

Implementation Report

Building a National Interoperable Rare Eye Disease Data Warehouse: Methodological Framework and Implementation Report From the French Rare Eye Disease Database (FREDD) Initiative

Camille Beluffi Marin¹, PhD; Marilyne Oswald², PhD; Laura Ratenet², MSci; Matthieu Stoll², MSci; Kirsley Chennen³, PhD; Nathalie Chabin⁴, MSci; Julie Hubert⁴, MSci; Grégoire Rey⁴, PhD; Guylène Le Meur⁵, MD; Vasily Smirnov⁶, MD; Isabelle Meunier⁷, Prof Dr Med; Isabelle Audo⁸, Prof Dr Med; Hélène Dollfus⁹, Prof Dr Med

¹Laboratoire de Génétique Médicale, Strasbourg, Grand Est, France

²F5MR SENSEGENE, Centre de Référence Pour les Affections Rares en Génétique Ophtalmologique (CRM R CARGO), Hôpitaux Universitaires de Strasbourg, Strasbourg, Grand Est, France

³Laboratoire de Génétique Médicale, Strasbourg, Grand Est, France

⁴France Cohortes, Inserm, Ined, Sorbonne Université, Université Paris Cité, Université Paris Saclay, Paris, France

⁵Nantes University, ophthalmology department Nantes University Hospital, Inserm, TARGET, Nantes, Pays de la Loire, France

⁶Exploration de la Vision et Neuro-Ophtalmologie, CHU de Lille, Lille, Hauts-de-France, France

⁷National Reference Centre for Inherited Sensory Diseases, University of Montpellier, Montpellier, Occitanie, France

⁸Centre Hospitalier National d'Ophtalmologie des Quinze-Vingts, Centre de Référence Maladies Rares REFERET and INSERM-DGOS CIC 1423, Paris, Île-de-France, France

⁹F5MR SENSEGENE, Centre de Référence Pour les Affections Rares en Génétique Ophtalmologique (CRM R CARGO), Institut de Génétique Médicale d'Alsace (IGMA), ERN-EYE, Hôpitaux Universitaires de Strasbourg, Strasbourg, France

Corresponding Author:

Camille Beluffi Marin, PhD
Laboratoire de Génétique Médicale
1 rue Eugène Boeckel
Strasbourg, Grand Est 67000
France
Phone: 33 368853065
Email: beluffi@unistra.fr

Abstract

Background: Rare eye diseases are characterized by low prevalence, clinical heterogeneity, and fragmented data collection, which limit the reliability of analysis results and multicenter research. In France, the development of health data warehouses is strictly regulated by national data protection authorities. While international initiatives aim to harmonize rare disease registries, capturing hyperspecialized, multimodal clinical records and ophthalmology-specific imaging within a fully compliant and sustainable infrastructure remains a major operational challenge.

Objective: This study aims to describe the methodological framework, regulatory implementation, and early operational outcomes of FREDD (French Rare Eye Disease Database), a national interoperable health data warehouse dedicated to rare eye diseases, and to analyze its key success factors, technical bottlenecks, and long-term financial sustainability models.

Methods: Developed under the oversight of the French National Institute of Health and Medical Research (Institut national de la santé et de la recherche médicale; INSERM), the warehouse implements a centralized 3-layer architecture encompassing data collection, processing, and research reuse environments. The dataset extends the national minimum rare disease dataset with detailed ophthalmology variables structured through a dynamic electronic case report form, as well as a dedicated collection of ophthalmic images. Interoperability is achieved using international domain ontologies and a custom parsing tool (FREDDEx) designed to automatically prefill clinical data from the existing French national rare disease registry. Centralized monitoring, data curation, and cross-center quality tracking are driven by a dedicated in-house dashboard (FREDDIn [FREDD Insights]), while unstructured retinal images are processed through a standardized pseudonymization and human-verified validation pipeline.

Implementation (Results): Official regulatory authorization was obtained in April 2024, and active data collection began in March 2025 across 5 pilot expert centers, focusing its initial phase on inherited retinal dystrophies, primarily retinitis pigmentosa. Over a 1-year period, the warehouse successfully integrated clinical data from 1649 patients—representing 20.3% of all retinitis pigmentosa cases registered nationally—and accumulated more than 10,000 ophthalmic images. The centralized dashboard maintained an overall data inconsistency rate of approximately 7%, chiefly reflecting logical dependencies. Among the included patients, 36.7% (605/1649) patients had associated imaging data, with fundus photography being the most widely available modality (768/1649, 46.6%).

Conclusions: The implementation of this database demonstrates that a highly specialized, multimodal health data warehouse can be successfully deployed in a highly regulated environment when regulatory anticipation, governance formalization, and technical flexibility are addressed in parallel. While initial development was supported by institutional grants, long-term sustainability will require a structured cost-recovery model. Future milestones will focus on expanding center coverage nationwide, refining image annotation frameworks for AI readiness, and deploying automated export pathways toward European registries.

JMIR Med Inform 2026;14:e99206; doi: [10.2196/99206](https://doi.org/10.2196/99206)

Keywords: rare eye diseases; retinitis pigmentosa; health data warehouse; secondary use of health data; clinical research; data interoperability; data governance; regulatory bodies; inherited retinal dystrophies

Introduction

Context

The French Rare Disease Ecosystem

Over the past 2 decades, successive French national plans for rare diseases (PNMRs) have structured national rare disease care and research. During PNMR2 (2011-2016) [1], the Filières de Santé Maladies Rares (FSMR) network was established to coordinate expert centers, while the National Registry for Rare Diseases (Banque Nationale de Données Maladies Rares [BNDMR]) centralized standardized clinical data through BaMaRa (Base maladies rares), a national data collection system supporting both clinical management and secondary data use [2-4]. PNMR3 (2018-2022) [5] further strengthened structured data collection by integrating BNDMR into the Health Data Hub (HDH) [6] to facilitate secure linkage with National Health Insurance data. Within this framework, the SENSGENE FSMR [7] coordinates 91 expert centers dedicated to sensory diseases. PNMR4 (2025-2030, “From Territories to Europe”) [8] now aims to further develop health data resources, biobanking, and international interoperability.

This ecosystem operates under the oversight of the French data protection authority (CNIL, Commission nationale de l’informatique et des libertés [9]) and complies with the French Data Protection Act (Loi Informatique et Libertés [10]) and the European General Data Protection Regulation (GDPR [11]).

