
Implementation Report

Implementing a Biomedical Data Warehouse From Blueprint to Bedside in a Regional French University Hospital Setting: Unveiling Processes, Overcoming Challenges, and Extracting Clinical Insight

Matilde Karakachoff¹, MSc; Thomas Goronflot¹, MSc; Sandrine Coudol¹, MSc; Delphine Toublant^{1,2}, MSc; Adrien Bazoge^{1,3}, MSc, PhD; Pacôme Constant Dit Beaufile^{1,4}, MD; Emilie Varey^{1,5}, MSc; Christophe Leux⁶, MD, PhD; Nicolas Mauduit⁶, MD; Matthieu Wargny¹, MD, PhD; Pierre-Antoine Gourraud^{1,7}, MPH, PhD

¹Centre d'Investigation Clinique 1413, INSERM, Clinique des données, Pôle Hospitalo-Universitaire 11: Santé Publique, Centre Hospitalier Universitaire Nantes, Nantes Université, Nantes, France

²IT Services, Centre Hospitalier Universitaire Nantes, Nantes Université, Nantes, France

³Unité Mixte de Recherche 6004, Laboratoire des Sciences du Numérique de Nantes, Centre National de Recherche Scientifique, École Centrale Nantes, Nantes Université, Nantes, France

⁴Institut du thorax, Service de neuroradiologie diagnostique et interventionnelle, Centre Hospitalier Universitaire Nantes, Nantes Université, Nantes, France

⁵Direction de la Recherche et de l'Innovation, Centre Hospitalier Universitaire Nantes, Nantes Université, Nantes, France

⁶Service d'information médicale, Centre Hospitalier Universitaire Nantes, Nantes Université, Nantes, France

⁷INSERM Center for Research in Transplantation and Translational Immunology, Nantes Université, Nantes, France

Corresponding Author:

Pierre-Antoine Gourraud, MPH, PhD

Centre d'Investigation Clinique 1413, INSERM, Clinique des données

Pôle Hospitalo-Universitaire 11: Santé Publique, Centre Hospitalier Universitaire Nantes, Nantes Université

30 bd Jean Monnet

Nantes, 44093

France

Phone: 33 2 44768234

Email: pierre-antoine.gourraud@univ-nantes.fr

Abstract

Background: Biomedical data warehouses (BDWs) have become an essential tool to facilitate the reuse of health data for both research and decisional applications. Beyond technical issues, the implementation of BDWs requires strong institutional data governance and operational knowledge of the European and national legal framework for the management of research data access and use.

Objective: In this paper, we describe the compound process of implementation and the contents of a regional university hospital BDW.

Methods: We present the actions and challenges regarding organizational changes, technical architecture, and shared governance that took place to develop the Nantes BDW. We describe the process to access clinical contents, give details about patient data protection, and use examples to illustrate merging clinical insights.

Implementation (Results): More than 68 million textual documents and 543 million pieces of coded information concerning approximately 1.5 million patients admitted to CHUN between 2002 and 2022 can be queried and transformed to be made available to investigators. Since its creation in 2018, 269 projects have benefited from the Nantes BDW. Access to data is organized according to data use and regulatory requirements.

Conclusions: Data use is entirely determined by the scientific question posed. It is the vector of legitimacy of data access for secondary use. Enabling access to a BDW is a game changer for research and all operational situations in need of data. Finally, data governance must prevail over technical issues in institution data strategy vis-à-vis care professionals and patients alike.

JMIR Med Inform 2024;12:e50194; doi: [10.2196/50194](https://doi.org/10.2196/50194)

Keywords: data warehouse; biomedical data warehouse; clinical data repository; electronic health records; data reuse; secondary use; clinical routine data; real-world data; implementation report

Introduction

The increasing use of electronic health records in research settings presents physicians with the systematic yet secondary use of data collected from multiple sources [1-3]. Indeed, hospital information systems (HISs) face a technical challenge to harmonize and integrate application systems and clinical databases that are highly heterogeneous, are based on editor-specific software formats, and use nonstandardized terminologies [3,4].

Moreover, institutions must face the legal and ethical challenge of granting secondary access to data due to national and international laws vis-à-vis patient privacy [5-7]. The reuse of data produced during the care process implies operational knowledge of ethics and legacy concepts that must be solved through well-defined data governance and access policies [7]. Repositories must ensure not only the technical aspects to data access but also the decision criteria granting access [6]. Indeed, institutional data governance is also pivotal when considering legal and ethical principles such as patient informed consent and privacy data protection [2,8,9]. Last but not least, following the line of traditional epidemiology, clinical data reuse requires treatment within a validated and standardized methodological framework to ensure a qualitative result from a scientific and clinical point of view [9].

Despite these difficulties, the rise of biomedical data warehouses (BDWs) is transforming research processes for epidemiology and clinical studies [5,10-12]. Patient data constitute well-defined profiles that can be used to facilitate the enrichment of cohorts [13,14], patient selection and follow-up for clinical trials [15], phenotypic detection, and detailed descriptions of symptoms. BDWs can facilitate the development and performance of personalized and precision medicine, including through the use of big data and artificial intelligence methods.

The implementation of BDWs is conducted at various geographical levels in France. To our knowledge, 24 active hospital BDWs were set up between 2008 and 2023 [16-21]. Regional coordination often occurs within specialized networks of BDWs such as the “Ouest Data Hub,” which is specifically designed for university hospitals in Western France [22]. This can take place in thematic networks and is well advanced in cancer data [23]. National or European Union-wide initiatives also propose a development and coordination framework to deal with the different challenges of implementation. In particular, France initiated in 2019 a national project called “Health Data Hub” [24,25] that promotes centralized coordination and increases the visibility of data sources on a nationwide level.

