BIOPRESERVATION AND BIOBANKING
Volume 00, Number 00, 2026
Mary Ann Liebert, A Part of Sage

DOI: 10.1177/19475535261428051

Requalification and Data Management of Pediatric Biological Samples Collected since 1984: A Case Study from a Neuroblastoma Collection

Laure Sanvee-Edoh,[1] Kouamé Ambroise Kintossou,[2] and Séverine Tabone-Eglinger[1]

Introduction: The Biological Sample Management Platform (PGEB) of the Centre Léon Bérard is responsible for the preparation, preservation, storage, and provision of the institution's sample collection. The PGEB was approached to integrate a neuroblastoma collection, one of the most common pediatric cancers. This collection was initiated through the diagnostic reference activity for neuroblastoma at the Centre Léon Bérard.
Objectives: To requalify pediatric biological samples collected between 1984 and 2024 and to make them available for pediatric research protocols.
Methods: This study began with an assessment of the data associated with each of the 21,389 samples in the collection. These data were then compared with relevant regulations and consensus documents related to biobank data management, as well as the minimum data requirements for research use. Based on this, a methodology for sorting samples—either for retention or disposal—was applied.
Results: A set of minimum information criteria was established to revalue the samples. Various texts applicable in France mention the minimum data required for making samples available, but no consensus exists, either nationally or internationally. Furthermore, 65% of the samples met the criteria and were retained for future research use.
Conclusions: This article describes the research work that led to the identification of minimum required data. However, further development is needed to harmonize practices regarding data management and interoperability.

Keywords: requalification of biological samples, preservation, quality criteria, clinical data, scientific research

Introduction

Biomedical research plays a crucial role in improving our understanding of biological and pathological mechanisms, with the aim of developing new therapeutic strategies. In the field of oncology, the study of biological samples is a key element in better understanding tumor mechanisms, identifying biomarkers, and testing new therapeutic approaches. These samples, obtained from patient-derived specimens, are stored in biobanks or biological resource centers, ensuring their availability for research purposes. However, their long-term storage and management require rigorous organization and the assurance that each sample can be properly identified and utilized.

The preservation of biological samples relies on various strategies tailored to their nature and the analyses to be performed in the future. Samples may be stored under different conditions: at room temperature (e.g., in paraffin), at −20°C, or under cryopreservation at various temperatures (−80°C, −196°C). While storage conditions directly affect sample viability, the quality of the associated data is equally critical to ensure their usability.[1] These data—comprising clinical, biological, diagnostic, and methodological information—are essential for ensuring sample traceability and scientific relevance. The absence or inconsistency of certain information (such as collection date, sample type, or patient consent) can compromise the usefulness of a sample, even if it has been stored under optimal conditions.[2–4]

Over the years, the evolution of biobank management practices has led to significant improvements in data structuring and tracking. The emergence of dedicated management software, or Biobank Information Management Systems (BIMS), has facilitated the organization and recording of sample

[1]Plateforme de Gestion des Échantillons Biologiques, Centre Léon Bérard, Cancer Research Center of Lyon, Université de Lyon, Université Claude Bernard Lyon 1, INSERM 1052, CNRS 5286, Lyon, France.
[2]Centre de Ressources Biologiques, Institut Pasteur de Côte d'Ivoire, Abidjan, Côte d'Ivoire.

lifecycle information. Legacy collections are sets of biological samples collected over an extended period, established before the digitization of data or the implementation of strict standards. As they are often stored in various formats (paper laboratory notebooks, Excel files, or Access databases), these data may contain gaps or errors that limit their reuse in current research.[5] But legacy collections can be valuable because they provide access to large sets of samples for research, whereas prospective collections often require substantial time to be established before supporting new projects. They can be essential for hypothesis validation and for conducting long-term follow-up studies. Thus, biobanks face a dual challenge: ensuring the physical preservation of samples while also guaranteeing the quality and completeness of the associated data over a long period of time.

In this context, the Biological Sample Management Platform (PGEB) of the Centre Léon Bérard (CLB) aimed to integrate a retrospective neuroblastoma collection that has been established since 1984. This collection was developed over the year by the CLB's Biopathology Department, which serves as a reference center for the diagnosis of this specific pathology. It represents a valuable resource for pediatric cancer research, but its value depends on the ability to requalify the samples based on the completeness and reliability of the available data.[6,7] The objective of this study is to identify the minimum essential criteria, based on the available data of the collection, to support the requalification of those legacy biological samples, and to propose a rigorous sorting methodology.

Although international neuroblastoma resources exist, many legacy diagnostic collections remain locally managed and heterogeneous, which supports the need for pragmatic requalification approaches based on minimum data requirements. Indeed, beyond this case study, the proposed framework aims to contribute to the development of general recommendations for managing legacy collections within biobanks, notably by supporting database interoperability and harmonized sample selection criteria. In this context, the integration of standardized clinical and biological classifications represents an important perspective for enhancing the accessibility and research value of requalified legacy collections. Such harmonization is particularly relevant for rare diseases such as neuroblastoma, where cross-cohort analyses may help explore relationships between clinical characteristics, biological markers, and therapeutic outcomes.

