data that could supersede stricter national rules such as Belgian medical secrecy legislation, but since this is not explicitly supported by the text of the regulation, room is left for divergent interpretations and potential conflicts in its implementation across Member States, and across institutions and maybe even departments of one institution. This may further complicate access and hinder the full realisation of its secondary use potential.

2.4.4 Data Identifiability and Technical Modalities

To mitigate privacy risks while enabling data sharing, the EHDS introduces specific safeguards regarding data identifiability. Article 44(3) stipulates that electronic health data should be provided in anonymised form wherever possible. Only when the intended purpose of the data user cannot be fulfilled with anonymised data, based on the justification provided by the user, may data be delivered in pseudonymised format. In such cases, the information necessary to reverse pseudonymisation must remain exclusively with the HDAB, and data users are strictly prohibited from attempting re-identification. Violations of these safeguards may result in penalties. This layered approach reflects the EHDS's effort to balance data utility and privacy, but also adds another level of operational complexity, particularly where anonymisation and pseudonymisation are not straightforward, as discussed above.

Lastly, despite its ambition to facilitate health data sharing, the EHDS remains vague on the technical modalities needed to operationalise its goals [32]. The regulation sets out overarching principles for interoperability, security, and access, but stops short of specifying how these should be technically implemented in real-world settings. However, this gap may be addressed through a series of implementing acts that the European Commission is mandated to adopt under the EHDS. These acts, which are expected by March 2027, will set out technical specifications and requirements for key aspects such as data quality standards (Art. 13), cross-border electronic health record exchange formats (Art. 15), and the technical implementation of individual rights (Art. 17), among others. As such, the implementing acts could provide the necessary detail to render the EHDS practically operable and technically actionable across Member States. For now, it remains unclear to researchers what infrastructure, standards, or safeguards for secondary use of health data will be expected to comply with the regulation. At the same time, the absence of overly prescriptive technical detail is understandable: digital infrastructures and data technologies evolve rapidly, and a regulatory framework which is too specific risks becoming obsolete before it is fully applied. Still, without clearer technical direction, the EHDS risks placing the burden of interpretation on national authorities, which could potentially trigger forum shopping when different countries use different standards,

and on local actors, who may overcompensate due to risk-aversion, potentially leading to overcompliance.

3 Discussion

This article set out to explore the legal, technical, and organisational barriers encountered during interdisciplinary research involving secondary use of clinical data, revealing the tension between regulatory compliance and the practical feasibility of data sharing. A key insight, echoing earlier research[33], is that the successful reuse of clinical data depends not only on legal frameworks but also on the technical and organisational setup. In practice, accessing and reusing real-world health data proves far more complex than expected, due to a combination of technical, legal, and institutional challenges, including limited financial resources.

Firstly, the disconnect between researchers' expectations, clinical practice and technical realities remains a major barrier. Researchers typically base data requests on what they see in the hospital information system interface, while data engineers must retrieve data from fragmented back-end systems. This mismatch results in vague or incomplete requests, requiring constant translation between clinical and technical language. Common data quality issues, such as limited availability, inconsistent formats, and reliance on free-text fields, further complicate access and preparation. A potential solution is to improve clinicians' data literacy, which is still under-addressed in medical education. As health care becomes increasingly data-driven, it may be time to revisit the curricula of all professionals involved in generating or using health data [34]. In parallel, continued efforts on the technical side to structure and harmonise data, where feasible, could significantly enhance interoperability and reduce the translation burden between domains.

Secondly, the fragmented nature of key datasets, spread across EMRs, imaging systems, genetic databases, and research registries, requires coordination between departments with unclear or overlapping responsibilities. Even when a research protocol involving multimodal data research is approved by a MREC, the absence of a central registry linking patient-level MRI and genetic data may prevent a project from moving forward. This raises questions about whether traditional ethics review processes remain adequate for data-driven research [29]. In our case, the planned integration had to be abandoned in favour of parallel tracks, resulting in missed opportunities. One proposed solution is the creation of a "data supermarket" that transparently maps available datasets and their conditions of use, making project planning more realistic and reducing the risk of failure. This concept is closely aligned with the EU Datasets Catalogue proposed in Article 57 of the EHDS, which envisions publicly accessible national and European catalogues to help researchers identify and access relevant datasets more efficiently.

