



HAL
open science

The Ouest Data Hub: An Interregional Health Data Sharing Ecosystem for Research

Marc Cuggia, Leslie Grammatico-Guillon, Pierre-antoine Gourraud, Jean-Marie Chrétien, François Bocquet, Valentine Guiton, Éric Stindel

► **To cite this version:**

Marc Cuggia, Leslie Grammatico-Guillon, Pierre-antoine Gourraud, Jean-Marie Chrétien, François Bocquet, et al.. The Ouest Data Hub: An Interregional Health Data Sharing Ecosystem for Research. *Studies in Health Technology and Informatics*, 2024, 316, pp.1679-1683. 10.3233/shti240746 . hal-04694424

HAL Id: hal-04694424

<https://hal.science/hal-04694424v1>

Submitted on 11 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

The Ouest Data Hub: An Interregional Health Data Sharing Ecosystem for Research

Marc CUGGIA^{a,1}, Leslie GUILLON-GRAMMATICO^b,

Pierre-Antoine GOURRAUD^c, Jean-Marie CHRETIEN^d, François BOCQUET^e,
Valentine GUITON^f and Eric STINDEL^g

^a Univ Rennes, CHU Rennes, INSERM, LTSI-UMR 1099, F-35000, Rennes, France

^b Epidemiology Unit EpiDclIC, CDC, Dpt of Public Health, Tours University Hospital, France

^cNantes Université, CHU Nantes, CDC, INSERM, CIC 1413, Nantes, France

^dCHU Angers, CDC, Angers, France

^eData Factory & Analytics Department, ICO, 44805 Nantes-Angers, France

^fGCS HUGO Hôpitaux Universitaire du Grand Ouest, 49000 Angers, France

^gCHU de Brest, CHU Brest, LATIM, UMR 1101, 2 Avenue Foch, 29200, Brest, France

Abstract. The Ouest Data Hub (ODH) a project lead by GCS HUGO which is a cooperation group of University Hospitals in the French Grand Ouest region represents a groundbreaking initiative in this territory, advancing health data sharing and reuse to support research driven by real-world health data. Central to its structure are the Clinical Data Warehouses (CDWs) and Clinical Data Centers (CDCs), essential for analytics and as the linchpin of the ODH's status as an interregional Learning Health System. Aimed at fostering innovation and research, the ODH's collaborative and multi-institutional model effectively utilizes both local and shared resources. Yet, the path is not without challenges, especially in data quality and interoperability, where ongoing harmonization and standard adherence are critical. In 2023, this facilitated access to extensive health data from over 9.3 million patient records, demonstrating the ODH's capacity for both monocentric and multicentric research across various clinical fields, in close collaboration with physicians. The integration of healthcare professionals is crucial, ensuring data's clinical relevance and guiding accurate interpretations. Future expansions of the ODH to new hospitals and data types promise to enhance its model further, already inspiring similar frameworks across France. This scalable model for health data ecosystems showcases the ODH's potential as a foundation for national and supranational data sharing efforts

Keywords. Health data re-use, Clinical Datawarehouse, Data Sharing

1. Introduction

In the era of digital health, the need for seamless health data sharing across institutions has led to the emergence of territorial health data hubs. These hubs serve as pivotal

¹ Corresponding author: Pr Marc Cuggia; E-mail: marc.cuggia@univ-rennes.fr.

platforms for innovation and collaboration at regional, national (e.g., the French HDH[1]), and international scales (e.g., European Health Data Space[2]). Designed to facilitate the aggregation, standardization, and analysis of health data across diverse healthcare settings, these hubs aim to foster innovation and research, support vigilance, and inform healthcare strategy. In this context, the Ouest Data Hub (ODH) platform launched in 2020 by the GCS HUGO which is a cooperation group of University Hospitals in the Grand Ouest region stands out as a significant initiative, leading the advancement towards an interregional Learning Health System (LHS)[3] in France. By enabling effective sharing and utilization of health data, the ODH seeks to bridge the gap between research and healthcare, ultimately fostering innovation for improved healthcare delivery. Clinical Data Warehouses (CDWs), managed by Clinical Data Centers (CDC), play a crucial role in this endeavor, serving as the backbone for the integration and processing of data from electronic health records (EHRs), laboratory information systems, imaging, and other clinical sources. The integration of CDWs into the architecture of the ODH underscores their importance in supporting health data analytics at scale. These infrastructures facilitate the extraction of meaningful insights from clinical data, thereby enabling various use cases such as predictive modeling, population health management, and personalized medicine. This paper explores the implementation of the ODH, focusing on its technological foundations, organizational structure, and governance to facilitate efficient data sharing and utilization in research and innovation.

2. Material and Methods

ODH integrates a comprehensive approach to health data management, underpinned by the deployment of the eHOP CDW, collaborative processes among CDCs, a structured governance framework, and initiatives aimed at fostering knowledge sharing and skills enhancement among data professionals.