This study primarily addresses the first step of sorting a legacy collection: defining the minimal dataset required to enable research use, a less-discussed part of the requalification of legacy collections. This minimal dataset should be understood as a baseline threshold for eligibility and traceability, not as the full set of information required for specific downstream research questions. Other key dimensions, including regulatory eligibility and physical quality assessment, were also addressed for the requalification project. They are therefore considered in the workflow but are only briefly described here to support the overall understanding of the study.

Materials and Methods

Workable protocol for requalifying a long-term legacy collection

The protocol for requalifying data from a long-term legacy collection was intended to support a retention or disposal decision based on predefined minimum data requirements and quality checks. It comprised the following steps:

- Data overview: Identification of available information, missing variables, and the structure of the Access' neuroblastoma database tables.
- Data correction, retrieval, and completion: Missing or incomplete data elements, including consent status, were retrieved, and inaccurate fields were corrected using structured query language (SQL) queries and manual source review. Storage location fields were corrected when needed, including through physical verification when inconsistencies could not be resolved from legacy records. Data gaps were not filled by inference; when information could not be verified, it was left missing and handled through the eligibility criteria.
- Definition of the minimum dataset: The minimum dataset used to support retention decisions was defined as *a priori* and applied consistently throughout the process.
- Eligibility sorting based on minimum criteria: Samples were screened against the predefined minimum data requirements. Samples with missing data elements and consent required for retention were excluded from conservation.
- Quality testing on a representative subset: A representative proportion of samples underwent physical quality testing, applied only after data and consent eligibility had been established. These assessments included morphological evaluation on 5 μm tissue cryosections stained and reviewed by a certified pathologist, nucleic acid extraction and quantification from tissue, peripheral blood lymphocytes, bone marrow, and whole blood, as well as cellular viability analysis for blood-derived samples. These tests were performed independently of data completion and served as an initial screening of material integrity. More advanced and resource-intensive analyses (such as polymerase chain reaction-based assays, next-generation sequencing, or fluorescence *in situ* hybridization) are performed only upon specific research or diagnostic requests. Detailed laboratory protocols are outside the scope of this article and are therefore not reported.
- Synthesis and final decision: The combined outputs of data requalification and quality testing were reviewed to support the final decision to conserve or dispose of samples.
- Transfer to the CLB BIMS: Retained samples were registered or reassociated (when applicable) in the BIMS, enabling controlled tracking of storage location, movements, and sample status throughout their lifecycle and supporting regulatory compliance.

This requalification workflow was designed to be consistent with International Organization of Standardization (ISO) 20387:2018, with particular attention to requirements related to traceability of biological material and associated data, controlled data management, and quality control activities supporting consistent biobank operation (Table 1).

As the PGEB operates under ISO 20387:2018 accreditation, the requalification activities were conducted within an

Table 1. Alignment of the Legacy Data Requalification Workflow Steps with ISO 20387:2018

Protocol steps	*ISO 20387:2018 chapters*[8]	*Objective evidence*
Data overview	7.8.3 Quality control of data 7.10 Management of information and data	Data completeness/missingness overview spreadsheet
Data correction, retrieval, and completion (SQL/manual) of missing information	7.8.3 Quality control of data 7.10 Management of information and data	SQL extraction spreadsheet Curation spreadsheet (sample list + corrections)
Definition of the minimum dataset	7.8.3 Quality control of data 7.10 Management of information and data 8.4 Control of records	Reference list Minimum dataset table Decision workflow
Eligibility sorting based on minimum criteria	7.5 Traceability of biological material and associated data 7.8.3 Quality control of data 7.11 Nonconforming output 7.12 Report requirements	Exclusion list with rationale (by sample ID)
Consent and ethical aspects	4.1 General (ethical principles) 4.3 Confidentiality 7.3 Reception and distribution of biological material and associated data 7.10 Management of information and data 7.11 Nonconforming output	Consent/conditions-of-use check Regulatory declarations
Quality testing subset	7.5 Traceability of biological material and associated data 7.8 Quality control of biological material and associated data	QC results spreadsheet (by sample ID) Comparison with historical QC results
Synthesis and final decision	7.5 Traceability of biological material and associated data 7.10 Management of information and data 7.11 Nonconforming output	Consolidated retained/excluded list (by sample ID) Validated biopathology pathway
Transfer to the institutional BIMS	7.5 Traceability of biological material and associated data 7.7 Storage of biological material 7.10 Management of information and data 8.4 Control of records	BIMS transfer/import files Storage location reconciliation (physical check)

Correspondence between workflow steps for the sample requalification, relevant ISO 20387:2018 requirements, and associated objective evidence supporting the requalification process.

It illustrates how data management, traceability, regulatory compliance, and decision-making processes were addressed within the requalification protocol.[8]

BIMS, Biobank Information Management Systems; ID, identification number; ISO, International Organization of Standardization; SQL, structured query language; QC, quality control.

organizational framework governed by this standard, including established procedures and qualified personnel.