Thirdly, pseudonymisation practices often lack consistency and reproducibility, with no clear standards provided by regulations or guidelines. When researchers request access to a dataset, there is frequently uncertainty about who is responsible for performing the pseudonymisation and whether the level of de-identification is sufficient to meet validation and compliance requirements. Whether data will be

anonymised or pseudonymised also depends on the nature of the research and what is necessary to answer the research questions set out in the protocol. Some types of research cannot be conducted on anonymised data, particularly given the high threshold for anonymity under the GDPR, and consequently under the EHDS. Current governance frameworks offer limited clarity, leaving uncertainty about whether researchers can guarantee secure and compliant data handling. While some risk of re-identification is unavoidable, particularly in multimodal datasets, which link different data types for every involved patient, these risks must be assessed and mitigated in advance through a Data Protection Impact Assessment (DPIA). The DPIA, carried out before any data processing begins, helps identify appropriate, proportionate safeguards and ensures that the residual risk is reduced to an acceptable level, with all measures documented as part of the assessment. Moreover, given that academic researchers are bound by scientific codes of conduct, such as the ALEA Code of Conduct[35], a more trust-based approach is warranted.

Finally, it is essential to recognise that making health data accessible involves real and ongoing costs. These include not only staff time but also expenses related to infrastructure, secure access, compute and storage, and quality control processes. Health data is not "free", and meaningful reuse depends on sustained investment and institutional support. This point aligns with recent discussions within the EHDS, which acknowledges the need for financial compensation and infrastructure development to enable compliant secondary use [36].

Even when access to health data is technically feasible and the costs can be carried by a defined research project, it can be delayed or blocked due to distrust in data use, legal uncertainty, and oversight concerns. Projects built around a specific, clearly defined research goal typically benefit from close collaboration with clinicians, manageable datasets, and shared responsibilities. In such cases, the benefits are evident and tend to outweigh the risks.

Looking at the final text of the EHDS, it seems like the legal issues we have touched upon will not be resolved by this regulation, as it barely mentions medical secrecy, and does not change the national, criminal, law in Belgium. Moreover, the regulation remains vague on the issue of intellectual property, leaving significant uncertainty about how proprietary interests will be balanced with the goal of facilitating secondary data use. Depending on how the provisions on intellectual property are interpreted, they could ultimately weaken the intended benefits of data sharing. If data holders choose to share only redacted or heavily restricted datasets to comply with confidentiality and IP concerns, the usefulness of secondary use provisions in health research could be significantly limited [37]. This is particularly problematic given that, in many cases, the data in question were generated using public resources and depend on patient participation. Ethically, there is a strong argument that such data should be reusable for research that benefits society. While it is understandable that companies wish to protect their intellectual property to ensure return on investment, this must be weighed against the collective investment made by hospitals and patients in generating the data in the first place. As highlighted by the European Group on Ethics in Science and New Technologies, research involving human subjects cannot be treated solely as an economic activity subject to market rules but must be guided by principles of solidarity and public good. This is particularly the

case in healthcare, which is to be regulated by fundamental ethical standards rather than commercial imperatives [38]. Without mechanisms that recognise and reconcile these competing contributions, the EHDS may simply reinforce existing barriers under a different legal framework.

Not only will this regulation not resolve some of the issues mentioned, but some aspects of health data processing may also become more restrictive. Under the GDPR, the processing of non-pseudonymised personal data for research purposes is permissible under certain conditions and safeguards, as previously discussed. In contrast, the EHDS proposes a more restrictive framework by permitting only the use of anonymised data as a default, or pseudonymised data when anonymisation is not feasible or appropriate for the specific research purpose. This shift towards more stringent data protection standards poses significant obstacles for working with data types that are inherently difficult, or even impossible, to pseudonymise, such as medical imaging data.