Infrastructure Deployment: Central to the ODH's infrastructure is the standard deployment of the eHOP CDWs across all participating healthcare facilities. This foundational step allows data aggregation, harmonization, and interoperability, of different source of data coming from the hospital information system. The CDW is used either for monocentric study or multicentric study. The CDW provides secure data export capabilities by project, enabling the transfer of pseudonymized datasets to the hub, thus facilitating a centralized platform for data analysis in alignment with the hub's research and innovation objectives. Interoperability between business software components and the CDW relies on national and international standards.

Clinical Data Center (CDCs): At the forefront of the ODH are the CDCs, multidisciplinary teams comprising healthcare professionals, data scientists, and data engineers. These teams play a pivotal role in managing and leveraging health data within their respective institutions. They are responsible for ensuring data quality and security, conducting analyses, and utilizing the CDW as a tool for research and innovation, along with acculturation/communication of the physicians to integrate this new approach in their research proposals. The collaboration among CDCs amplifies the capacity for multicentric studies, allowing for the exchange of expertise and the harmonization of data management practices across the network. To enhance the analytical capabilities within the hub, thematic workshops are organized regularly for data scientists and data engineers. These sessions are designed to share best practices, tools, and methodologies

pertinent to data science and engineering, promoting a culture of continuous learning and professional development among data professionals.

Data Sharing Infrastructure: The data sharing infrastructure, inspired by the methodologies and framework of the INSHARE project[4], supports the integration of various data sources for comprehensive health data analysis. Collaborative processes among CDCs, including roles as 'leader' or 'delegates' and conducting feasibility studies, are developed in collaboration with other centers. These processes ensure efficient and secure data exchange, enabling multicentric research studies and innovation projects.

Governance Structure: Adhering to the principle of subsidiarity, the ODH's governance comprises a strategic committee, a scientific and ethical committee, and a one-stop-shop for queries and project submissions. This governance framework is replicated at both local and interregional levels to support multi-institutional projects, ensuring alignment with ethical, regulatory requirements, and strategic goals.

3. Results

In 2024, the ODH incorporated data from five university hospitals and a Cancer Center into its network. The CDW of these institutions feature a variety of health data sources. Although the type of data available may differ from one hospital to another, a comprehensive catalogue outlines the data specifics and the standards guiding the ETL/ELT processes and data exportation to the ODH's data sharing platform. Collectively, the CDWs within the Grand Ouest hold extensive health data from approximately 9.3 million patients, detailing 11 million hospital admissions and encompassing over 365 million clinical documents. In total, there are around 2.5 billion individual data entries. These CDWs are updated continuously, providing a real-time data flow (Table 1). However, the extent of historical data varies across the CDWs depending on the scope and success of past data integration efforts.

Table 1. Data Sources available in the centers

Data Source	Available in n centers	Interoperability	Structured / Unstructured
Patient demographics	6/6	HL7 PAM, OMOP	S
DRG (Diagnosis, Procedure, stay)	6/6	RSS, OMOP	S
Clinical and discharge reports	6/6	HL7 CDA, OMOP	U
Drug prescription-administration	6/6	PN13, OMOP	S
Biology	6/6	HL7H-PRIM, OMOP	S
Radiology reports	6/6	HL7 CDA, OMOP	U
Imaging	1/6	DICOM, DICOM	S/U
Emergency data	1/6	HL7 CDA, OMOP	S/U
Biobank	1/6	Local Standard	S
Intensive care data	1/6	OMOP,OMOP	S/U
Chemotherapy	1/6	PN13/OMOP	S
Medical Device data	2/6	Local Standard	S

The following diagram outlines the overall infrastructure (figure 1). The CDWs are integrated into the information systems of each hospital and adhere to security standards set forth by guidelines published by the CNIL in November 2021[5]. For each project, the CDCs prepare the data on-site by creating datamarts that targets the data of interest within the CDW based on the eligibility criteria of the study protocol. Linkage is ensured through a secure process using the social security number combined with other demographic information, which is hashed. The connection between the CDWs and the

data sharing platform is also secure. On the platform, a workspace is prepared for each project, and preprocessing is carried out upon receipt of data from different centers to check and possibly link data sets. Tools and resources necessary for analysis are made available in this space. Additionally, the platform allows the integration of other data sources from institutions or actors, such as registry data or knowledge bases.