In this paper we report our 5-year experience regarding organizational changes, technical architecture, and

governance, supporting the implementation of the Nantes University Hospital Biomedical Data Warehouse (NBDW). We describe the process to access clinical content. We also give figures concerning sources and contents included in the repository, and provide some insight into the projects to which the NBDW contributed. Finally, we propose three indicators to measure the effectiveness of the setting up operation.

Methods

Overview

In 2018, Centre Hospitalo-Universitaire de Nantes (CHUN; University Hospital of Nantes) implemented a BDW to facilitate secondary use of personal health data originally collected in the context of patient care for research and to offer single secure access to up-to-date data from different sources within the CHUN HIS, accommodating a wide range of data types, including demographic and clinical information, consultations, billing codes, diagnoses, laboratory results, medical notes, and drug administration in a unified view.

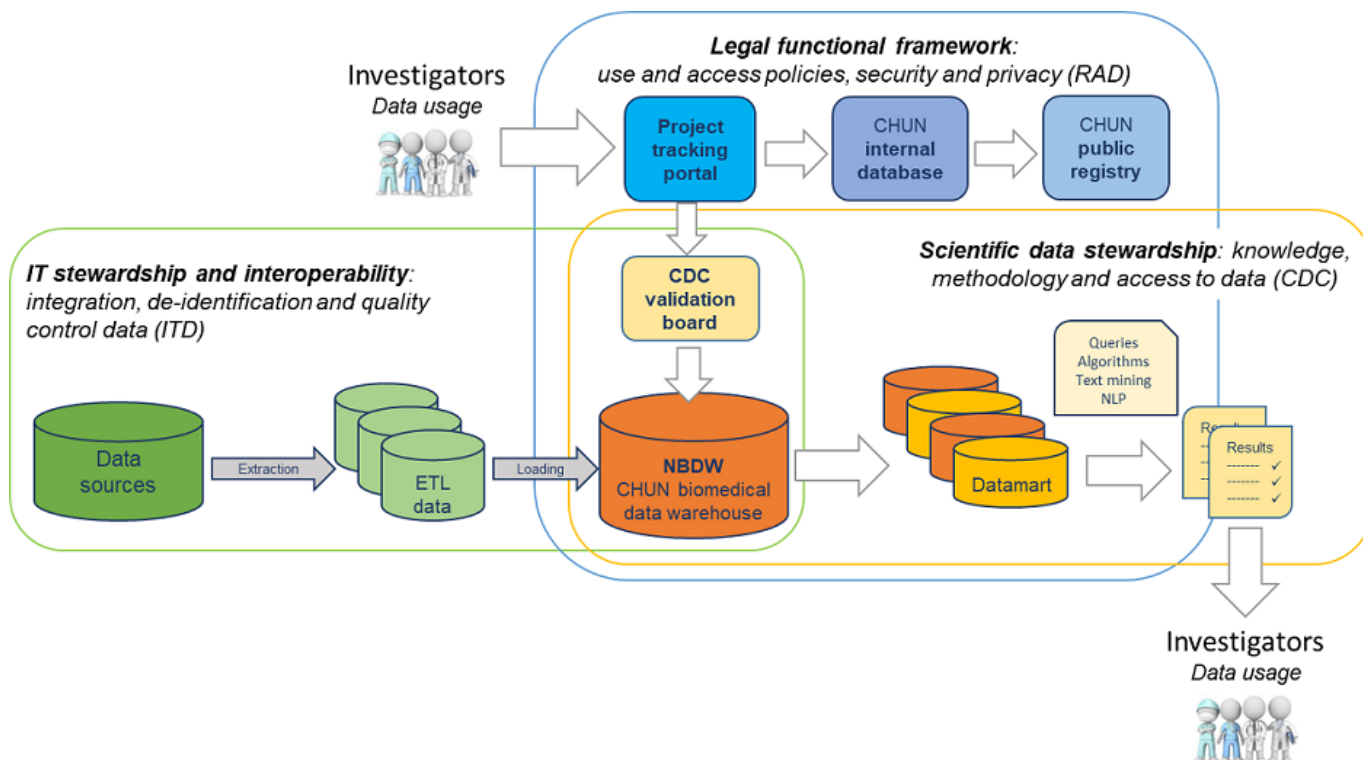
The focal point of the intervention geared toward implementing the NBDW entailed orchestrating the reorganization across the involved hospital research entities. This restructuring initiative encompassed not only the resolution of technical hurdles but also the delineation of governance structures, regulatory frameworks, and parameters governing data access.

This implementation report adheres to the iCHECK-DH (Guidelines and Checklist for the Reporting on Digital Health Implementations) reporting guidelines [26].

From Blueprint to the Functional Organization of the NBDW

Implementation of the NBDW required the de novo organization of a functional research network within a 3-fold structure (Figure 1). First, enforcement of the governance policy and NBDW legal framework was achieved by the Research Administration Department (RAD), which assumed legal and ethical responsibility. Second, the creation of a clinical data center (CDC) took charge of the scientific question and the potential need for NBDW data to coassume scientific responsibility with project investigators. The last layer of technical responsibility was established by the hospital’s Information Technology Department (ITD). IT priorities are defined by legal and governance constraints, and address the data needs of each project. The new organization allows the distribution of tasks and responsibilities in the implementation and management of the NBDW.

Figure 1. A functional framework dedicated to the CHUN NBDW. A 3-fold governance structure constitutes the functional research framework accompanying the development of the NBDW. The ITD ensures the IT infrastructure's quality, transparency, and data deidentification. Through an ETL process, data flows are extracted and loaded to the NBDW. The RAD is an administrative department that assumes legal and ethical responsibility, and lays the regulatory framework that will define data use and access policies. The CDC assumes scientific data stewardship, is in charge of scientific responsibility, supports the methodology of the study, and grants data access. Those three structures are organized and structured to ensure research quality, ethics, and technical transparency. White arrows represent the data request and access process undertaken by investigators. CDC: clinical data center; CHUN: Centre Hospitalo-Universitaire de Nantes; ETL: extract, transform, and load; ITD: Information Technology Department; NBDW: Nantes Biomedical Data Warehouse; NLP: natural language processing; RAD: Research Administration Department.