Another challenge is the role of HDABs in implementing these provisions. While transparency in their policies will be essential to ensure a consistent approach, national deviations are likely to emerge, given the discretion left to Member States. A good thing is that, at a minimum, commercial companies will be required to disclose the datasets they hold, but other data holders, such as hospitals, may struggle to comply. Without dedicated funding, some data holders may lack the resources to catalogue and share their data effectively, a time-consuming and complex task typically handled by data teams who are already operating under significant workload pressures, as discussed before. An undesirable side effect could be that hospitals eventually prioritise the monetisation of their data out of financial necessity, while their own researchers find it difficult or impossible to use the data.

While the regulation includes enforcement mechanisms, it remains unclear how they will be applied in practice. For example, imposing monetary fines on underfunded public hospitals for failing to make their datasets available may be possible, but it would be neither effective nor appropriate. These institutions already operate under financial strain, and adding such a burden could further compromise their ability to fulfil their public health mission. The extent to which data holders will collaborate depends largely on the local context, raising doubts about whether the EHDS can truly facilitate large-scale data sharing in practice.

These broader uncertainties around data protection laws, the EHDS, intellectual property, IT security laws and medical secrecy also play out at the local level. In practice, the interpretation and implementation of existing legal frameworks, particularly the GDPR, can sometimes lead to a cautious approach, especially in settings where legal responsibilities are unclear or shared across departments. While these laws are vital for protecting patient rights, their application across multi-stakeholder settings frequently lead to over-compliance. This aligns with previous studies showing that institutions often default to highly restrictive data practices, not solely due to legal obligation but due to a lack of clarity and internal capacity to assess risk in a proportionate way [16]. The result is not so much legal obstruction as inertia: a hesitance to act in the absence of clear, shared frameworks [15]. Confusion over data ownership, lack of confident legal guidance, and differing interpretations between departments mean that what can be

perceived as legal barriers often originates in organisational barriers, with fragmented decision-making when multiple hospital departments are involved. Compounding these challenges, institutions lack the financial resources to develop a proof of concept for a secure processing environment or to invest in the necessary infrastructure to support this type of research.

A striking example of the challenges in balancing data utility and privacy is the treatment of genetic and imaging data, which are routinely considered inherently identifiable, despite technical safeguards such as pseudonymisation. This rigid perception, combined with uncertainty over what constitutes adequate de-identification, can lead to a reluctance to share. This has also been reported in other studies: departments may be unwilling to release data without absolute security guarantees, often due to a perceived responsibility for the data they manage, regardless of the legal or technical feasibility of secure reuse [15, 39, 40]. In practice, these risk-averse attitudes, exacerbated by the absence of practical guidance, reinforces binary thinking around data sharing: either everything is shared, or nothing is. The experiences reported in our case study suggest that research efforts were at times delayed or revised, not necessarily due to legal prohibitions, but rather due to a lack of shared understanding of how data sharing should be technically implemented. Moving forward, proactive steps, such as engaging legal and ethical experts early, mapping data governance roles clearly, and adopting a contextual, risk-based model for de-identification, could reduce friction in future data-sharing initiatives. With the EHDS poised to mandate data sharing, institutions should proactively enable researchers to work with their own data now. Once the EHDS is fully in effect, sharing will be compulsory. Early preparation, by structuring and analysing existing datasets, not only eases the future burden but also allows institutions to extract value from their data today and maintain a strategic head-start once data sharing becomes mandatory.

Indeed, challenges often labelled as "legal" may, in fact, stem from institutional hesitance or a lack of shared frameworks. When no clear pathway exists for cross-departmental data governance, decision-making becomes fragmented. Without proper guidance, it is difficult to blame research teams and medical professionals for erring on the side of caution, as they often feel personally responsible for ‘their’ data and its protection. As others have noted, stakeholder engagement, legal-ethical audits, and ongoing training are essential for building trust and shared responsibility [15, 41, 42]. A risk-based, iterative approach to de-identification, tailored to the specific context, could enable more proportionate decision-making, as opposed to a binary logic of ’share’ versus ’don’t share’ [15, 39]. These findings also echo recent developments under the NIS2 Directive [28].