European Integration: ERN-EYE and REDgistry

At the European level, rare disease expertise is coordinated through European Reference Networks (ERNs). For rare eye diseases (REDs), ERN-EYE [12] connects over 60 specialized centers across 24 member states to facilitate cross-border expertise and collaborative research. The REDgistry [13], launched in 2022, provides a European registry framework to harmonize RED data collection and enable multicentric analyses. These initiatives align with the emerging European

Health Data Space (EHDS [14]), which aims to support secure and interoperable reuse of health data across the European Union.

The Digital Health and AI Opportunity in Rare Diseases

National and European platforms, such as the BNDMR and REDgistry, have become pivotal for large-scale, data-driven multicenter research, cohort building, and AI applications [15-22]. However, while broader initiatives such as RD-Connect [23], BBMRI-ERIC [24], EPIRARE [25], and the ERNs have established foundational frameworks for data standardization and cross-border sharing, they typically operate at a generic rare disease level or focus primarily on biological sample governance and federated access.

Identified Gap

Despite this flourishing national and European context, a critical limitation remains. While the BNDMR collects data across all rare diseases, it is primarily designed for care coordination and epidemiological monitoring, thus lacking ophthalmology-specific details such as imaging data or visual function measurements. Conversely, REDgistry aggregates data at the European level but cannot capture the full spectrum of national patient cohorts, potentially overlooking local research priorities and multicenter project feasibility. To address these limitations, comprehensive national rare disease registries should be integrated with disease-specific European infrastructures, creating a continuum from routine clinical data collection to highly specialized ophthalmic research while maintaining international interoperability.

Aims and Objectives

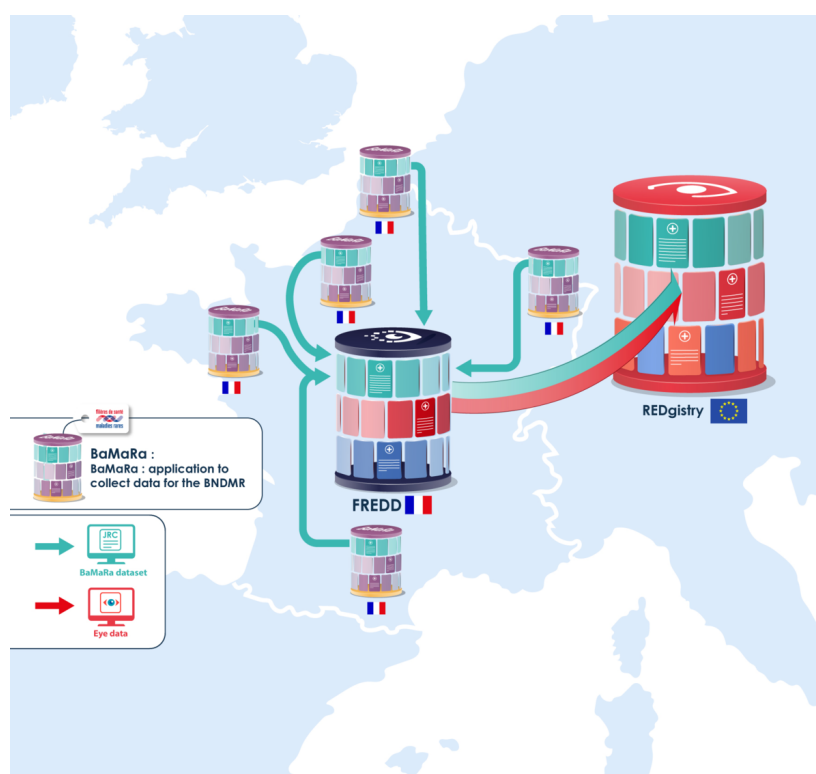
To address this gap, clinicians from the SENSGENE network created FREDD (French Rare Eye Disease Database) [26], a dedicated health data warehouse (HDW) designed to centralize, structure, and enrich clinical data from RED patients across multiple expert centers. FREDD is intentionally designed to follow, as closely as possible, existing and widely adopted principles of rare disease data infrastructures

in terms of technical architecture, interoperability standards, and data governance models to ensure alignment with established national and European ecosystems. FREDD aims to provide a nationally coordinated repository for RED data; enable national multicenter research, cohort building, and longitudinal studies; facilitate interoperability with the French BNDMR, ERN-EYE, and the emerging EHDS; and support

the development of new diagnostic and therapeutic strategies tailored to RED.

This multiscale articulation—from local clinical data collection to national structuring and European integration—is illustrated in Figure 1.

Figure 1. The rare eye disease ecosystem at the national and international levels. BaMaRa (Base maladies rares) and the BNDMR (Banque nationale des maladies rares) are French databases funded by the French Ministry of Health, and REDgistry is publicly funded through the European Commission. FREDD (French Rare Eye Disease Database) represents the bridge between the national and European ecosystems.



This manuscript presents the concrete deployment of this national infrastructure, detailing its methodological framework while providing a critical appraisal of both the enabling factors and the operational challenges encountered during its initial phases. To ensure a rigorous and standardized presentation of our system architecture, governance, and deployment outcomes, the manuscript has been formatted as an implementation report and adheres to the iCHECK-DH reporting framework [27].

Methods

Ethical Considerations

This article describes the development and implementation of the FREDD database and the assessment of its data collection and integration processes. This did not involve a research project using FREDD data for secondary analysis or the generation of new clinical research findings. The FREDD database has been authorized by the French data protection authority (CNIL) for health data processing under decision DT-2024-008 dated April 6, 2024. Access to FREDD data

for secondary research use is subject to a separate governance and data access procedure.

Blueprint Summary

The CNIL HDW reference framework [28] underpins FREDD's design and operation, guiding both its technical architecture and regulatory implementation. It defines more than 100 organizational, technical, and legal requirements across 12 domains. A data protection impact assessment (DPIA) serves as the core regulatory document demonstrating compliance.