Data overview

The PGEB of CLB, BB-0033-00050, CRB CLB, Lyon, France, was entrusted with a neuroblastoma collection comprising 21,389 samples of various types: tumor tissue, DNA, blood-derived resources (plasma, serum, etc.), and nontumoral material. These samples were collected and stored between 1984 and 2024. The objective was to requalify these care-derived samples to make them available, with patient consent, for pediatric cancer research. However, the age of the collection has led to deficiencies in data management, and some essential associated data—such as diagnosis, sample collection date, sample type, sample location, or consent—were missing, rendering the samples unusable for research protocols.

The Access database constituted a legacy dataset and was treated as an initial working source rather than a validated reference, as it contained known inaccuracies and incomplete fields. For this reason, no precleaning was performed before initiating the protocol; data cleaning, correction, and completion were embedded as the first steps of the requalification work.

To requalify the samples, a minimum set of information needed to be retrieved and included in a new, more suitable database for research use. The state of the data, the relevance of the available information, and the methods used for recording and tracking traceability have been examined.

Methods for assessing data completeness

The data analyzed were extracted from a Microsoft Access database containing six different tables: constitutional DNA, tumor-derived DNA, nontumoral material preserved as dry pellet (NT MAT pellet), nontumoral material preserved in dimethyl sulfoxide (NT MAT DMSO), serum bank, and tumor bank. This software allows data to be compiled and linked to cross-reference information across the different tables. However, a junction table—necessary for this process—was missing and could not be created given the current state of the Microsoft Access database. The goal was therefore to identify missing and inconsistent fields and assess whether

they were resolvable from the available internal sources: the Microsoft Access tables, paper patient records, and the center's database (containing, e.g., name, surname, diagnosis, patient consent, and sample collection date).

Processing missing information and updating data

Because the Microsoft Access database lacked a junction table, missing information was retrieved using a semiautomated approach combining SQL queries and manual cross-checking across internal sources. Therefore, SQL queries were used to retrieve verified patient-related information from the CLB institutional electronic medical records (EMRs) based on a predefined list of unique patient identifiers extracted from the original Access database. The SQL outputs provided authoritative patient attributes, which were then used to correct incomplete or inconsistent records in the Access dataset.

Using this verified information, data were exported to cleaned Excel sheets and cross-checked across sources to consolidate patient-related fields and support consistent linkage between biospecimens originating from the same patient (using XLOOKUP or VLOOKUP formulas). This process was conducted using only internally available CLB records, with the institutional database serving as the single reference source and with the help of an information technology (IT) engineer. No external datasets were queried.

All linkages were manually reviewed. No automated or probabilistic matching, inference, or extrapolation was performed to avoid identity-related errors (identity vigilance). When associations between patient and sample could not be established with certainty, no linkage was created.

The recovered information was then added accordingly, allowing the process to move on to the sample sorting stage.

Minimum data completeness criteria

The minimum data completeness criteria were based on:

- Regulatory requirements, including:
  - The General Data Protection Regulation (GDPR) and the guidelines of the French Data Protection Authority (CNIL), as well as the French Public Health Code, which govern patient information and regulate the protection of personal data.[9–11]
  - Normative standards, such as ISO 20387:2018, which defines a data management system from ethical, regulatory, and quality perspectives for biobanking.[8]
- Health data standards from consortiums applicable to biobanks:
  - The Minimum Data Set for Rare Diseases (SDM MR) from the French National Rare Disease Data Bank, which establishes a core set of data that proved partially suitable for the current case.[12,13]
  - Groups working on data standardization and interoperability in biobanking, including:
    - The GrOupe inter-SIRIC sur le paRtage et l'Intégration des donnéeS clinico-biologiques en cancérologie (OSIRIS group) from the French National Cancer Institute. This is a French standard in the form of a theoretical data schema defining a minimal set of representative clinical data for cancer patients.[14,15]
    - The Biobanking and BioMolecular resources Research Infrastructure – European Research Infrastructure Consortium group (BBMRI-ERIC) and its Minimum Information About Biobank data Sharing (MIABIS) project, which aims to establish data collection criteria that facilitate information exchange and cooperation between biobanks.[16]

However, these systems do not specifically define the mandatory and necessary minimal clinical data required for biomedical research. They are more focused on the minimum criteria for patient care or for establishing interoperability between biobanks. To determine the relevant data, we also relied on the needs expressed by physician–researchers working with the PGEB. This work is intended to be applicable to any collection of biological samples derived from clinical care that was not managed by a biobank and requires requalification for research use, except in highly specific cases.

Criteria and process for sample sorting

Data available in the Microsoft Access tables were heterogeneous. Initial data retrieval therefore focused on key identifiers required to enable further data completion and sorting: the patient's first and last name, their EMR at CLB, consent for the use of this resource in the context of research, and the sample identification number (ID).

Patient identifiers and their EMR numbers, both available in the Access database, were cross-linked with the CLB database to retrieve consent information and allow early exclusion of samples not eligible for research reuse. With the sample ID, which ensured traceability of the biological material, these four initial pieces of information constituted key identifiers to retrieve additional associated data, including diagnosis, date of sample collection or freezing, and type of sample. Together with storage location, those elements defined the minimum "complete data" set used for eligibility assessment.