Moving forward, one key question is how to balance legal safeguards with the need to enable meaningful research. DPIAs, for example, are most effective when used to compare and evaluate specific technical options. Accepting a reasonable level of residual risk, particularly when supported by mitigation measures such as contracts, user training, and access logging, as well as strong governance, can help create a more balanced and functional model of data stewardship [15, 42]. These safeguards are also in line with the broader risk management framework required under NIS2, suggesting that privacy and cybersecurity compliance can be addressed in tandem.

Within the broader question of governance for the use of real-world health data in research, the role of MRECs deserves renewed attention. As mentioned earlier, it is worth asking whether these committees, originally established in the context of clinical research in the early 2000s, are still adequately equipped to assess complex, data-driven projects. In today's research landscape, are HDABs with expertise in information technology security and data protection perhaps better suited to evaluate such proposals? At the same time, the use of patient data in research raises ethical questions that go beyond compliance and data security. It is not only a matter of whether data can be used, but whether it should be used for a particular purpose [43].

There is a strong case to be made for involving both types of review. However, requiring researchers to pass through two separate and often uncoordinated approval processes adds to the already substantial regulatory burden in a sector that should ideally focus on innovation rather than administrative complexity [44]. A more integrated approach is needed. MRECs, information technology security experts and data protection officers should work together in reviewing data-driven research. When all relevant aspects of a proposal are assessed collaboratively by experts in their respective fields, trust in decisions about data access is likely to increase. This would also result in more feasible and realistic research proposals and a more efficient review process overall.

Finally, the question of transparency and public trust cannot be overlooked. While legal and ethical safeguards are essential, it is important to remember why this work matters in the first place: to improve patient outcomes. Several studies show that patients are not inherently opposed to the reuse of their data[45]; on the contrary, conversations with patients in brain tumour studies revealed many expect their data to be used, especially for medical research that serves the public good. What matters most is that they are clearly informed, feel that appropriate safeguards are in place, and have some degree of oversight or control [12, 32, 39]. This idea of contextual acceptability, that willingness to share data depends on who will use it and for what purpose, is a recurring theme. When used in a medical context for clearly defined research, patients are generally supportive, particularly if transparency and accountability are ensured.

4 Conclusion

The ongoing debate around new legislation and its interpretation reveals a divide between schools of thought, ranging from strict protectionist views to more pragmatic, risk-based approaches. Yet amidst this legal complexity, it is important to return to the core purpose of these frameworks: the protection of patients, as well as the ability to (re)use data in the EU. This article highlights the complex interplay of legal, technical, organisational, and financial factors that shape the secondary use of health data in research. While legal and ethical considerations are often foregrounded, our experience shows that the technical and infrastructural demands are equally important. Meaningful data reuse requires more than regulatory permission: it depends on strong internal collaboration, adequate resources, and clear, shared

processes across departments. Equally important is ensuring that medical professionals, who are bound by medical secrecy obligations, have clear and practical guidance on what is permitted regarding data sharing for scientific research, as such clarity is essential to overcoming hesitance and promoting confidence in responsible data sharing.

The reported experiences from our case study highlight the practical challenges of integrating diverse data sources, managing pseudonymisation, and ensuring data quality. These tasks require time, expertise, and sustained support—resources that are often stretched in both healthcare and research environments. However, these challenges are not insurmountable. They point to opportunities for investment in infrastructure, training, and new roles that can bridge clinical, technical, and research perspectives. As institutions prepare for wider data sharing under the EHDS, these capacity-building efforts will be essential.

Crucially, our experience also underlines the importance of internal alignment. Legal, ethical, IT, and clinical stakeholders all play essential roles in enabling responsible data reuse. Where barriers appeared, they often stemmed not from conflict but from a lack of shared language or structured collaboration. By promoting interdisciplinary dialogue and a proactive approach to governance, institutions can turn existing safeguards into enablers of innovation, rather than hurdles to overcome.