Technical Design

To support multicenter data collection, postcollection quality control, and secure research reuse, FREDD relies on a centralized 3-layer HDW architecture:

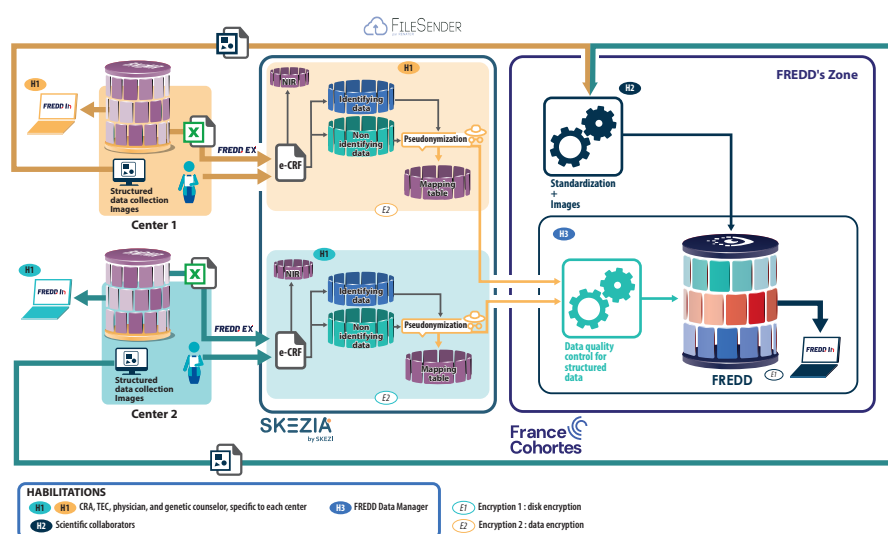
First, the collection layer uses a dedicated electronic case report form (eCRF) developed with SKEZIA, a compliant and cost-effective commercial platform developed by SKEZI [29], for standardized clinical data capture and early pseudonymization at participating sites level. Second, the storage and processing layer is hosted within the French National Institute of Health and Medical Research (Institut

national de la santé et de la recherche médicale; INSERM) France Cohortes infrastructure [30], where clinical datasets, genetic data, and high-volume imaging batches are securely ingested, harmonized, and curated. Third, the research layer provides project-specific, isolated virtual environments for governance-approved secondary data analysis under strict access permissions.

Data transfers between collection and processing environments currently rely on standardized, semiautomated

batch exports via FileSender [31]. Access control follows the principle of least privilege, enforced through a formalized data access matrix. The overall system architecture and end-to-end data life cycle are illustrated in Figure 2. Granular technical specifications are detailed in Multimedia Appendix 1.

Figure 2. Diagram of the FREDD (French Rare Eye Disease Database) data life cycle within the information system, illustrating the different data processing steps and the measures taken to ensure data security and confidentiality, in accordance with the national regulatory framework and guidelines for health data warehouses. The diagram highlights the data life cycle from the collection layer, managed within the SKEZIA environment, to the data quality, storage, and historization layer in the France Cohortes environment. It also specifies who can access each type of data depending on their role. CRA: clinical research associate; eCRF: electronic case report form; FREDDIn: FREDD-Insight; TEC: Clinical Research Technician.



Target

While FREDD is designed to ultimately aggregate data for all patients affected by any RED nationwide, the initial pilot phase implements a progressive and controlled deployment strategy. To validate the infrastructure, evaluate the warehouse's scientific value, and assess center engagement and adoption of the new data-sharing workflows, initial data collection focuses on a selected group of pilot centers and restricts its clinical scope to inherited retinal dystrophies (IRDs). Within this broader IRD population, particular emphasis was placed on retinitis pigmentosa (RP), which represented the largest patient cohort available for early evaluation.

Data

Structured Data

In line with regulatory requirements, FREDD captures only variables strictly necessary for its research objectives. The dataset encompasses up to 447 variables across 13 categories, including administrative data, clinical and genetic diagnoses, disease history, visual function, treatments, disability, clinical protocols, and imaging deposits. To ensure baseline research completeness, a maximum of 153 (34.2%) variables are designated as mandatory. Crucially, this framework

is dynamic rather than static: field display and mandatory status adapt conditionally to each patient's clinical pathway, preventing clinical team overburden. Detailed variable distributions and complex categories are provided in Multimedia Appendices 2 and 3.

Imaging Data

A distinctive feature of FREDD is the collection of retinal imaging data, including fundus photography, autofluorescence, and optical coherence tomography (OCT) single cross-sectional scans (B-scans) rather than full 3D volumes, which are essential for accurate diagnosis. To overcome heterogeneity across acquisition devices and formats while preserving structural features for AI research, images undergo a standardized 4-step processing pipeline: (1) resizing and color harmonization, (2) pseudonymized DICOM (Digital Imaging and Communications in Medicine) object reconstruction, (3) hybrid automated text detection and human-verified screening for embedded identity leaks, and (4) secure deletion of raw source files after validation.

To mitigate any risk of retaining directly identifying information, original source images are securely deleted from the processing environment immediately after processing. Consequently, only images that successfully complete this

full pipeline are permanently stored and uploaded into the final FREDD warehouse.

A step-by-step description of the image-processing pipeline is detailed in [Multimedia Appendix 4](#).

Structured Data Management and Monitoring

To ensure data accuracy and intercenter harmonization from the point of entry, standardized data entry procedures are disseminated to all participating centers. These procedures provide detailed operational definitions for each variable, standardized coding rules, and guidance for handling ambiguous clinical situations, thereby minimizing interoperator variability and reducing downstream data cleaning requirements.

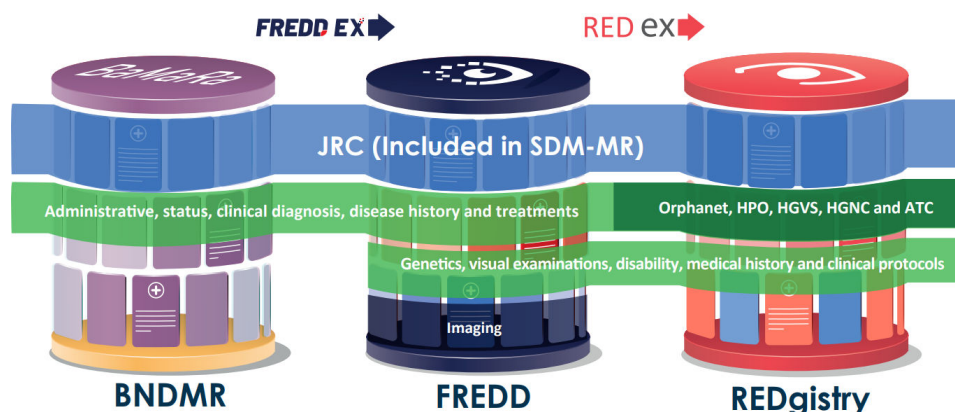
Postcollection curation, quality control, and tracking are driven by FREDDIn (FREDD-Insight), a centralized, in-house dashboard developed via Python scripts. FREDDIn

automatically performs data standardization, temporal consistency checks, anomaly detection, and patient deduplication. It is deployed both locally, to support participating centers, and centrally within France Cohortes for national quality monitoring.

Interoperability

To maximize semantic and technical interoperability, the dataset is built upon national and international standards. As illustrated in [Figure 3](#), the FREDD data model integrates the French BNDMR core framework (via the Set de Données Minimum–Maladies Rares [32]) and overlaps with the European REDgistry schema, which is conceptually derived from the Joint Research Centre common data elements [33,34]. To address the limitations of these generic rare disease frameworks, FREDD extends this core model with ophthalmology-specific variables and imaging data.