The sorting process was designed to progressively exclude samples based on objective constraints, starting with data traceability and availability, followed by regulatory and scientific eligibility, and finally by material-related considerations. Once the maximum amount of missing information was filled in, the samples could be sorted accordingly (Table 2 and Fig. 1). The samples were classified into four categories:

- To be retained: These are samples with complete associated data, as well as those considered valuable despite the absence of certain information (such as only the year of collection being known, meaning we only have an approximate collection date).
- To be discarded
  - Digitally: Samples for which the biological material is depleted or for which the physical location/sample ID is unknown. The information will not be transferred to the data management system of CLB.
  - Physically: Samples whose recorded information makes them unsuitable, meaning they have undergone an event that compromised their quality (temperature excursions, sample thawed upon receipt, unknown sample type, plasma sample without information on

TABLE 2. RETENTION CRITERIA CONCERNING COMPLETENESS OF ASSOCIATED DATA

Criteria based on	*To be discarded: digitally*	*To be discarded: physically*	*Retained as a tester*	*To be retained for transfer to the institutional database*
Data traceability and availability	Unknown sample ID or storage location	Unknown sample type (e.g., tumor, DNA, ...)	Unknown date of sample or diagnosis	Full data availability[a]
Regulatory and scientific eligibility	N/A	Opposition of patient to reuse sample or data for research purpose[b]	No opposition Low scientific interest for research	No opposition Potential scientific interest for research
Material-related considerations	No available material	Degraded samples	Not supposed to be degraded	Not supposed to be degraded

[a]For date of sample: year accepted.
[b]Except primary samples retained for medico-legal reasons.
To determine sample eligibility for requalification, retention criteria were identified and categorized according to data traceability, regulatory eligibility, and expected research usability.
N/A, not applicable.

the anticoagulant used, and degraded samples), or for which consent is missing.

- Tester: samples with missing information that make them unsuitable for research uses but still useful for assessing the overall quality of the collection (e.g., unknown diagnosis, irrelevant diagnosis, and approximate collection date). Their role is to attest to the physical quality of the collection. They were selected