Looking ahead, the EHDS provides a framework for broader data access, but its practical impact will depend on how it is implemented. As the EHDS is transposed into national legislation, there is an opportunity to fine-tune the approach. Belgium's relatively generous allowances for research under the GDPR, when compared to some other Member States, offer a hopeful starting point. However, it is essential that the transposition process reflects the *spirit*, and not just the letter, of the law. This means avoiding overly rigid interpretations that could paralyse innovation. Even after the GDPR introduced significant legal hurdles, the legislator has continued to issue new laws that reaffirm the importance of enabling responsible data sharing; for example, in support of research and AI model development. A strict interpretation of data protection principles and national law would likely have prevented such developments.

If researchers and institutions do not proactively engage with their own data now, they risk losing ground as access expands to a wider range of users. At the same time, national implementation must strike a delicate balance between legal safeguards and research needs. Enforcement mechanisms should be effective but not punitive to underfunded institutions essential to public health. The true success of the EHDS in the context of secondary use will be measured not only by the accessibility of data but by whether it effectively fosters meaningful, high-quality research that translates into tangible benefits for patients and society.

Despite the obstacles discussed in this article, we have used this analysis as a foundation to move forward with the project. Our continued belief in the importance and feasibility of this project and this type of interdisciplinary research is reflected not only in the commitment of our growing team but also in the additional funding the project has received. Importantly, the first phase of the project was far from unsuccessful: two PhD candidates successfully completed their work under its umbrella, contributing valuable insights and demonstrating the viability

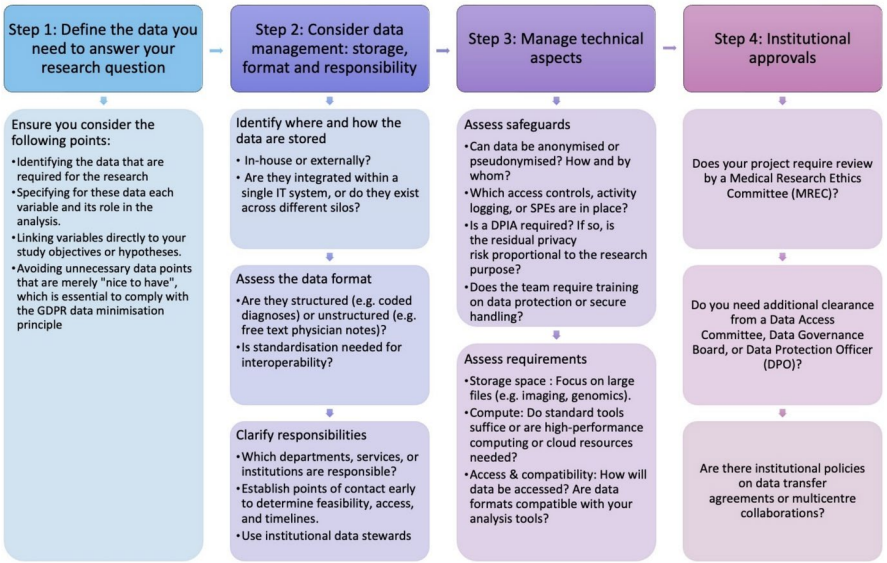

Fig. 9 Proposed preparatory steps for data-driven research involving health data

of this research model. We are convinced that researchers can anticipate and prepare for many of the challenges that arise, even though some obstacles remain beyond their control. Careful preparation, particularly in relation to data-driven research, is essential for moving forward. We therefore propose the workflow outlined in Fig. 9, intended to support researchers in navigating similar projects with greater clarity and confidence.