Legal Functional Framework

On behalf of the hospital, the RAD defines the framework of data use and access policies, security, and patient consent and privacy, and is in charge of the construction, organization, and enforcement of the NBDW regulatory framework. It works in conjunction with the hospital data protection officer, represented by an officer specifically dedicated to research data. Furthermore, the RAD structure requires documented monitoring to ensure a constant adaptation of the NBDW regulatory framework to answer to changes in legal standards. All projects are publicly described on the web [27] for all potential patients contributing data.

IT Stewardship and Interoperability

The ITD is in charge of the collection, storage aggregation, quality, and integrity of clinical data. This structure carries out a continuous process to conform to technical standards. Setting up the addition of new hospital data sources is an ongoing process that started in July 2018 and is still relevant.

The NBDW is set up on virtual machines. An Oracle database enables massive data storage, integrating multiple hospital data sources into a unique set of tables containing data on patients, consultations, diagnoses, laboratory results, medical notes, and inpatient drug administration. The NBDW uses eHOP [22,28] software to organize and query the database. eHOP is a platform developed

by Rennes University using a public-private partnership on the enterprise application integration Enovacom Suite Version 2 (ESV2) of Enovacom (Orange Business Service-Santé). eHOP carries out the acquisition, transformation, and integration of the data coming from various HISs with different formats and standards (Health Level 7 [HL7]; Harmoniser et promouvoir l'informatique médicale [HPRIM]; PN13; Logical Observation Identifiers Names and Codes [LOINC]; and Word documents, PDF, CSV, and text). ESV2 provides an automatic, scheduled (daily, weekly, or monthly), and monitored data supply from the NBDW.

Scientific Data Stewardship and Mediated Access to BDW

Beginning in 2018, the CDC has had a specific team dedicated to the reuse of health data for research, relying on the expertise of public health doctors, data and computer scientists, epidemiologists, biostatisticians, and project managers promoting epidemiology analyses and providing support to clinical investigators and researchers in data access. In conjunction with the ITD, the CDC defines the standards necessary for data integration and data use practices, and ensures data control and the scientific quality of the analyses.

Target and Data Access

The end users of the NBDW are investigators (internal or external to CHUN) aiming to advance the institution's research. A per-project access is created for CHUN investigators through a standardized approval process consisting of 4 main steps (Figure 1). First, investigators register their request and submit a research protocol through a portal hosted by the CHUN intranet. Second, the CDC validation board, which meets once a week, analyzes three dimensions of the submitted research projects: compliance with ethical principles, scientific relevance, and feasibility. On completion of the approval process, the project is registered on an internal database for the completion of legal requirements. Third, the CDC processes all the necessary queries on the NBDW to select a group of patients relevant to the scientific question. During this step, ongoing collaboration with the investigator is necessary, in particular, to precisely define the scientific purpose, eligible population, and data of interest. Fourth, data are made available internally through a data mart, hosted by the CHUN intranet. The data mart system provides regulated, parsimonious (only the required data regarding the targeted population), and time-limited access to investigators. Data are completely deidentified. Moreover, investigators can make simple queries and seek data on patients contained in the data marts through eHOP, which is enriched with a set of tools for simple textual and structured data queries.

Projects requiring data management and extraction to integrate a research database are declared to the public registry of CHUN projects as a guarantee of transparency and to allow patient opposition. At this step, more complex methods for the extraction of information through natural language processing (NLP) [29], regular expression tools, or other structured data [30] may be applied. Finally, data extraction is constrained to strictly necessary data, following the parsimony principle, and only if access to data can be done in a secure environment.

In the case of a project supported by an external project leader from CHUN (academic or private partner), the same process as described above takes place with the exception of the following differences: the project might be supported by a clinical team that submits the research protocol through the portal; a partnership agreement must be signed between the hospital and the partner, and the data mart is only available through a specific virtual working space (data are still internally hosted).

Data Protection and Patient Consent

To comply with national and international privacy regulations, data integration is subjected to a deidentification algorithm. Data are stored in two independent and separate Oracle schemas to separate pseudonymized data from nominative or other directly reidentifying information to which access is strongly limited. Data separation is supplemented by access management and traceability of the actions carried out (ie, AuditLog). Most notably, the platform includes a functionality for collecting and applying patient consent to the use of personal data, ensuring compliance with

French law and European General Data Protection Regulation requirements [31].

Regulatory Approval

In alignment with the French Data Protection Act (Loi Informatique et Libertés, 1978), the use of personal data for health research and evaluation requires compliance with a reference methodology, representing good practices. Without such compliance, personal data use must be authorized by the Commission Nationale de l'Informatique et des Libertés (CNIL; French National Commission for Information Technology and Civil Liberties). At the launch of the NBDW, no research methodology existed for data warehouses in the field. Therefore, approval from the CNIL was mandatory to initiate implementation. Submission to the CNIL covered legal responsibilities, data processing details, access, governance, and more. Comprehensive data access details were provided, extending to researchers whether affiliated with CHUN or not. Private entities are permitted to engage in research projects based on the NBDW, ensuring adherence to this resolution and French regulations. The Data Protection Impact Assessment for NBDW was an integral part of the submission to the CNIL, serving as a mandatory document. The authorization to set up and use the NBDW was granted on July 19, 2018, by the CNIL (resolution 2018-295).