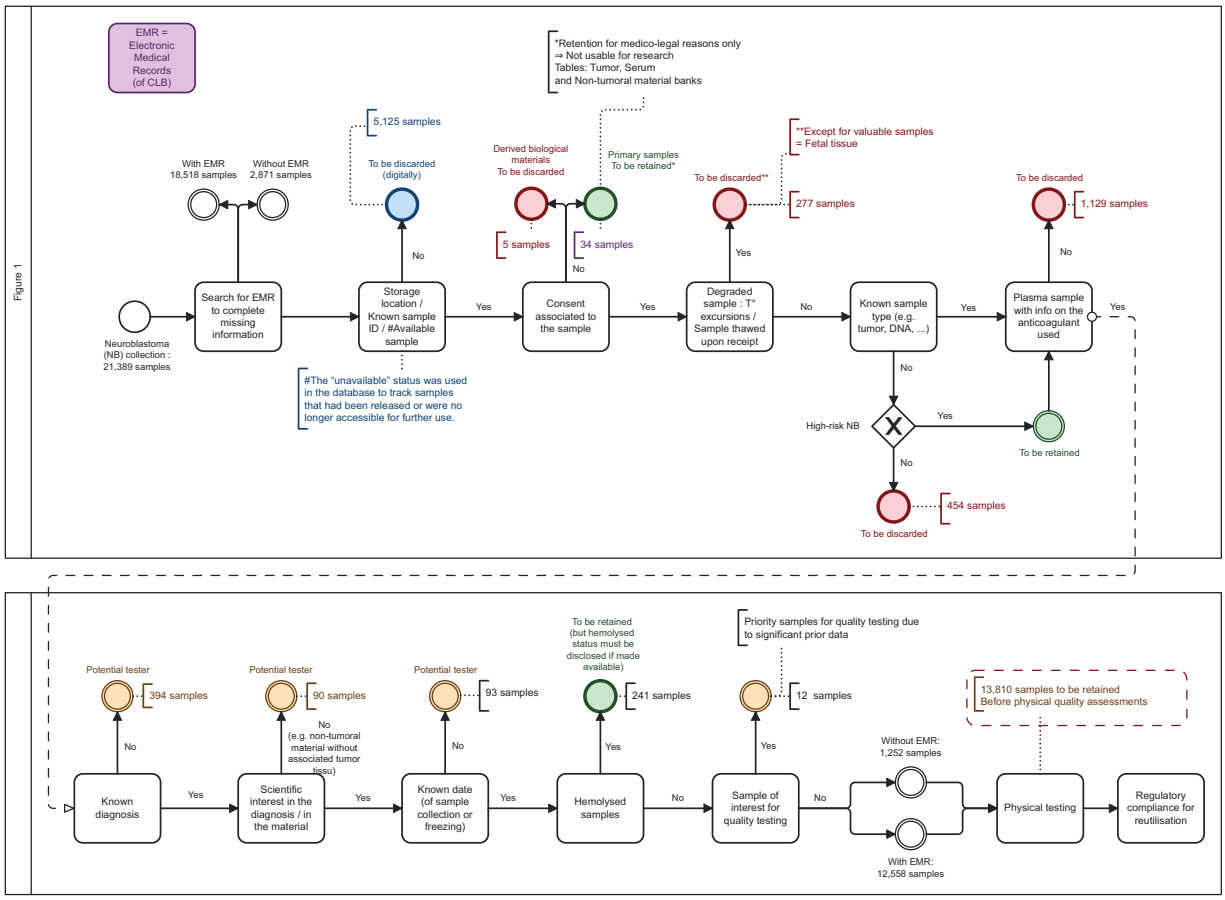

FIG. 1. Sample requalification workflow. Workflow illustrating the requalification and sorting of samples from a long-term legacy collection based on associated data and predefined decision criteria. Data traceability and completeness are first used to establish an evaluable sample set, followed by the regulatory and scientific eligibility criteria (including consent) and sample-related constraints to determine suitability or prioritization, with physical quality assessment applied last to evaluate material usability. While these steps are presented sequentially for clarity, data retrieval and completion were performed in parallel where possible, and the order reflects a logical decision framework rather than a strictly chronological process. Physical testing and the regulatory compliance steps are out of the range of the present study but are still essential elements of this requalification work.

in particular because multiple samples from the same patient of different types existed (tumor DNA and associated tumor, nontumor material, and associated DNA), based on their collection date (from 1989 to 2014) and the presence of prior data (quality control of DNA or RNA via Nanodrop).

Operational resources

The project mobilized multiple teams across the CLB. The PGEB team (biobank technicians and a biobank engineer) and the IT team (including a data manager) carried out the core of the requalification work, including data curation, verification, sample testing, and data transfer to the BIMS. The Biopathology Department contributed through technician and researcher support for data follow-up and handling of prospective samples. Regulatory compliance was ensured in collaboration with the Data Protection Officer team.

Overall, the requalification process was completed over an approximate period of 2 years.

Results

Data completeness

Some samples were found in multiple aliquots within the storage units, which explains the increase in the number of samples as well as the associated data (Table 3).

Diagnoses, which represent the category with the highest amount of recovered data (+66%), required significant time. However, they were actually the easiest information to retrieve when the EMR was available (see Table 3).

Sample distribution

Samples to be retained represent 65% of the total collection, indicating that most of the samples met the predefined minimum dataset and quality criteria (Fig. 2).

Samples to be physically discarded account for only 9%. Therefore, the space freed by these 1865 samples in the storage units is not sufficient to justify a major reorganization of sample storage at the box level.

Moreover, the samples are scattered across different storage units (3 freezers and 1 nitrogen tank with 5 racks of 12 boxes), making individual box rearrangement disproportionate in terms of time and effort. However, because the racks containing these samples are movable, the frozen samples will be consolidated into one and a half freezers to simplify access.

Tester samples represent approximately 3% of the collection and will either be kept for training purposes (e.g., technician practice in extraction techniques or cryostat sectioning) or used for quality control to assess the overall quality of the collection.

Research potential

As the collection was managed by the biopathology service, historical use of the collection could not be systematically quantified due to the absence of traceability in the management of sample releases prior to this work. However, some data were retrieved as follows:

- In 2024, approximately 5000 samples were flagged as released in the Access database, although the associated projects were not documented.

Table 3. Percentage of Available Data Before and After Data Completion Work

Data completeness before search	*Constitutional DNA*	*Tumor-derived DNA*	*NT MAT pellet*	*NT MAT DMSO*	*Serum bank*	*Tumor bank*	*Total (before gain)*	*Total*
Number of samples	632	1694	2919	1061	8162	6868	21,336	21,336
% Samples ID	100%	97%	99%	100%	100%	100%	95%	95%
% EMR	*0%*	24%	*0%*	*0%*	91%	57%	55%	55%
% Diagnosis	*0%*	*0%*	*0%*	*0%*	*0%*	96%	31%	31%
% Date of sample collection or freezing	96%	98%	98.3%	99.7%	100%	100%	99.5%	99.5%
% Type of sample	*0%*	97%	94%	66%	99.8%	96%	93%	93%

Data completeness after search	*Constitutional DNA*	*Tumor-derived DNA*	*NT MAT pellet*	*NT MAT DMSO*	*Serum bank*	*Tumor bank*	*Total (+ gain)*	*Total*
Number of samples	632	1694	2919	**1072**	8162	**6910**	**21,389 (+53)**	**21,389**
% Samples ID	100%	97%	99%	100%	100%	100%	95%	95%
% EMR	**77%**	**86%**	**89%**	**70%**	**92%**	**84%**	**87% (+32%)**	**87%**
% Diagnosis	**83%**	98%	**96%**	**82%**	**97%**	**99%**	**96% (+66%)**	**96%**
% Date of sample collection or freezing	**100%**	**100%**	**98.6%**	**100%**	100%	100%	**99.8 (+0.03%)**	**99.8%**
% Type of sample	**98%**	97%	94%	**96%**	100%	**96%**	**97% (+4%)**	**97%**

0% indicates information initially missing in the Microsoft Access database prior to data harmonization and search.

In bold are the data fields that were completed, notably through SQL queries and manual searches. The data provided in the table correspond to the minimum information for sample storage.

The EMR includes the patients' first name, last name, and consent.

EMR, electronic medical record; NT MAT DMSO, nontumoral material preserved in dimethyl sulfoxide.