Ultimately, it is important to remember that these legal frameworks were created not only to protect patients but also to enable research based on health data, and not to obstruct progress. This research is generally supported, and even anticipated, by patients, who in return expect transparency and accountability in its conduct. Legislators continue to affirm the value of data sharing, as seen in recent initiatives that promote its use for purposes that include training AI models. A rigid, overly cautious interpretation of legal principles can risk stalling important work. But with a balanced, forward-looking approach that honours both the letter and the spirit of the law, we can responsibly unlock the potential of health data, to the benefit of science, clinical practice, and, most importantly, patients themselves.

Acknowledgements We want to acknowledge the ICT Department of the UZ Brussels, as members of the broader TumorScope consortium, for their practical guidance and review of the manuscript.

Author Contributions All authors contributed to the research design. The first draft of the manuscript was written by Audrey Van Scharen and Karen Cruyt and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This study was funded by Vrije Universiteit Brussel, under IRP27 "TumorScope: Digital health research on data, AI and legal challenges at the VUB and UZB".

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing Interests The authors declare no competing interests.

References

1. Barzegar Behrooz A, Latifi-Navid H, da Silva Rosa SC, Swiat M, Wiechec E, Vitorino C, Vitorino R, Jamalpoor Z, Ghavami S (2023) Integrating multi-omics analysis for enhanced diagnosis and treatment of glioblastoma: a comprehensive data-driven approach. Cancers 15:3158
2. Fathi Kazerooni A, Saxena S, Toorens E, Tu D, Bashyam V, Akbari H, Mamourian E, Sako C, Koumenis C, Verginadis I (2022) Clinical measures, radiomics, and genomics offer synergistic value in AI-based prediction of overall survival in patients with glioblastoma. Sci Rep 12:8784
3. Teodoro D, Teodoro D, Sundvall E (2018) ORBDA: An openEHR benchmark dataset for performance assessment of electronic health record servers. PLoS ONE. https://doi.org/10.1371/journal.pone.0190028
4. 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). OJ L 119:1–88
5. Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 (December 2022) on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive). OJ L 333:80–152
6. Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 (February 2025) on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847. OJ L 2025(327):1–96
7. European Commission. Directorate General for Communications Networks, Content and Technology (2020) Shaping Europe's digital future. Publications Office, LU
8. Steeg K, Bohrer E, Schäfer SB, Vu VD, Scherberich J, Windfelder AG, Krombach GA (2024) Re-identification of anonymised MRI head images with publicly available software: investigation of the current risk to patient privacy. eClinMed 78:102930
9. Eke D, Aasebø IEJ, Akintoye S (2021) Pseudonymisation of neuroimages and data protection: increasing access to data while retaining scientific utility. Neuroimage: Reports 1:100053
10. GDPR Brief: can genomic data be anonymised?
11. Cai R, Hao Z, Winslett M, Xiao X, Yang Y, Zhang Z, Zhou S (2015) Deterministic identification of specific individuals from GWAS results. Bioinformatics 31(11):1701–1707
12. Rahnasto J (2023) Genetic data are not always personal—disaggregating the identifiability and sensitivity of genetic data. J Law Biosci 10:lsad029
13. Tzavella K, Diaz A, Olsen C, Vranken W (2025) Combining evolution and protein language models for an interpretable cancer driver mutation prediction with D2Deep. Brief Bioinform 26:bbae664

14. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A et al (2020) Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat Genet 52:306–319
15. Thoral P, Thoral PJ, Thoral PJ (2021) Sharing ICU patient data responsibly under the Society of Critical Care Medicine/European Society of Intensive Care Medicine joint data science collaboration: the Amsterdam University Medical Centers Database (AmsterdamUMCdb) example. Crit Care Med. https://doi.org/10.1097/ccm.0000000000004916
16. Quinn P, Ellyne E, Yao C (2024) Will the GDPR restrain health data access bodies under the European Health Data Space (EHDS)? Comput Law Secur Rev 54:105993
17. (2023) Gesamtverband Autoteile-Handel eV v Scania CV AB (C-319/22).
18. (2023) Single Resolution Board v European Data Protection Supervisor (T-557/20).
19. European Data Protection Board (2025) Guidelines 01/2025 on Pseudonymisation.
20. Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 (April 2014) on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC. OJ L 158:1–76
21. Goffin T (2008) Toestemming in het medisch recht. Een nieuwe lezing van een oud probleem. Rechtskd Weekbl 1306–1317
22. Vansweevelt T, Broeckx N (2022) Privacy, persoonsgegevens en beroepsgeheim. In: Handb. Gezondheidsrecht Vol. 2Vansweevelt Thierry Ed. Al. pp 641–826
23. Hooghiemstra T (2018) Informationele zelfbeschikking in de zorg.
24. World Health Organization (2021) The protection of personal data in health information systems-principles and processes for public health. Prot. Pers. Data Health Inf. Syst.-Princ. Process. Public Health
25. (2021) Deontologische aspecten van het gebruik van big data en artificiële…. In: Ordomedic. https://ordomedic.be/nl/adviezen/telematica/telematica/deontologische-aspecten-van-het-gebruik-van-big-data-en-artificiële-intelligentie-voor-biomedisch-onderzoek. Accessed 21 May 2025
26. European Data Protection Board (2019) Opinion 3/2019 concerning the questions and answers on the interplay between the clinical trials regulation (CTR) and the general data protection regulation (GDPR)(art. 70.1. b).
27. Opgenhaffen T, de Koning (2024) Strikt persoonlijk? Het beroepsgeheim in tijden van gegevensbescherming. Tijdschr Psychiatr 8:421–425
28. Vandezande N (2024) Cybersecurity in the EU: how the NIS2-directive stacks up against its predecessor. Comput Law Secur Rev 52:105890
29. Ferretti A, Ienca M, Sheehan M et al (2021) Ethics review of big data research: what should stay and what should be reformed? BMC Med Ethics 22:51
30. Lynch HF, Nicholls S, Meyer MN, Taylor HA, For the Consortium to Advance Effective Research Ethics Oversight (AEREO) (2019) Of Parachutes and Participant Protection: Moving Beyond Quality to Advance Effective Research Ethics Oversight. J Empir Res Hum Res Ethics 14:190–196
31. Dal-Ré R, Morejón E, Ortega R (2004) Nature and extent of changes in the patient's information sheets of international multicentre clinical trials as requested by Spanish research ethics committees. Med Clin (Barc) 123:770–774
32. Raab R, Küderle A, Zakreuskaya A et al (2023) Federated electronic health records for the European Health Data Space. Lancet Digit Health. https://doi.org/10.1016/s2589-7500(23)00156-5
33. Van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, Heymann D, Burke DS (2014) A systematic review of barriers to data sharing in public health. BMC Public Health 14:1144
34. Doll J, Anzalone AJ, Clarke M, Cooper K, Polich A, Siedlik J (2024) A call for a health data-informed workforce among clinicians. JMIR Med Educ 10:e52290–e52290
35. ALLEA - All European Academies (2023) The European Code of Conduct for Research Integrity. ALLEA - All European Academies, DE
36. (2021) Unlocking the full benefits of health data Recommendations from MedTech Europe.
37. McMahon A, Staunton C (2024) Managing access to health data for research and innovation in the EU: is a better regulatory approach possible? In: Confidentiality Priv. Data Prot. Biomed. Routledge, pp 105–126
38. Commission E, Science EG on E in, Technologies N (2003) Opinion Nr 17 on ethical aspects of clinical research in developing countries. Publications Office
39. Rodemund N, Wernly B, Jung C, Cozowicz C, Koköfer A (2023) Striking the balance: privacy protection and data accessibility in critical care research. Intensive Care Med 49:1029–1030