Budget Planning and Sustainability

Estimating the costs associated with implementing and maintaining a data warehouse is challenging owing to several factors. First, the NBDW is part of an institutional strategy, making it difficult to consider it as a stand-alone entity. Second, implementing it involves the collaboration and coordination of multiple structures and experts, complicating the estimation of resource use. Third, hidden costs are difficult to anticipate and consider, including system failures and delays, unplanned license renewals and upgrades, adjustments to regulatory and legal requirements, unplanned changes in HISs, and infrastructure upgrades. We made an estimation by considering three budget lines—infrastructure, license, and human resources—and two different periods—completion in 5 years (2018-2022) and maintenance in 2 years (2022 and 2023)—for a total of €2.6 million (US \$2.8 million).

In terms of sustainability, an annual operational budget is allocated for maintenance and updates. Specific needs for the integration of new hospital data sources are financed through project-based funding. Moreover, an economic model is currently being defined to incorporate additional charges for infrastructure costs in the case of external research projects.

Ethical Considerations

An ethics statement is included in the regulatory approval granted by CNIL with resolution number 2018-295 [32].

Implementation (Results)

Description of the Sources, Concepts, and Contents

The NBDW integrates multiple hospital data sources into a unique and structured set of tables containing data on patient demographic and administrative records, inpatient drug administration, inpatient constants and anthropometric scores and metrics, anatomic pathology notes, inpatient and outpatient medical laboratory results, narrative

medical notes (including admission/discharge summaries, inpatient anesthesia notes, outpatient consultation notes, nurse notes), *International Classification of Diseases, 10th Revision (ICD-10)* and French Classification Commune des Actes Médicaux (CCAM; Common Classification of Medical Procedures) codes for inpatient diagnoses and procedures, and medical imaging reports. [Table 1](#) shows principal concepts and contents according to different HIS data sources or software integrated up to now. Some data sources contain only narrative notes, and some sources contain both unstructured and structured data.

Table 1. Nantes University Hospital Biomedical Data Warehouse principal sources and concepts. Data extracted May 10, 2023.

Concepts	Software	Period	Patients, n	Documents, n	Documents in 2022, n	Structured data, n
Inpatient drug prescriptions	MILLENNIUM	2015-today	318,456	39,681,513	5,673,000	248,289,146
Cardiology narrative notes	CARDIOREPORT	2015-today	9278	38,041	6644	— ^a
Consultation clinical narrative notes	GAM-CLINICOM	2002-today	1,053,386	6,507,860	22,982	—
Constants and anthropometric data	MILLENNIUM	2015-today	440,250	3,306,589	638,969	61,142,757
Anatomic pathology notes	DIAMIC	2015-today	131,812	229,081	27,244	1246
Biology laboratory results	DXLAB	2012-today	701,804	10,752,384	1,672,218	155,825,060
Clinical narrative notes	MILLENNIUM	2015-today	569,114	3,967,073	907,573	5,395,233
<i>ICD-10</i> ^b and clinical procedure codes	CLINICOM	2006-today	725,802	5,318,712	401,607	105,246,620
Radiology reports	QDOC	2015-today	284,937	904,625	122,900	—
Nurse transmissions	TRANSMISSIONS	2017-today	131,508	1,546,114	320,293	1,546,114

^aNot applicable.

^b*ICD-10: International Classification of Diseases, 10th Revision.*

NBDW Figures and Populations

CHUN, a tertiary care hospital ranked seventh in France in terms of activity [33], provides care over a population catchment area of 1.4 million inhabitants. It provides follow-up and long-term health care for both in- and outpatients. It has 2993 hospital beds, delivers 4380 babies, and conducts more than 1 million consultations and external medical procedures per year [34]. It also carries out practical teaching for 1200 medical students, 800 medical residents, and over 2000 non-medical students.

The NBDW includes information on approximately 1.5 million patients admitted between 2003 and 2022 ([Table 2](#)).

More than 1.2 million hospitalizations are associated with approximately 12.3 million *ICD-10*-coded diagnoses and 7.3 million clinical procedure codes. Together with more than 6.3 million external consultations, the NBDW contains more than 11 million textual documents. These narrative notes integrated as free-text documents can be interrogated and turned into structured data for research.

The yearly number of patients, hospitalizations, consultations, and narrative notes has increased over time ([Multimedia Appendix 1](#)) with growth rates between 2003 and 2019 ranging from 109% (hospitalizations) to 430% (outpatient consultations).

Table 2. Nantes University Hospital Biomedical Data Warehouse figures and contents, 2003–2022.

Contents	Since 2003, n	2022 only, n
Patients ^a	1,597,498	300,804
Hospitalizations ^b	2,635,809	183,361
Outpatient consultations	6,358,271	524,948
Clinical narrative notes ^c	11,634,761	826,670
Diagnoses ^d	12,251,148	956,688
Clinical procedures ^e	7,272,346	504,571

^aPatients with ≥ 1 clinical narrative note or a structured document, including inpatient and patients admitted for outward consultations.

^bInpatient hospitalizations in medical, surgical, and obstetric services, including complete hospitalizations, day-hospital admissions, and recurring visits.

^cClinical narrative notes (with the exclusion of vital signs and anthropometric data; *International Classification of Diseases, 10th Revision [ICD-10]* and clinical procedure codes; laboratory results; inpatient drug administrations; and nurse transmissions).

^dMedical diagnoses following the *ICD-10* for medical, surgical, and obstetric hospitalizations.

^eClassification Commune des Actes Médicaux for medical, surgical, and obstetric hospitalizations.

Projects and Effectiveness Outcomes

The availability of NBDW data makes it possible to provide data in response to a wide array of scientific questions and the need for data in the analysis and management of care and organization. Prior to the creation of NBDW data, researchers were limited to the interrogation of medico-administrative-structured information out of the scope of data reuse consent. It only covered structured data such as *ICD-10* diagnoses associated with hospitalizations; medical procedures through CCAM codes; and, to a lesser extent owing to availability issues, laboratory results. Obtaining results from different queries performed independently on different HISs and data manager services was time-consuming if not impossible for both legal and technical reasons.