Figure 3. Diagram illustrating FREDD's (French Rare Eye Disease Database) architectural role as a technical and semantic bridge between the French national rare disease registry (Banque Nationale de Données Maladies Rares [BNDMR] and Base Maladies Rares [BaMaRa]) and the European registry (REDgistry). The blue bar highlights the European Joint Research Centre (JRC) core variables shared across all repositories. Green bars delineate the specific clinical data categories captured and the international used for standardized semantic mapping. The dark blue bar highlights hyperspecialized ophthalmology extensions and imaging data unique to FREDD. Red arrows highlight automated data mapping pipelines powered by custom parsing tools: FREDDEX for data ingestion from BaMaRa, and REDDEX for data exports toward REDgistry. ATC: Anatomical Therapeutic Chemical Classification; HGNC: HUGO Gene Nomenclature Committee; HGVS: Human Genome Variation Society; HPO: Human Phenotype Ontology; JRC: joint research center; SDM-MR: Set de Données Minimum–Maladies Rares.



To significantly reduce manual data entry, FREDD uses FREDDEX [35], a dedicated parsing and mapping tool that automatically retrieves clinical data already available in each patient's BaMaRa record upon prospective enrollment into FREDD. Among the 447 variables defined in the FREDD dataset model, 281 are potentially retrievable from BaMaRa, corresponding to a theoretical maximum prefill capacity of 62.9%. This represents the upper interoperability potential of the pipeline, while the effective completion rate remains dependent on source data completeness and quality. Considering duplicated or repeated variable structures within FREDD (eg, main and additional diagnoses), these 281 target variables correspond to 83 unique BaMaRa source variables that can be extracted and mapped to FREDD. Among these source variables, 23 are mandatory BaMaRa fields and are therefore systematically available, providing a guaranteed baseline availability of 28% of BaMaRa-derived variables.

FREDDEX extracts BaMaRa exports, maps the variables to their corresponding eCRF fields, and enforces international

domain-specific ontologies and standards, including disease classification using Orphanet (ORPHA) codes [36], phenotypic annotation using Human Phenotype Ontology (HPO) terms [37], and genetic nomenclature following HUGO Gene Nomenclature Committee (HGNC) [38] and Human Genome Variation Society [39] standards for gene and variant identification.

Furthermore, a complementary tool currently under development, REDDEX, leverages these identical interoperability principles to enable automated, structured data transfer from FREDD back toward the European REDgistry infrastructure. REDDEX is currently undergoing production testing and is scheduled to be fully operational by late 2026.

Participating Entities

FREDD is an initiative developed and led by the SENS-GENE rare disease network, which drove the inception of the project. Administratively and legally, FREDD operates under

the responsibility of the INSERM [40], acting as the central data controller.

To demonstrate the feasibility of the architecture and gather early validation data, FREDD is currently deployed across 5 pilot expert centers within the SENSGENE network, which actively feed data into the warehouse:

1. CARGO (Rare Disease Center for Ocular Genetics) at Strasbourg University Hospital
2. REFERET (Rare Disease Center for Hereditary Retinal Dystrophies) at National Ophthalmology Hospital (Quinze-Vingts), Paris
3. REFERET at Nantes University Hospital
4. MAOLYA (Rare Disease Center for Severe Myopia, Neuro-Ophthalmology, and Hereditary Optic Neuropathies) at Lille University Hospital
5. MAOLYA at Montpellier University Hospital

Regulatory Aspects, Governance, and Phased Deployment

The FREDD legal framework governs all data life cycle operations through 3 central pillars:

First, governance and data reuse, where a steering committee oversees strategic policies twice yearly, data access for reuse requires a formal review of all research proposals by the scientific and ethics committee, and researchers receive variable-specific completeness reports before extraction to guide data quality strategies.

Second, legal and ethical formalization, where operational standards are defined in a governance charter approved in 2025, complemented by data-sharing agreements with participating centers.

Third, patient rights and transparency, where to ensure full GDPR and CNIL compliance, patients are informed through a dedicated transparency portal and physical brochures in clinical centers, with accessible formats specifically provided for visually impaired patients.

Following the DPIA submission in December 2022, FREDD obtained official CNIL authorization in April 2024. The collection layer became operational and compliant in March 2025, allowing structured data entry to begin while the full France Cohortes infrastructure finalized compliance in November 2025. During this interim phase, collected datasets remained confined to the collection layer; imaging integration was initiated thereafter. The detailed timeline is available in [Multimedia Appendix 5](#).

Budget Planning and Sustainability

The launch and development of FREDD were made possible by the RaReTiA project (ANR-21-PMRB-0009), initiated in October 2022. This project was funded through a specific call for proposals launched by the French Ministries of Health and Research under PNMR3 and the France 2030 program, titled Accelerating Research and Innovation on Rare Diseases through Databases.

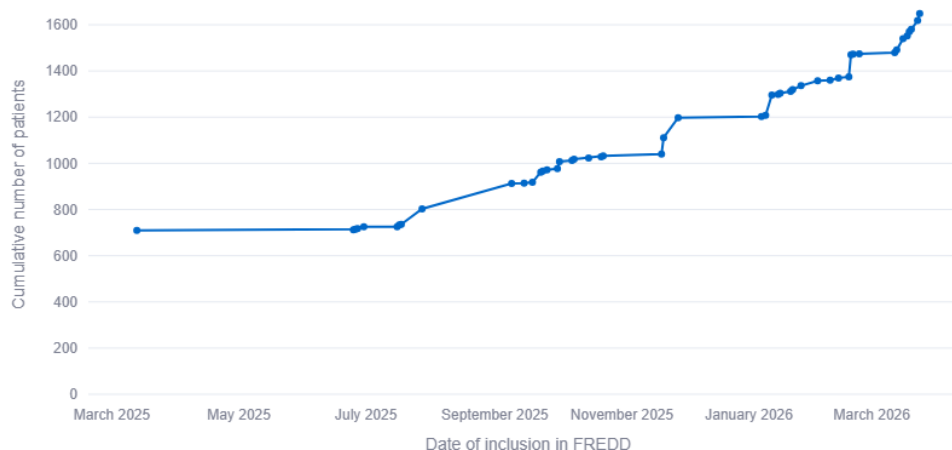
This dedicated budget successfully covered the entire “build” phase of FREDD, as well as its initial deployment and the beginning of the “run” phase. However, because this institutional funding is time-limited, it does not cover the expenses associated with FREDD beyond the scope and duration of the RaReTiA project.