40. Pollard T, Pollard TJ, Pollard TJ et al (2018) The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data 5:180178–180178
41. Rajamäki J, Jarzemski D, Kucera J, Nyman V, Pura I, Virtanen J, Herlevi M, Karlsson L (2024) Implications of GDPR and NIS2 for cyber threat intelligence exchange in hospitals. WSEAS Trans Comput 23:1–11
42. Johnson AEW, Johnson AEW, Pollard T et al (2016) MIMIC-III, a freely accessible critical care database. Sci Data 3:160035–160035
43. Benfatto G, Regulatory Group, Longo L (2021) Regulatory, scientific, and ethical issues arising from institutional activity in one of the 90 Italian Research Ethics Committees. BMC Med Ethics 22:40
44. Nicholls SG (2021) Is the Red Queen Sitting on the Throne?: Current Trends and Future Developments in Human Health Research Regulation. In: Laurie G, Dove E, Ganguli-Mitra A, McMillan C, Postan E, Sethi N, Sorbie A (eds) Camb. Handb. Health Res. Regul., 1st ed. Cambridge University Press, pp 197–204
45. (2020) Algoritmes En Artificiële Intelligentie in Een Medische Context : Een Studie Naar de Perceptie, Mening En Houding van Vlaamse Burgers.

Authors and Affiliations

Audrey Van Scharen[1,2,3,13] · Karen Cruyt[1,3,13] · Jeroen Colon[4,20] · Selene De Sutter[5,19] · Johnny Duerinck[6,7] · Ramses Forsyth[8,9] · Catharina Olsen[10,11,12] · Paul Quinn[1,13] · Konstantina Tzavella[12] · Sonia Van Dooren[3,20] · Wim Waelput[8,9] · Arne Witdouck[20] · Pieter Cornu[3,4] · Jef Vandemeulebroucke[5,14,19] · Wim Vranken[12,15,16,17,18]

✉ Audrey Van Scharen
Audrey.van.scharen@vub.be

[1] Medical Ethics Committee, Vrije Universiteit Brussel (VUB), Universitair Ziekenhuis Brussel (UZ Brussel), Brussels, Belgium

[2] Gerontological Sciences (GERO), Vrije Universiteit Brussel (VUB), Brussels, Belgium

[3] Research Centre for Digital Medicine, Vrije Universiteit Brussel (VUB), Brussels, Belgium

[4] ICT Department, Universitair Ziekenhuis Brussel (UZ Brussel), Brussels, Belgium

[5] Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB), Brussels, Belgium

[6] Department of Neurosurgery, Universitair Ziekenhuis Brussel (UZ Brussel), Brussels, Belgium

[7] Research Group Center For Neurosciences (C4N-NEUR), Vrije Universiteit Brussel (VUB), Brussels, Belgium

[8] Department of Pathology, Universitair Ziekenhuis Brussel (UZ Brussel), Brussels, Belgium

[9] Experimental Pathology Research Group, Vrije Universiteit Brussel (VUB), Brussels, Belgium

[10] Clinical Sciences, Research Group Genetics, Reproduction and Development (GRAD), Vrije Universiteit Brussel (VUB), Universitair Ziekenhuis Brussel (UZ Brussel), Brussels, Belgium

[11] Brussels Interuniversity Genomics High Throughput Core (BRIGHTcore), Université Libre de Bruxelles, Vrije Universiteit Brussel (ULB-VUB), Brussels, Belgium

[12] Interuniversity Institute of Bioinformatics (IB2), Université Libre de Bruxelles, Vrije Universiteit Brussel (ULB-VUB), Brussels, Belgium

[13] Health and Ageing Law Lab (HALL), Research Group Law, Science, Technology & Society (LSTS), Vrije Universiteit Brussel (VUB), Brussels, Belgium

[14] Department of Radiology, Universitair Ziekenhuis Brussel (UZ Brussel), Brussels, Belgium

[15] Structural Biology Brussels, Vrije Universiteit Brussel (VUB), Brussels, Belgium

[16] AI Lab, Vrije Universiteit Brussel (VUB), Brussels, Belgium

[17] Chemistry Department, Vrije Universiteit Brussel (VUB), Brussels, Belgium

[18] Biomedical Sciences, Vrije Universiteit Brussel (VUB), Brussels, Belgium

[19] Imec, Kapeldreef 75, Louvain, Belgium

[20] Health Innovation Hub, Universitair Ziekenhuis Brussel (UZ Brussel), Brussels, Belgium