The NBDW currently facilitates queries of clinical concepts in both structured and unstructured free-text notes in an integrated environment and with respect to data protection policies and laws. BDW data requests may be divided into three different types of research projects according to their purpose: (1) optimize patient screening in both clinical trials and observational studies, (2) enrich case reports or electronic case report forms for disease surveillance, and (3) evaluate and improve clinical practices and resource management. To illustrate different types of studies, three concrete examples of data use are described in [Multimedia Appendix 2 \[35-37\]](#).

The first outcome to measure the effectiveness of the NBDW was defined as the number of studies supported through the project tracking portal ([Figure 1](#)). Since 2018, 577 requests have been made and treated by the CDC. Among them, 269 projects involved patients included thanks to NBDW queries and research tools (second outcome), and for 115 of them, data marts were created to give investigators access to data (third outcome).

Discussion

Lessons Learned

The development of clinical data warehouses has provided unprecedented access to a large amount of diverse data from clinical care. However, it requires a dedicated effort in terms of governance, data access rules with respect to patient consent and data protection, and technical challenges. The reorganization of structures within a functional research framework is the first factor in the success of the NBDW. Collaboration between departments has not only facilitated seamless communication but also engendered the innovation necessary to deal with the complexities of health care data management. It was an opportunity to test new IT

technologies such as distributed infrastructure and anonymization techniques such as deidentification. The interaction between structures has also been a fundamental element in the process of obtaining BDW authorization from the CNIL. Indeed, the obtention of regulatory approval is the result of a long negotiation that required expertise in addressing legal, technical, and scientific research requirements (see Regulatory Approval section for more details).

The creation of the CDC, the result of multidisciplinary teamwork composed of computer scientists, NLP engineers, statisticians, physicians, epidemiologists, and project managers, has probably been the second factor of success. The weekly CDC validation boards, supported by the RAD and ITD structures, verify project compliance vis-à-vis three aspects: ethics, scientific relevance, and practical feasibility, giving support and access to the NBDW in a secure context. However, it is also a leading driver in the shift toward data-driven governance in hospitals.

Implementation of the NBDW has required a continuous and still relevant process to conform to new regulatory and technical standards, and to add new hospital sources and ongoing improvements. An important lesson learned was that each of the 269 NBDW projects in the past 5 years has been an opportunity to revisit the contents of the BDW, further the quality control process, and lead the data transformation process, creating data use value.

Establishing networks and working together is probably the best lesson learned. In France, some experiences have led to changes in health data foresight [28] and promoted the implementation of interregional [22] and national hubs. The NBDW has benefited from and contributed to the “Ouest Data Hub,” a network of Western France university hospital data warehouses. The aim is both to facilitate the reproducibility of data analyses, share resources and best querying practices, and promote adapted and standardized terminologies and nomenclatures between centers. Moreover, BDWs use the same software for both integration and querying, allowing them to be interrogated using consistent queries and rules. This approach ensures a high level of interoperability and accessibility, facilitating seamless interaction and adherence to the Findable, Accessible, Interoperable, Reusable (FAIR) principles.

In hindsight, if this process were to be revisited, there are certain facets that we would address differently. Specifically, in the initial stages of implementing the NBDW, primary emphasis was placed on deploying the software and IT infrastructure recommended by the regional network, ODH, aimed at facilitating the establishment of the repository and its querying system. While acknowledging the benefits

inherent in this strategy, a more thorough examination of alternative IT solutions is warranted to mitigate reliance on a singular approach and to streamline potential transitions to alternative and varied options.

Conclusions

In conclusion, conducting health studies using electronic health records requires careful attention to ensure accurate results owing to a lack of a systematic quality process. The data quality control procedure is a long and necessary process [17,38,39]. The future challenge will be the setting up of standardized and shared quality control pipelines to ensure quality results, not only at the local level but also in a regional and national context of future data sharing. By extending

long-term investments in IT and data in care institutions, the development of NLP and text-mining tools will further accelerate the use of BDW, facilitating data-driven decision-making discussions from top management to patients.

Any research project and analysis of care-based research performed in health management institutions could benefit from the deployment of organized data access. The collaborative nature of data production and the information and privacy protection for patients require mediated and expert access to the BDW. It demonstrates that technical solutions are partial answers to better data-driven practices and must lead to a clear governance strategy

Acknowledgments

The authors wish to thank P Boistard, M Lebigre, C Cartau, A Magnan, P Sudreau, P Lecerf, Dr S Sacher-Huvelin, Dr V Guardiolle, Dr J Esbelin, M Lazarevic, A Royer, and AC de Reboul for their assistance.

This work was financially supported, in part, by the Agence Nationale de la Recherche AIBy4 under contract ANR-20-THIA-0011 and the cluster DELPHI - NEXt under contract ANR-16-IDEX-0007, and integrated into the France 2030 plan by Région Pays de la Loire and Nantes Métropoles.

Conflicts of Interest

PAG is the founder of Methodomics (2008) and the cofounder of Big data Santé (2018). He consults for major pharmaceutical companies and start-ups, all of which are handled through academic pipelines (AstraZeneca, Biogen, Boston Scientific, Cook, Docaposte, Edimark, Ellipses, Elsevier, Methodomics, Merck, Mérieux, Octopize, and Sanofi-Genzyme). PAG is a volunteer board member at the AXA not-for-profit mutual insurance company (2021). He has no prescription activity with either drugs or devices. The other authors declare no potential conflicts of interest to disclose.

Multimedia Appendix 1

Nantes University Hospital Biomedical Data Warehouse yearly data volume by type of data.

[\[DOCX File \(Microsoft Word File\), 132 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Three projects as an example of case experiences based on the Nantes University Hospital Biomedical Data Warehouse.