Implementation (Results)

Coverage

Over a 1-year data acquisition period spanning from March 2025 to March 2026, FREDD successfully included a total of 1649 patients across the 5 participating pilot centers, demonstrating a steady and robust inclusion dynamic ([Figure 4](#)). Image collection was initiated in November 2025, and more than 10,000 ophthalmic images were subsequently integrated into the warehouse during the study period.

Figure 4. Evolution of patient inclusions in FREDD (French Rare Eye Disease Database) from the start of data collection in March 2025 to March 2026, representing 1 year of data acquisition.



Among the 1649 included IRD patients, 763 (46.3%) had a diagnosis of RP (ORPHA:791). According to the February 2026 BNDMR national report [41], 3764 patients with RP

have been registered nationwide since the deployment of the registry. Therefore, within its first year of implementation and across only 5 expert centers, FREDD has already enriched

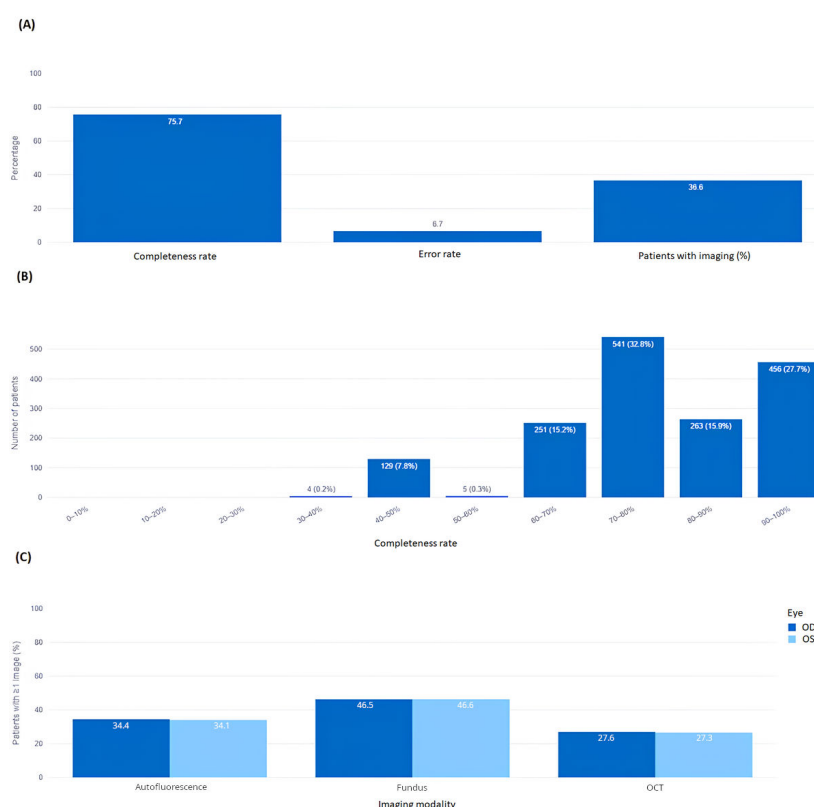
data corresponding to 20.3% (763/3764) of the national BNDMR RP cohort. These patients originated from existing BaMaRa records and were further characterized through the systematic collection of disease-specific ophthalmological variables, multimodal imaging data, centralized quality control procedures, and research-oriented governance. This rapid enrichment of a substantial proportion of the national RP cohort highlights the potential of disease-specific data infrastructures to transform routine rare disease records into structured, research-ready resources.

Outcomes

Overall, the FREDD database demonstrates high completeness for mandatory variables (Figure 5A). Figure 5B shows a

heterogeneous completeness distribution, with many patients between 30% and 60%, primarily corresponding to FRED-DEX-imported records that were prefilled according to the information available in BaMaRa and are progressively enriched through routine curation. Approximately one-quarter (456/1649, 27.7%) of patients have fully completed (100%) the mandatory data.

Figure 5. Overview of data quality, completeness scores, and ophthalmic imaging coverage. The charts provide a descriptive evaluation of the integrated dataset during the first year of data collection. (A) Overall data quality metrics showing the mean completeness rate, error rate, and the percentage of the cohort with associated imaging data. (B) Distribution of the patient cohort (N=1649) across categorical data completeness score ranges, detailing both absolute counts and percentages for each bin. (C) Proportional imaging coverage across the primary modalities, stratified by eye lateralization (right eye [OD] vs left eye [OS]). OCT: optical coherence tomography.



Data quality analysis (Figure 5A) identified inconsistencies in 7% (110/1649) of patients. These mainly reflect logical inconsistencies related to incomplete records rather than incorrect data entry, such as missing dependent variables.

Slightly less than 40% (605/1649, 36.7%) of patients had at least one associated image (Figure 5A). Image completeness varied across modalities (Figure 5C), reaching nearly 50% (767/1649, 46.5% and 768/1649, 46.6%) for fundus photography versus approximately 30% (455/1649, 27.6% and 451/1649, 27.3%) for OCT. This difference reflects both routine clinical practice and operational constraints, as OCT and autofluorescence are acquired less systematically and

imaging integration into France Cohortes requires additional workflows outside SKEZIA.

Lessons Learned

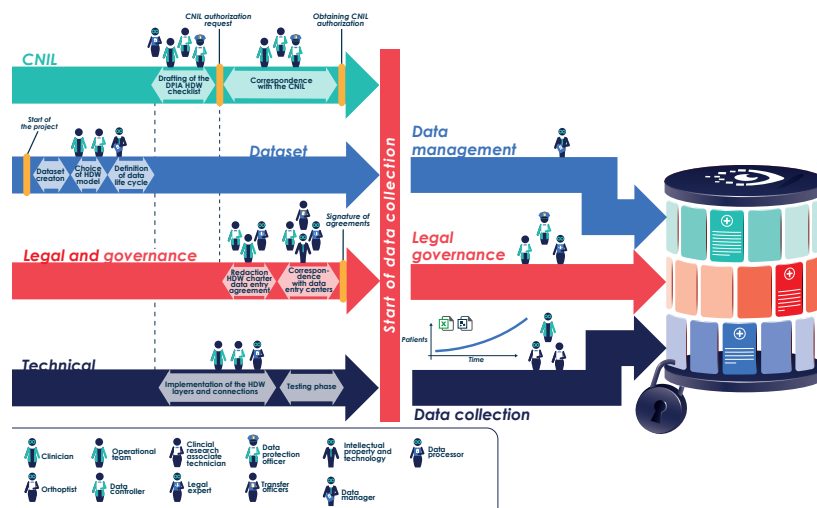
Operational Bottlenecks and Challenges

Institutional and Regulatory Workstreams

Setting up FREDD within the highly demanding CNIL HDW framework introduced significant decision-making latency due to the need to coordinate numerous institutional stakeholders. To mitigate these administrative hurdles, our experience demonstrates that regulatory, legal, and techni-

cal workstreams must progress concurrently rather than sequentially to maintain project momentum (Figure 6).