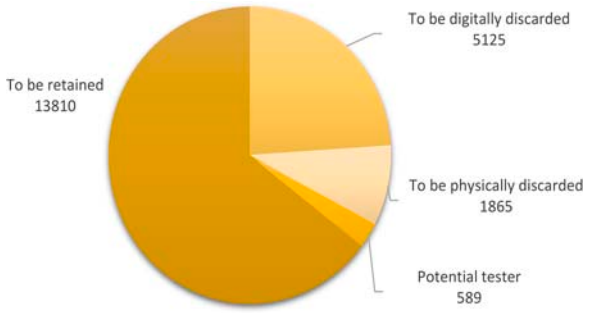

FIG. 2. Distribution of samples after sorting. The samples were dispatched in four categories after being sorted out: the ones to be retained and useful to pediatric research, the ones to be digitally or physically discarded, and the ones usable as tester (DNA extraction or cryostat sectioning).

- Since 2018, eight external access requests were reviewed by the institutional medico-technical committee, with an average of 50 samples released per request.
- In addition, 12 publications published between 2009 and 2015 cited the participation of the CLB through this collection.

Process outcomes and organizational implementation

As a direct outcome of this retrospective requalification work, associated data for retained samples are being migrated into the institutional CLB database to ensure traceability, regulatory compliance, and future research use, while the original Access database is preserved as an archive.

Importantly, the lessons learned from this large-scale retrospective effort have been incorporated into current biobank practices. Newly requalified neuroblastoma samples are now directly registered in the CLB's BIMS through the routine workflow of the Biopathology Department, avoiding the creation of parallel or intermediate databases. This prospective integration aims to prevent the recurrence of large-scale requalification needs.

At present, no other legacy collections requiring similar large-scale requalification have been identified within the institution.

Discussion

In this work, the minimum dataset was defined as a first-level qualification to ensure reliable cataloguing and screening against common request criteria (e.g., diagnosis, preanalytical variables, treatments, outcome, ...). More detailed clinical linkage remains possible and will be performed on demand at the time of sample request using the data transferred in the institutional BIMS; the present study focused on an *a priori* triage of a large legacy collection based on general, verifiable minimum requirements.

There is currently no single international consensus on the minimum data management requirements for samples intended for research. The diversity of practices and regulatory frameworks across countries and institutions makes it difficult to establish such a standard. However, several initiatives have been under development or trial for the past few years:

- The previously mentioned projects—MIABIS, SDM-RD, ISO standards, CNIL, etc.—serve as essential foundations for achieving this long-awaited goal.[5,8,9,12–16]
- Initiatives such as the Practical Guide to the International Alignment of Research Data Management by Science Europe aim to harmonize data management requirements. This document highlights the importance of reliable databases, data collection, preservation, and sharing throughout the research process, as well as adherence to ethical and regulatory standards.[17]
- The European Union Regulation (EU) 2025/327 on the European Health Data Space, in effect since March 2025, is being progressively implemented. Its goals include enabling the secure reuse of data for secondary purposes, including research, and creating a single digital health market, providing a framework for system interoperability.[18]
- From a regulatory perspective, France has several legal texts (GDPR, Loi Jardé, Public Health Code, ISO 20387:2018, etc.) that define biobank activities. Countries such as Belgium, Spain, and Portugal follow similar principles. However, the Finnish Biobank Act is unique in that it centralizes the methods for creating, operating, and using biobanks for research purposes, making the implementation of such guidelines clearer.[19] Establishing a similar text at the international level would be ideal.
- In parallel, several research groups and biobank networks (including the Foundation for Innovative New Diagnostics and the Biological Resource Center of Institut Pasteur biobanks) have recently proposed structured frameworks to assess the value of existing biospecimen collections. These approaches focus on collection-level valuation and management, using staged evaluation strategies, scoring systems, and predefined eliminatory criteria to support decisions related to collection long-term conservation or culling. Some frameworks distinguish regulatory and ethical feasibility from biospecimen and data-related characteristics, while others emphasize fit-for-purpose assessment, anticipated scientific use, and operational considerations such as storage costs and sustainability. While these initiatives aim to provide standardized and transparent tools to support strategic decision-making at the collection level, our experience and results address sample-level sorting and data rescue within a single large legacy collection, thereby highlighting distinct levels of analysis that can be combined when adapted to specific operational contexts. To be clearer, the criteria described in those papers are relevant and complementary to our work (data management, consent, storage condition, legal requirement, quality control, published research, ...). The difference is that those articles used these criteria to assess their collections as a whole, whereas we evaluated the same type of criteria independently for each sample.[20–22]

Identifying minimum data requirements to facilitate data sharing and reuse is a substantial task, requiring collaboration among biobanking entities. Table 4 illustrates current differences. Although these initiatives do not address all the same