[\[DOCX File \(Microsoft Word File\), 233 KB-Multimedia Appendix 2\]](#)

Checklist 1

iCHECK-DH (Guidelines and Checklist for the Reporting on Digital Health Implementations).

[\[DOCX File \(Microsoft Word File\), 30 KB-Checklist 1\]](#)

References

1. Meystre SM, Lovis C, Bürkle T, Tognola G, Budrionis A, Lehmann CU. Clinical data reuse or secondary use: current status and potential future progress. *Yearb Med Inform.* Aug 2017;26(1):38-52. [doi: [10.15265/IY-2017-007](https://doi.org/10.15265/IY-2017-007)] [Medline: [28480475](https://pubmed.ncbi.nlm.nih.gov/28480475/)]
2. Safran C, Bloomrosen M, Hammond WE, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association white paper. *J Am Med Inform Assoc.* 2007;14(1):1-9. [doi: [10.1197/jamia.M2273](https://doi.org/10.1197/jamia.M2273)] [Medline: [17077452](https://pubmed.ncbi.nlm.nih.gov/17077452/)]
3. Haarbrandt B, Tute E, Marschollek M. Automated population of an i2b2 clinical data warehouse from an openEHR-based data repository. *J Biomed Inform.* Oct 2016;63:277-294. [doi: [10.1016/j.jbi.2016.08.007](https://doi.org/10.1016/j.jbi.2016.08.007)] [Medline: [27507090](https://pubmed.ncbi.nlm.nih.gov/27507090/)]
4. Gagalova KK, Leon Elizalde MA, Portales-Casamar E, Görges M. What you need to know before implementing a clinical research data warehouse: comparative review of integrated data repositories in health care institutions. *JMIR Form Res.* Aug 27, 2020;4(8):e17687. [doi: [10.2196/17687](https://doi.org/10.2196/17687)] [Medline: [32852280](https://pubmed.ncbi.nlm.nih.gov/32852280/)]
5. Chute CG, Beck SA, Fisk TB, Mohr DN. The enterprise data trust at Mayo Clinic: a semantically integrated warehouse of biomedical data. *J Am Med Inform Assoc.* 2010;17(2):131-135. [doi: [10.1136/jamia.2009.002691](https://doi.org/10.1136/jamia.2009.002691)] [Medline: [20190054](https://pubmed.ncbi.nlm.nih.gov/20190054/)]