Figure 6. Parallel timelines leading to the creation of a health data warehouse (HDW) through 4 key areas: regulatory (involving the French regulation bodies), data-related (dataset definition), legal (establishment of legal documentation), and technical (implementation of tools for data processing). The key actors are also displayed. CNIL: French data protection authority; DPIA: data protection impact assessment.



Beyond central governance, preparing the database deployment across multiple hospital networks requires substantial time. In this new AI era, healthcare institutions have become acutely aware of the strategic value of clinical and imaging data for training AI models. Consequently, negotiating data-sharing agreements is no longer an administrative formality; it involves protracted, detailed discussions to protect institutional assets and ensure clear returns on investment.

Crucially, these combined delays create an operational bottleneck where project staff are fully funded but data collection cannot yet begin, risking budgetary exhaustion prior to final deployment. Project planning must therefore build strict financial and temporal flexibility into timelines to absorb the impact of these modern negotiation landscapes.

Human Resources, User Engagement, and Tool Flexibility

The implementation of FREDD underscored an acute scarcity of specialized personnel—such as data protection officers, health data lawyers, data engineers, and clinical research technicians—within French public health care institutions, highlighting a systemic need to professionalize these hybrid roles. This resource scarcity directly challenged on-the-ground user engagement.

To prevent the platform from being perceived as an administrative encumbrance by clinical teams, securing long-term medical commitment was essential. We overcame this by implementing role-specific onboarding protocols, providing continuous on-site support, simplifying user interfaces, and demonstrating the direct scientific value of each module to build investigator trust.

Furthermore, data collection tools must be engineered with sufficient flexibility to align with diverse real-world

clinical routines, as rigid architectures inevitably cause friction and workflow disruptions. Establishing a continuous feedback loop between participating pilot sites and the central operational team proved vital for iteratively refining the eCRF, the FREDDEX parsing tool, and automated quality control modules.

Key Success Factors

Rapid Early Outcomes and Onboarding Dynamics

Once the regulatory and technical foundations were finalized, platform adoption across the pilot centers was remarkably swift, reflecting strong initial engagement by the clinical teams. This effective uptake is illustrated by a sharp, nonlinear increase in data collection, creating an unprecedented high-density repository for REDs.

Because strict regulatory frameworks and data quality pipelines were fully embedded into the infrastructure during the “Build” phase, ongoing operational efforts can now shift from structural setup to continuous adaptation. Future resources will focus entirely on microadjustments driven by emerging research queries and evolving data regulations, ranging from national compliance updates to broader European alignments.

Downstream Research and Public Health Impact

By systematically aggregating data from geographically dispersed expert centers, the FREDD repository is designed for continuous expansion, steadily increasing the statistics essential for RED research. This centralized, harmonized approach has already unlocked initial opportunities for rapid scientific use. The immediate availability of curated data

has supported the launch of early-stage initiatives, including preliminary multicenter epidemiological analyses and pilot AI-driven predictive modeling projects, which are currently ongoing.

Ultimately, by bridging this longitudinal clinical depth with a fully compliant regulatory framework, FREDD delivers a uniquely valuable asset for both clinical research and public health impact. It serves as a functional example for how modern HDWs can structurally transform multicenter collaboration, accelerating discoveries in fields where data scarcity remains the primary bottleneck.

A comprehensive synthesis of the key barriers, enabling factors, resource requirements, and actionable recommendations derived from this implementation process is provided in [Multimedia Appendix 6](#).

Discussion

Scalability, Imaging Standardization, and AI Readiness

Expanding FREDD from 5 pilot centers to a nationwide infrastructure will introduce organizational and technical challenges. Our implementation experience suggests that clinical data harmonization is unlikely to become the main bottleneck, as the common data model and standardized collection procedures proved transferable across centers despite differing local workflows. Instead, scalability will primarily depend on the storage and processing of ophthalmic imaging, the largest component of the dataset volume. Progressive onboarding should nevertheless allow infrastructure scaling over time.

Beyond structured clinical variables, ophthalmic imaging presents additional challenges for AI applications. Although the current workflow enables secure image integration, image standardization and normalization still require validation through ongoing research projects. Future work will therefore focus on refining these processes while developing scalable, multiexpert annotation frameworks for rare ocular manifestations.

Moreover, as FREDD expands beyond inherited retinal disorders, the data model will require disease-specific extensions. Its modular architecture, however, allows new ophthalmology-specific modules to be integrated while preserving the shared rare disease data core.

Long-Term Sustainability and Financial Models

Sustaining FREDD beyond its initial funding period presents significant financial and organizational challenges. While

maintaining the France Cohortes hosting environment and the SKEZIA platform generates recurring costs, human resources—including project management, data curation, regulatory oversight, quality control, and local data collection—represent the largest long-term expenditure.

As disease-specific HDWs currently receive limited governmental support for routine operations, FREDD is developing a data access and cost-recovery framework aligned with the HDH Starter Kit for HDWs [42]. This framework relies on three principles:

1. Differentiated access models—separating cost-recovery mechanisms for academic and industrial users
2. Public-interest preservation—ensuring that cost-recovery and industrial partnerships do not compromise the infrastructure's scientific mission
3. Support for participating centers—reinvesting cost-recovery fees into local data collection, storage, and curation activities

As disease-specific HDWs remain relatively new infrastructures, sustainable funding models are still evolving. FREDD's long-term viability will therefore depend on iterative refinement based on operational experience, user demand, and evolving national and European funding priorities.

Conclusions and Future Perspectives

The FREDD initiative demonstrates that strict adherence to the CNIL HDW framework from the very outset does not hinder technological innovation. Instead, this early regulatory integration serves as a foundational catalyst for institutional trust—providing the exact governance guarantees that health care networks now demand before releasing their high-value datasets into collaborative research ecosystems.

While establishing this framework required substantial upfront organizational and financial investments over several years, the rapid inclusion of 1649 patients within the first year of active data collection signals that FREDD has successfully moved past its implementation phase.

At this stage, the critical challenge shifts from infrastructure setup to scientific activation. The true value of the warehouse will now depend on the systematic launch of multicenter research projects, the deployment of AI-driven clinical models, and its deep integration into national and European data ecosystems. By strengthening direct interoperability with cross-border infrastructures such as the REDgistry, FREDD serves as a scalable research engine whose long-term impact will transform collaborative research in the field of rare diseases.