Table 4. Comparison of Minimum Data Between Different Systems

Mandatory information from	*SDM MR*[12,13]	*OSIRIS (INCa)*[14,15]	*ISO standards*[8] */CNIL*[9]	*MIABIS (BBMRI)*[16] *Practical guide (Science Europe)*[17] *Regulation (EU) 2025/327*[18]	*Research groups and biobank networks*[20–22]	*Criteria retained for the requalification work*
Patient consent	Yes	Yes	Yes	Aims to ensure interoperability between different infrastructures. As such, there are no mandatory criteria, only suggested ones. These criteria are essential to the success of the project and have therefore been used to define a set of minimum standards for the purposes of this article	Yes	Yes
Diagnosis	Yes	Yes (cancer-specific device)	Regulations governing the quality and compliance of resource management with respect to patient data protection. They define methods for safeguarding personal data, such as patients' rights to object or consent, as well as data anonymization or pseudonymization		Yes	Yes
Disease occurrence (initial tumor, recurrence, metastasis)	N/A	Yes			Yes	No, but available for certain cases
Patient type (if fetus)	Yes	N/A			Yes (as rare and scarce samples)	Yes
Patient identification information, personal information	Yes	Yes			Yes	No, sample provision is deliberately anonymous
Family history	Yes	No			N/A	No, sample provision is deliberately anonymous
Medical history	Yes	No			Yes	No, but available for certain cases
Date (of sample collection or freezing)	N/A	Yes			Yes	Yes
Sample type	N/A	Yes			Yes	Yes
Sample ID and storage location	N/A	Yes			Yes	Yes
Vital status	Yes	Yes			N/A	No, but available for certain cases
Transmitting center ID	N/A	Yes			N/A	No, but available for certain cases
Clinical trial data	N/A	Yes			N/A	N/A

The criteria adopted by the PGEB regarding minimum information for the requalification of biological samples have been established according to the Rare Diseases Minimum Data Set (SDM MR), the OSIRIS group of INCa, ISO, and CNIL standards, and the Minimum Information About Biobank Data Sharing (MIABIS) project from the BBMRI group. The second-to-last column compares our retained predefined minimal data (last column) with collection-level work conducted by research groups and biobank networks.[8,9,12–18,20–22]

CNIL, French Data Protection Authority; INCa, French National Cancer Institute; N/A, not applicable; OSIRIS, GrOupe inter-SIRIC sur le paRtage et l'Intégration des donnéeS clinico-biologiques en cancérologie; PGEB, Biological Sample Management Platform of CLB.

aspects, they share a common ground. Yet, since each research project is different, some will inevitably require more or less information. Implementing national or international standardization remains a major challenge for biomedical research. Currently, customized and individual approaches, combined with strict ethical frameworks, are what ensure the quality and reliability of biobank collections.

This study shows that for broad-scope research, and specifically in the context of research based on the requalification of biological samples collected for diagnostic purposes (nonhuman subject research according to CNIL), the minimum required information includes the patient's diagnosis, their consent for reuse, the sample collection or freezing date, the sample type, and its specific characteristics (e.g., the anticoagulant used for plasma or the tumor origin of a tumor DNA sample). From a logistical standpoint, a unique ID and the physical location of the sample are also mandatory. While this information may not be sufficient for all biomedical research, it remains fundamental, and samples from this collection that did not meet this baseline were excluded. However, in practice, the samples made available for research will likely have more associated data than this minimum. They will be accompanied, when accessible, by additional clinically relevant annotations accessible in institutional records, including standardized neuroblastoma classifications (e.g., International Neuroblastoma Staging System/International Neuroblastoma Risk Group Staging System [INSS/INRGSS] stage), key risk markers such as MYCN oncogene status, ploidy, segmental chromosomal aberrations (e.g., 1p/11q) or vital status, and years of follow-up from diagnosis.[23]

This study also highlights other limitations, such as the importance of having reliable and qualified data managers as well as the importance of centralizing all available data in secure and well-structured digital systems from the beginning of collection projects.[18–20]

In this case, paper records were extremely helpful in completing data that had not been digitized (EMRs and Access database). However, the transfer of paper data into the Access database and the manual searches conducted during this study introduced a nonnegligible risk of human error. This risk, even with digital management tools, cannot be entirely eliminated, but it can be minimized. The initial error rate of the Access database was not quantified. However, the protocol was made to minimize this risk, notably through verification against the institutional EMR used as the reference source, defined linkage rules, and strict handling of ambiguous cases (no linkage when uncertainty remained).

To improve the management of both legacy and new collections and to ensure their long-term usability, several recommendations can be made as follows:

- First, implementing a data verification and standardization process at the time of integration into a biobank would ensure the quality and scientific relevance of the samples.
- Second, adopting digital tools that allow for cross-referencing and centralization of information would reduce errors and improve database interoperability.
- Finally, collaboration between biobanks at both national and international levels would promote standard harmonization and strengthen access to samples for the scientific community.

Conclusions

Although the historical use of the collection cannot be fully quantified due to incomplete traceability in the collection handling system over the years, available indicators such as documented access requests and publications demonstrate sustained scientific interest.

This underestimation of past use was a central motivation for the requalification effort and highlights the critical importance of rigorous data management associated with biological samples to ensure their usability and scientific relevance in biomedical research. It requires a methodical approach to data sorting and completion. Implementing tailored solutions for legacy collections will help maximize their scientific value and ensure their contribution to advancing knowledge in oncology and other biomedical research fields.