6. Pavlenko E, Strech D, Langhof H. Implementation of data access and use procedures in clinical data warehouses. A systematic review of literature and publicly available policies. *BMC Med Inform Decis Mak*. Jul 11, 2020;20(1):157. [doi: [10.1186/s12911-020-01177-z](https://doi.org/10.1186/s12911-020-01177-z)] [Medline: [32652989](https://pubmed.ncbi.nlm.nih.gov/32652989/)]
7. Holmes JH, Elliott TE, Brown JS, et al. Clinical research data warehouse governance for distributed research networks in the USA: a systematic review of the literature. *J Am Med Inform Assoc*. 2014;21(4):730-736. [doi: [10.1136/amiajnl-2013-002370](https://doi.org/10.1136/amiajnl-2013-002370)] [Medline: [24682495](https://pubmed.ncbi.nlm.nih.gov/24682495/)]
8. Bloomrosen M, Detmer D. Advancing the framework: use of health data--a report of a working conference of the American Medical Informatics Association. *J Am Med Inform Assoc*. 2008;15(6):715-722. [doi: [10.1197/jamia.M2905](https://doi.org/10.1197/jamia.M2905)] [Medline: [18755988](https://pubmed.ncbi.nlm.nih.gov/18755988/)]
9. Rosenbaum S. Data governance and stewardship: designing data stewardship entities and advancing data access. *Health Serv Res*. Oct 2010;45(5 Pt 2):1442-1455. [doi: [10.1111/j.1475-6773.2010.01140.x](https://doi.org/10.1111/j.1475-6773.2010.01140.x)] [Medline: [21054365](https://pubmed.ncbi.nlm.nih.gov/21054365/)]
10. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. May 2, 2012;13(6):395-405. [doi: [10.1038/nrg3208](https://doi.org/10.1038/nrg3208)] [Medline: [22549152](https://pubmed.ncbi.nlm.nih.gov/22549152/)]
11. Fennelly O, Cunningham C, Grogan L, et al. Successfully implementing a national electronic health record: a rapid umbrella review. *Int J Med Inform*. Dec 2020;144:104281. [doi: [10.1016/j.ijmedinf.2020.104281](https://doi.org/10.1016/j.ijmedinf.2020.104281)] [Medline: [33017724](https://pubmed.ncbi.nlm.nih.gov/33017724/)]
12. Constant D, Beaufils P, Karakachoff M, Gourraud PA, Bourcier R. Management of unruptured intracranial aneurysms: how real-world evidence can help to lift off barriers. *J Neuroradiol*. Mar 2023;50(2):206-208. [doi: [10.1016/j.neurad.2023.01.156](https://doi.org/10.1016/j.neurad.2023.01.156)] [Medline: [36724868](https://pubmed.ncbi.nlm.nih.gov/36724868/)]
13. Kurian AW, Mitani A, Desai M, et al. Breast cancer treatment across health care systems: linking electronic medical records and state registry data to enable outcomes research. *Cancer*. Jan 1, 2014;120(1):103-111. [doi: [10.1002/cncr.28395](https://doi.org/10.1002/cncr.28395)] [Medline: [24101577](https://pubmed.ncbi.nlm.nih.gov/24101577/)]
14. Greenberg AE, Hays H, Castel AD, et al. Development of a large urban longitudinal HIV clinical cohort using a web-based platform to merge electronically and manually abstracted data from disparate medical record systems: technical challenges and innovative solutions. *J Am Med Inform Assoc*. May 2016;23(3):635-643. [doi: [10.1093/jamia/ocv176](https://doi.org/10.1093/jamia/ocv176)] [Medline: [26721732](https://pubmed.ncbi.nlm.nih.gov/26721732/)]
15. Meystre SM, Heider PM, Kim Y, Aruch DB, Britten CD. Automatic trial eligibility surveillance based on unstructured clinical data. *Int J Med Inform*. Sep 2019;129:13-19. [doi: [10.1016/j.ijmedinf.2019.05.018](https://doi.org/10.1016/j.ijmedinf.2019.05.018)] [Medline: [31445247](https://pubmed.ncbi.nlm.nih.gov/31445247/)]
16. Artemova S, Madiot PE, Caporossi A, PREDIMED group, Mossuz P, Moreau-Gaudry A. PREDIMED: clinical data warehouse of Grenoble Alpes University Hospital. *Stud Health Technol Inform*. Aug 21, 2019;264:1421-1422. [doi: [10.3233/SHTI190464](https://doi.org/10.3233/SHTI190464)] [Medline: [31438161](https://pubmed.ncbi.nlm.nih.gov/31438161/)]
17. Jannot AS, Zapletal E, Avillach P, Mamzer MF, Burgun A, Degoulet P. The Georges Pompidou University Hospital Clinical Data Warehouse: a 8-years follow-up experience. *Int J Med Inform*. Jun 2017;102:21-28. [doi: [10.1016/j.ijmedinf.2017.02.006](https://doi.org/10.1016/j.ijmedinf.2017.02.006)] [Medline: [28495345](https://pubmed.ncbi.nlm.nih.gov/28495345/)]
18. Wack M. Installation d'un entrepôt de données cliniques pour la recherche au CHRU de Nancy: déploiement technique, intégration et gouvernance des données [Doctoral thesis]. Université de Lorraine; Oct 7, 2017. URL: http://docnum.univ-lorraine.fr/public/BUMED_T_2017_WACK_MAXIME.pdf [Accessed 2024-06-17]
19. Pressat-Laffouilhère T, Balayé P, Dahamna B, et al. Evaluation of Doc'EDS: a French semantic search tool to query health documents from a clinical data warehouse. *BMC Med Inform Decis Mak*. Feb 8, 2022;22(1):34. [doi: [10.1186/s12911-022-01762-4](https://doi.org/10.1186/s12911-022-01762-4)] [Medline: [35135538](https://pubmed.ncbi.nlm.nih.gov/35135538/)]
20. Entrepôts de données de santé hospitaliers en France. Haute Autorité de Santé. Nov 17, 2022. URL: https://www.has-sante.fr/jcms/p_3386123/fr/entrepots-de-donnees-de-sante-hospitaliers-en-france [Accessed 2023-01-03]
21. Doutreligne M, Degremont A, Jachiet PA, Lamer A, Tannier X. Good practices for clinical data warehouse implementation: a case study in France. *PLOS Digit Health*. Jul 6, 2023;2(7):e0000298. [doi: [10.1371/journal.pdig.0000298](https://doi.org/10.1371/journal.pdig.0000298)] [Medline: [37410797](https://pubmed.ncbi.nlm.nih.gov/37410797/)]
22. Madec J, Bouzillé G, Riou C, et al. eHOP clinical data warehouse: from a prototype to the creation of an inter-regional clinical data centers network. *Stud Health Technol Inform*. Aug 21, 2019;264:1536-1537. [doi: [10.3233/SHTI190522](https://doi.org/10.3233/SHTI190522)] [Medline: [31438219](https://pubmed.ncbi.nlm.nih.gov/31438219/)]
23. Bocquet F, Raimbourg J, Bigot F, Simmet V, Campone M, Frenel JS. Opportunities and obstacles to the development of health data warehouses in hospitals in France: the recent experience of comprehensive cancer centers. *Int J Environ Res Public Health*. Jan 16, 2023;20(2):1645. [doi: [10.3390/ijerph20021645](https://doi.org/10.3390/ijerph20021645)] [Medline: [36674399](https://pubmed.ncbi.nlm.nih.gov/36674399/)]
24. Cuggia M, Combes S. The French Health Data Hub and the German Medical Informatics initiatives: two national projects to promote data sharing in healthcare. *Yearb Med Inform*. Aug 2019;28(1):195-202. [doi: [10.1055/s-0039-1677917](https://doi.org/10.1055/s-0039-1677917)] [Medline: [31419832](https://pubmed.ncbi.nlm.nih.gov/31419832/)]
25. Goldberg M, Zins M. Health data hub: why and how? *Med Sci (Paris)*. Mar 2021;37(3):271-276. [doi: [10.1051/medsci/2021016](https://doi.org/10.1051/medsci/2021016)] [Medline: [33739275](https://pubmed.ncbi.nlm.nih.gov/33739275/)]