Acknowledgments

The authors used ChatGPT (OpenAI) to assist with English language editing and improve readability. All outputs were carefully reviewed and validated by the authors, who take full responsibility for the final content of the manuscript.

Funding

This work was supported by funding from the French government managed by the National Research Agency (ANR) under the France 2030 program, through the RaReTiA project, reference ANR-21-PMRB-0009.

Data Availability

The individual-level pseudonymized data compiled within the French Rare Eye Disease Database (FREDD) are not publicly available due to legal, national regulatory, and patient privacy constraints surrounding health data warehouses. However, the data can be made accessible for secondary research use upon reasonable request, subject to approval by the FREDD governance board.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Detailed technical architecture, data security, and storage workflows.

[DOCX File (Microsoft Word File), 14 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Structure of the FREDD (French Rare Eye Disease Database) dataset and distribution of variables by category.

[DOCX File (Microsoft Word File), 14 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Detailed composition of complex categories of FREDD (French Rare Eye Disease Database) dataset.

[DOCX File (Microsoft Word File), 14 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Detailed imaging processing pipeline.

[DOCX File (Microsoft Word File), 14 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Chronological timeline of FREDD (French Rare Eye Disease Database) infrastructure phases, regulatory milestones, and deployment.

[DOCX File (Microsoft Word File), 14 KB-Multimedia Appendix 5]

Multimedia Appendix 6

Synthesis of the FREDD (French Rare Eye Disease Database) implementation framework: barriers, enablers, resources, and recommendations.

[DOCX File (Microsoft Word File), 16 KB-Multimedia Appendix 6]

Checklist 1

i-CHECK-DH checklist.

[DOCX File (Microsoft Word File), 32 KB-Checklist 1]

References

1. Évaluation du plan national maladies rares 2011 - 2016 [Article in French]. Haut Conseil de la Santé Publique. 2016. URL: https://sante.gouv.fr/IMG/pdf/hcsp_pnmr2_evaluation.pdf [Accessed 2026-08-02]
2. Banque Nationale de Données Maladies Rares. URL: <https://www.bndmr.fr/> [Accessed 2026-03-16]
3. Jannot AS, Messiaen C, Khatim A, Pichon T, Sandrin A, BNDMR infrastructure team. The ongoing French BaMaRa-BNDMR cohort: implementation and deployment of a nationwide information system on rare disease. J Am Med Inform Assoc. Jan 29, 2022;29(3):553-558. [doi: [10.1093/jamia/ocab237](https://doi.org/10.1093/jamia/ocab237)] [Medline: [34741516](https://pubmed.ncbi.nlm.nih.gov/34741516/)]
4. Pichon T, Messiaen C, Soussand L, et al. Overview of patients' cohorts in the French National Rare Disease Registry. Orphanet J Rare Dis. Jul 3, 2023;18(1):176. [doi: [10.1186/s13023-023-02725-2](https://doi.org/10.1186/s13023-023-02725-2)] [Medline: [37400917](https://pubmed.ncbi.nlm.nih.gov/37400917/)]
5. Plan national maladies rares 2018-2022: partager l'innovation, un diagnostic et un traitement pour chacun [Article in French]. Ministère des Solidarités et de la Santé; 2018. URL: https://sante.gouv.fr/IMG/pdf/pnmr_3_v25-09pdf.pdf [Accessed 2026-08-02]
6. Health Data Hub. URL: <https://www.health-data-hub.fr/> [Accessed 2026-03-16]
7. SENSEGENE - Filière de Santé Maladies Rares. URL: <https://www.sensgene.com/> [Accessed 2026-03-16]

8. Plan national maladies rares 2025-2030: des territoires vers l'Europe. Ministère du Travail, de la Santé, des Solidarités et des Familles. 2025. URL: https://sante.gouv.fr/IMG/pdf/pnmr4_vf6.pdf [Accessed 2026-08-02]
9. Particulier. CNIL. URL: <https://www.cnil.fr/fr> [Accessed 2026-03-16]
10. Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés [Article in French]. Légifrance. URL: <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000886460/> [Accessed 2026-06-21]
11. Le règlement général sur la protection des données - RGPD [Article in French]. CNIL. URL: <https://www.cnil.fr/fr/reglement-europeen-protection-donnees> [Accessed 2026-07-21]
12. European Reference Network. URL: <https://www.ern-eye.eu/> [Accessed 2026-03-16]
13. REDgistry. URL: <https://redgistry.eu/> [Accessed 2026-03-16]
14. Proposal for a regulation - the European Health Data Space. European Commission. 2022. URL: https://health.ec.europa.eu/publications/proposal-regulation-european-health-data-space_en [Accessed 2026-08-02]
15. 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/)]
16. 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/)]
17. Karakachoff M, Goronflot T, Coudol S, et al. 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*. Jun 24, 2024;12:e50194. [doi: [10.2196/50194](https://doi.org/10.2196/50194)] [Medline: [38915177](https://pubmed.ncbi.nlm.nih.gov/38915177/)]
18. 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 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/)]
19. Lamer A, Popoff B, Delange B, et al. Barriers encountered with clinical data warehouses: recommendations from a focus group. *Comput Methods Programs Biomed*. Nov 2024;256:108404. [doi: [10.1016/j.cmpb.2024.108404](https://doi.org/10.1016/j.cmpb.2024.108404)] [Medline: [39241462](https://pubmed.ncbi.nlm.nih.gov/39241462/)]
20. Kaspar M, Liman L, Morbach C, et al. Querying a clinical data warehouse for combinations of clinical and imaging data. *J Digit Imaging*. Apr 2023;36(2):715-724. [doi: [10.1007/s10278-022-00727-3](https://doi.org/10.1007/s10278-022-00727-3)] [Medline: [36417023](https://pubmed.ncbi.nlm.nih.gov/36417023/)]
21. La Rosa A, Verdoux M, Riebler P, et al. Multimodal identification of a rare head and neck cancer patient cohort in the clinical data warehouse of Greater Paris Teaching Hospital. *ESMO Real World Data Digit Oncol*. May 2025;8:100151. [doi: [10.1016/j.esmorw.2025.100151](https://doi.org/10.1016/j.esmorw.2025.100151)] [Medline: [41647696](https://pubmed.ncbi.nlm.nih.gov/41647696/)]
22. Petit-Jean T, Gérardin C, Berthelot E, et al. Collaborative and privacy-enhancing workflows on a clinical data warehouse: an example developing natural language processing pipelines to detect medical conditions. *J Am Med Inform Assoc*. May 20, 2024;31(6):1280-1290. [doi: [10.1093/jamia/ocae069](https://doi.org/10.1093/jamia/ocae069)] [Medline: [38573195](https://pubmed.ncbi.nlm.nih.gov/38573195/)]
23. Lochmüller H, Badowska DM, Thompson R, et al. RD-Connect, NeurOmics and EUREnOmics: collaborative European initiative for rare diseases. *Eur J Hum Genet*. Jun 2018;26(6):778-785. [doi: [10.1038/s41431-018-0115-5](https://doi.org/10.1038/s41431-018-0115-5)] [Medline: [29487416](https://pubmed.ncbi.nlm.nih.gov/29487416/)]
24. Reihs R, Proynova R, Maqsood S, et al. BBMRI-ERIC Negotiator: implementing efficient access to biobanks. *Biopreserv Biobank*. Oct 2021;19(5):414-421. [doi: [10.1089/bio.2020.0144](https://doi.org/10.1089/bio.2020.0144)] [Medline: [34182766](https://pubmed.ncbi.nlm.nih.gov/34182766/)]
25. Taruscio D, Vittozzi L, Choquet R, et al. National registries of rare diseases in Europe: an overview of the current situation and experiences. *Public Health Genomics*. 2015;18(1):20-25. [doi: [10.1159/000365897](https://doi.org/10.1159/000365897)] [Medline: [25228300](https://pubmed.ncbi.nlm.nih.gov/25228300/)]
26. Fredd. URL: <https://fredd-eds.fr/> [Accessed 2026-03-16]
27. 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:e46694. [doi: [10.2196/46694](https://doi.org/10.2196/46694)] [Medline: [37163336](https://pubmed.ncbi.nlm.nih.gov/37163336/)]
28. Référentiel: relatif aux traitements de données à caractère personnel mis en œuvre à des fins de création d'entrepôts de données dans le domaine de la santé [Article in French]. CNIL; 2021. URL: https://www.cnil.fr/sites/cnil/files/atoms/files/referentiel_entrepot.pdf [Accessed 2026-08-03]
29. SKEZI. URL: <https://skezi.eu/> [Accessed 2026-03-17]
30. France Cohortes. 2026. URL: <https://francecohortes.org/> [Accessed 2026-03-16]
31. FileSender. Renater. URL: <https://www.renater.fr/services/collaborer-simplement/filesender/> [Accessed 2026-03-17]
32. Le set de données minimal national [Article in French]. Banque Nationale de Données Maladies Rares. URL: <https://www.bndmr.fr/les-donnees-collectees/le-set-de-donnees-minimal/> [Accessed 2026-06-21]
33. European platform on rare disease registration. European Commission. URL: <https://eu-rd-platform.jrc.ec.europa.eu> [Accessed 2026-06-21]