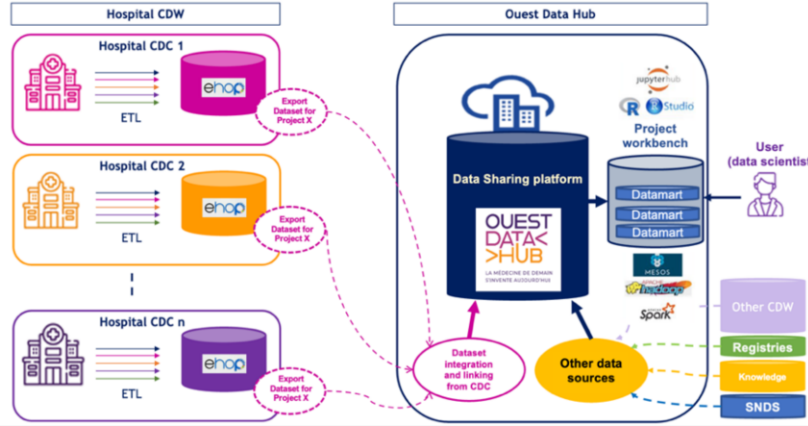


Figure 1. Overall infrastructure

Table 2. Stages involved in managing a study

Stage	Process	Details	Responsible Party
Needs Analysis	Scientific Validation	Review of the request's scientific merit.	ODH Scientific Committee
	Cost Evaluation	Assessment of WP costs for industrial requests.	ODH Project Manager
Methodological Framework	Contracting if industrial	Formal agreements for methodological phase	Legal Department
	Methodology Determination	Establishment of research methodology.	Project Lead
Authorization and Contract	Feasibility Study	Assessment of population, data availability, and quality.	CDC Data Science Team
	Cost Evaluation	Estimation of the project costs.	ODH Project Manager
	Study Documentation	Drafting of essential research documents.	Project lead +/- CDC lead
	Obtaining Authorizations	Securing approvals from relevant bodies.	ODH Scientific & Ethic Comity +/- Cesres Cnil
	Partnership Contracts	Formalization of collaborations with funders and data providers.	Legal Department
Data Preparation	Query Preparation	Development of queries for variable selection by a lead CDC.	Lead CDC
	Utilization of Queries	Adoption of standard queries by partner CDCs.	Partner CDCs
	Project Space Preparation	Preparation of a dedicated environment on the platform.	ODH IT Support Team
	Data Transfer	Secure uploading to the platform.	CDC Data Manager
Project Analysis	Analysis Execution	Conducting the research analysis.	Data Analysts
Valorization and Closure	Publications & Closure	Dissemination through scholarly articles. Formal completion including all regulatory steps.	Lead Researcher & Project Manager

The table above outlines the various stages involved in managing a study, from the initial contact with the one-stop-shop to the valorization of the project. The Table 2

presents the different actors and local resources, i.e., within the CDCs, as well as those pooled at the technical platform level of the ODH. In March 2024, 10 collaborative projects are active on the ODH platform, spanning various clinical domains: pharmacovigilance, genetics, cardiology/endocrinology, hepatology, critical care medicine, materiovigilance and surgery, neurology and internal medicine infectious diseases, pedo-psychiatry and neurodevelopmental troubles. These projects are multicentric studies necessitating the use of datasets from several data warehouses. In parallel, a significant number of monocentric projects have been executed locally.

4. Discussion and Conclusions

The ODH has established itself as a critical platform for data sharing between data providers and users, supporting large-scale, data-driven research initiatives. Its success is built on a governance structure that respects the subsidiarity principle, fostering trust among institutions, project leads, and patients while balancing autonomy and network-wide coordination. While the Hub has made significant strides, challenges with data quality and interoperability persist. Adherence to international standards is uneven across source software systems, requiring ongoing efforts to harmonize data and terminology usage. Project-specific harmonization, such as with LOINC for biological data, has yielded reusable mappings that enhance data integration and facilitate cross-project research. This Hub model deeply involves physicians and other healthcare professionals, harnessing their expertise to optimize the use of real-world data and mitigate biases in analysis. This clinical input is crucial for the pertinence and applicability of research findings, along with highly relevant questions unsolved by classical experimental approaches due for instance to the lack or the latency of events. Future growth includes integrating four additional hospitals into the network, allowing a more comprehensive territorial coverage and the assimilation of new data types. This expansion and the Hub's replicable organizational model across France signal a shift toward more interconnected clinical data management systems, paving the way for national and even supranational data hubs. Beyond the data sharing paradigm, federated analysis, which allows calculations to be performed on aggregated data without transferring it out of the institution, is a promising approach to improving data security. The ODH is currently exploring this approach notably through the European project INDICATE. The ODH's framework demonstrates the potential for scalable and adaptable health data ecosystems.

References

- [1] 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.* août 2019;28(1):195-202.
- [2] The European Health Data Space (EHDS) [Internet]. [cited 2024 Jun 7]. Available from: <https://www.european-health-data-space.com/>
- [3] Friedman CP, Lomotan EA, Richardson JE, Ridgeway JL. Socio-technical infrastructure for a learning health system. *Learn Health Syst.* janv 2024;8(1):e10405.
- [4] Bouzillé G, Westerlynck R, Defossez G, Bouslimi D, Bayat S, Riou C, et al. Sharing Health Big Data for Research - A Design by Use Cases: The INSHARE Platform Approach. *Stud Health Technol Inform.* 2017;245:303-7.
- [5] La CNIL adopte un référentiel sur les entrepôts de données de santé [Internet]. [cited 2024 Jun 7]. Available from: <https://cnil.fr/fr/la-cnil-adopte-un-referentiel-sur-les-entrepots-de-donnees-de-sante>