26. Perrin Franck C, Babington-Ashaye A, Dietrich D, et al. iCHECK-DH: Guidelines and Checklist for the Reporting on Digital Health Implementations. *J Med Internet Res*. May 10, 2023;25(1):e46694. [doi: [10.2196/46694](https://doi.org/10.2196/46694)] [Medline: [37163336](https://pubmed.ncbi.nlm.nih.gov/37163336/)]
27. Liste de recherches sur données et/ou échantillons menées au CHU de Nantes, notamment à partir de l'entrepôt. Centre Hospitalier Universitaire de Nantes. URL: <https://www.chu-nantes.fr/liste-des-etudes-menees-au-chu-de-nantes-utilisant-des-donnees-de-l-entrepot-de-recherche> [Accessed 2024-01-26]
28. Delamarre D, Bouzille G, Dalleau K, Courtel D, Cuggia M. Semantic integration of medication data into the EHOP Clinical Data Warehouse. *Stud Health Technol Inform*. 2015;210:702-706. [Medline: [25991243](https://pubmed.ncbi.nlm.nih.gov/25991243/)]
29. Labrak Y, Bazoge A, Dufour R, et al. DrBERT: a robust pre-trained model in French for biomedical and clinical domains. In: Rogers A, Boyd-Graber J, Okazaki N, editors. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics; 2023:16207-16221. [doi: [10.18653/v1/2023.acl-long.896](https://doi.org/10.18653/v1/2023.acl-long.896)]
30. Guardiolle V, Bazoge A, Morin E, et al. Linking biomedical data warehouse records with the national mortality database in France: large-scale matching algorithm. *JMIR Med Inform*. Nov 1, 2022;10(11):e36711. [doi: [10.2196/36711](https://doi.org/10.2196/36711)] [Medline: [36318244](https://pubmed.ncbi.nlm.nih.gov/36318244/)]
31. General Data Protection Regulation (GDPR). URL: <https://gdpr-info.eu/> [Accessed 2023-06-06]
32. Délibération 2018-295 Du 19 Juillet 2018. Légifrance. Oct 23, 2018. URL: <https://www.legifrance.gouv.fr/cnil/id/CNILTEXT000037509951> [Accessed 2024-06-11]
33. Bases statistiques SAE. Données statistiques publiques en santé et social. Jan 6, 2015. URL: https://data.drees.solidarites-sante.gouv.fr/explore/dataset/708_bases-statistiques-sae/ [Accessed 2021-12-28]
34. En chiffres. Centre Hospitalier Universitaire de Nantes. URL: <https://www.chu-nantes.fr/activite-et-chiffres-cles> [Accessed 2023-01-03]
35. Bourcier R, Chatel S, Bourcereau E, et al. Understanding the pathophysiology of intracranial aneurysm: the ICAN Project. *Neurosurgery*. Apr 1, 2017;80(4):621-626. [doi: [10.1093/neuros/nyw135](https://doi.org/10.1093/neuros/nyw135)] [Medline: [28362927](https://pubmed.ncbi.nlm.nih.gov/28362927/)]
36. Cotton F, Kremer S, Hannoun S, Vukusic S, Dousset V, Imaging Working Group of the Observatoire Français de la Sclérose en Plaques. OFSEP, a nationwide cohort of people with multiple sclerosis: consensus minimal MRI protocol. *J Neuroradiol*. Jun 2015;42(3):133-140. [doi: [10.1016/j.neurad.2014.12.001](https://doi.org/10.1016/j.neurad.2014.12.001)] [Medline: [25660217](https://pubmed.ncbi.nlm.nih.gov/25660217/)]
37. Lucas DN, Yentis SM, Kinsella SM, et al. Urgency of caesarean section: a new classification. *J R Soc Med*. Jul 2000;93(7):346-350. [doi: [10.1177/014107680009300703](https://doi.org/10.1177/014107680009300703)] [Medline: [10928020](https://pubmed.ncbi.nlm.nih.gov/10928020/)]
38. Jantzen R, Rance B, Katsahian S, Burgun A, Looten V. The need of an open data quality policy: the case of the "transparency - health" database in the prevention of conflict of interest. *Stud Health Technol Inform*. 2018;247:611-615. [Medline: [29678033](https://pubmed.ncbi.nlm.nih.gov/29678033/)]
39. Looten V, Kong Win Chang L, Neuraz A, et al. What can millions of laboratory test results tell us about the temporal aspect of data quality? Study of data spanning 17 years in a clinical data warehouse. *Comput Methods Programs Biomed*. Nov 2019;181:104825. [doi: [10.1016/j.cmpb.2018.12.030](https://doi.org/10.1016/j.cmpb.2018.12.030)] [Medline: [30612785](https://pubmed.ncbi.nlm.nih.gov/30612785/)]

Abbreviations

- BDW:** biomedical data warehouse
- CCAM:** Classification Commune des Actes Médicaux (English: Common Classification of Medical Procedures)
- CDC:** clinical data center
- CHUN:** Centre Hospitalo-Universitaire de Nantes (English: University Hospital of Nantes)
- CNIL:** Commission Nationale de l'Informatique et des Libertés (English: French National Commission for Information Technology and Civil Liberties)
- ESV2:** Enovacom Suite Version 2
- FAIR:** Findable, Accessible, Interoperable, Reusable
- HIS:** hospital information system
- HL7:** Health Level 7
- HPRIM:** Harmoniser et promouvoir l'informatique médicale
- ICD-10:** *International Classification of Diseases, 10th Revision*
- iCHECK-DH:** Guidelines and Checklist for the Reporting on Digital Health Implementations
- ITD:** Information Technology Department
- LOINC:** Logical Observation Identifiers Names and Codes
- NBDW:** Nantes University Hospital Biomedical Data Warehouse
- NLP:** natural language processing
- RAD:** Research Administration Department

Edited by Caroline Perrin; peer-reviewed by Dianna Reuter, Kerstin Gierend, Olivier Steichen; submitted 22.06.2023; final revised version received 08.04.2024; accepted 17.04.2024; published 24.06.2024

Please cite as:

Karakachoff M, Goronflot T, Coudol S, Toublant D, Bazoge A, Constant Dit Beaufils P, Varey E, Leux C, Mauduit N, Wargny M, Gourraud PA

Implementing a Biomedical Data Warehouse From Blueprint to Bedside in a Regional French University Hospital Setting: Unveiling Processes, Overcoming Challenges, and Extracting Clinical Insight

JMIR Med Inform 2024;12:e50194

URL: <https://medinform.jmir.org/2024/1/e50194>

doi: [10.2196/50194](https://doi.org/10.2196/50194)

© Matilde Karakachoff, Thomas Goronflot, Sandrine Coudol, Delphine Toublant, Adrien Bazoge, Pacôme Constant Dit Beaufils, Emilie Varey, Christophe Leux, Nicolas Mauduit, Matthieu Wargny, Pierre-Antoine Gourraud. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 24.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.