34. JRC partnerships. European Commission. URL: https://joint-research-centre.ec.europa.eu/about-us/partnerships_en [Accessed 2026-03-16]
35. Beluffi Marin C, Oswald M, Ratenet L, Stoll M, Chennen K, Dollfus H. Interoperable integration of a national rare disease registry into a rare eye disease data warehouse: implementation study. *JMIR Med Inform.* May 26, 2026;14:e79378. [doi: [10.2196/79378](https://doi.org/10.2196/79378)] [Medline: [42190267](https://pubmed.ncbi.nlm.nih.gov/42190267/)]
36. Orphanet. URL: <https://www.orpha.net/fr> [Accessed 2026-03-17]
37. Human Phenotype Ontology. URL: <https://hpo.jax.org/> [Accessed 2026-03-17]
38. Tweedie S, Braschi B, Gray K, et al. Genenames.org: the HGNC and VGNC resources in 2021. *Nucleic Acids Res.* Jan 8, 2021;49(D1):D939-D946. [doi: [10.1093/nar/gkaa980](https://doi.org/10.1093/nar/gkaa980)] [Medline: [33152070](https://pubmed.ncbi.nlm.nih.gov/33152070/)]
39. Hart RK, Fokkema IF, DiStefano M, et al. HGVS Nomenclature 2024: improvements to community engagement, usability, and computability. *Genome Med.* Dec 20, 2024;16(1):149. [doi: [10.1186/s13073-024-01421-5](https://doi.org/10.1186/s13073-024-01421-5)] [Medline: [39702242](https://pubmed.ncbi.nlm.nih.gov/39702242/)]
40. Inserm. URL: <https://www.inserm.fr/> [Accessed 2026-03-17]
41. Nombre de cas par maladie rare dans la BNDMR [Article in French]. Banque Nationale de Données Maladies Rares. URL: <https://www.bndmr.fr/publications/nombre-de-cas-par-mr/> [Accessed 2026-06-22]
42. Créer son entrepôt de données de santé (EDS). Plateforme des données de santé. URL: <https://www.health-data-hub.fr/starter-kit-EDS> [Accessed 2026-06-22]

Abbreviations

BaMaRa: Base Maladies Rares
BNDMR: Banque Nationale de Données Maladies Rares
CNIL: Commission nationale de l'informatique et des libertés
DICOM: Digital Imaging and Communications in Medicine
DPIA: Data Protection Impact Assessment
eCRF: electronic case report form
EHDS: European Health Data Space
ERN: European Reference Network
FREDD: French Rare Eye Disease Database
FREDDIn: FREDD Insights
FSMR: Filières de santé maladies rares
GDPR: General Data Protection Regulation
HDH: Health Data Hub
HDW: health data warehouse
HGNC: HUGO Gene Nomenclature Committee
HPO: Human Phenotype Ontology
INSERM: Institut national de la santé et de la recherche médicale
IRD: inherited retinal dystrophy
OCT: optical coherence tomography
PNMR: National Plan for Rare Diseases
RED: rare eye disease
RP: retinitis pigmentosa

Edited by Caroline Perrin; peer-reviewed by Krishna Sai Dheeraj Kalluri, Mohammad Al-Agil; submitted 23.Apr.2026; final revised version received 21.Jul.2026; accepted 22.Jul.2026; published 25.Aug.2026

Please cite as:

Beluffi Marin C, Oswald M, Ratenet L, Stoll M, Chennen K, Chabin N, Hubert J, Rey G, Le Meur G, Smirnov V, Meunier I, Audo I, Dollfus H

Building a National Interoperable Rare Eye Disease Data Warehouse: Methodological Framework and Implementation Report From the French Rare Eye Disease Database (FREDD) Initiative
JMIR Med Inform 2026;14:e99206

URL: <https://medinform.jmir.org/2026/1/e99206>

doi: [10.2196/99206](https://doi.org/10.2196/99206)

in JMIR Medical Informatics (<https://medinform.jmir.org>), 25.Aug.2026. 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.