Acknowledgments

The authors give special thanks to Ms. Isabelle Iacono Di Cacito, Advanced Laboratory Technician, who supported the authors throughout the management of this project; to Mr. Nawar Lwaissa, Data Manager, whose expertise enabled the retrieval of most patient data; and to the PGEB team for their assistance and kindness.

Authors' Contributions

L.S.-E.: Conceptualization, methodology, visualization, formal analysis, investigation, writing—original draft, visualization, and project administration. K.A.K.: Investigation, methodology, formal analysis, writing—review and editing, and validation. S.T.-E.: Conceptualization, resources, writing—review and editing, and supervision.

Ethics Statement

This study describes internal biobank management and data governance procedures applied to an existing collection of biological samples obtained during routine clinical care. No interventional procedures were performed, and no biological material was used for research analyses as part of this work. In accordance with the national regulatory framework governing the secondary use and conservation of human biological materials, these activities did not constitute research involving human participants and did not require prior ethics committee review or approval.[10,24] Outside the scope of this study, any subsequent use of samples within a defined research project, or any transfer of samples to third parties, would be subject to the appropriate ethical and regulatory authorizations under the applicable national framework.[10,24]

Author Disclosure Statement

No competing financial interests exist.

Funding Information

No funding was received for this article.

References

1. Speirs V. Quality considerations when using tissue samples for biomarker studies in cancer research. Biomark Insights 2021;16.
2. Shabihkhani M, Lucey GM, Wei B, et al. The procurement, storage, and quality assurance of frozen blood and tissue biospecimens in pathology, biorepository, and biobank settings. Clin Biochem 2014;47(4–5):258–266.
3. Hubel A, Spindler R, Skubitz APN. Storage of human biospecimens: Selection of the optimal storage temperature. Biopreserv Biobank 2014;12(3):165–175.
4. Haute Autorité de Santé (HAS). Cryopréservation – Recommandations. 2010. Available from: https://www.has-sante.fr/upload/docs/application/pdf/2010-02/cryopreservation_-_recommandations.pdf [Last accessed: October 8, 2025].
5. Olund G, Lindqvist P, Litton J-E. BIMS: An information management system for biobanking in the 21st century. IBM Syst J 2007;46(1):171–182.
6. Cianflone A, Savoia F, Parasole R, et al. Pediatric biobanks to enhance clinical and translational research for children. Eur J Pediatr 2023;182(4):1459–1468.
7. Valle-Simón P, Borobia AM, Pérez-Martínez A. Clinical research with targeted drugs in paediatric oncology. Drug Discov Today 2023;28(8):103672.
8. ISO. ISO 20387:2018. Biotechnology – Biobanking. 2018. Available from: https://www.iso.org/standard/67888.html
9. Commission nationale de l'informatique et des libertés (CNIL). Le règlement européen sur la protection des données. 2016. Available from: https://www.cnil.fr/fr/reglement-europeen-protection-donnees [Last accessed: September 18, 2025].
10. Légifrance. Code de la santé publique – Article L. 1211-2. 2021. Available from: https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000043895792/2024-06-12 [Last accessed: September 18, 2025].
11. Légifrance. Code de la santé publique – Article L. 1245-2. 2021. Available from: https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000041721161 [Last accessed: September 18, 2025].
12. Banque Nationale de Données Maladies Rares (BNDMR). Le set de Donnees minimal national. 2024. Available from: https://www.bndmr.fr/les-donnees-collectees/le-set-de-donnees-minimal/ [Last accessed: October 2, 2025].
13. Banque Nationale de Données Maladies Rares (BNDMR). Le set de Données minimal national. 2024. Available from: https://www.bndmr.fr/wp-content/uploads/2024/01/Set-de-donnees-minimum-national-maladies-rares_v1.12-29012024.pdf [Last accessed: October 2, 2025].
14. Guérin J, Laizet Y, Le Texier V, et al. OSIRIS: A minimum data set for data sharing and interoperability in oncology. JCO Clin Cancer Inform 2021;5:256–265.
15. Institut National du Cancer (INCA). OSIRIS data elements concepts. GitHub; 2024. Available from: https://github.com/InstitutNationalduCancer/OSIRIS/wiki/Home [Last accessed: October 2, 2025].
16. Eklund N, Engels C, Neumann M, et al. Update of the minimum information about BIobank data sharing (MIABIS) core terminology to the 3rd version. Biopreserv Biobank 2024;22(4):346–362.
17. Science Europe. Practical guide to the international alignment of research data management – Extended edition. Zenodo; 2021. Available from: https://www.scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf; doi: 10.5281/zenodo.4915862
18. Official Journal of the European Union. Regulation (EU) 2025/327 of the European parliament and of the council of 11 February 2025 on the European health data space and amending directive 2011/24/EU and regulation (EU) 2024/2847 (Text with EEA relevance). Official Journal of the European Union 2025. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202500327
19. Salokannel M, Tarkkala H, Snell K. Legacy samples in Finnish biobanks: Social and legal issues related to the transfer of old sample collections into biobanks. Hum Genet 2019;138(11–12):1287–1299.
20. El Idrissi I, Piquard V, Fransman W, et al. A quantitative tool for culling collections of human specimens; Proof of concept. Biopreservation Biobanking 2025; doi: 10.1177/19475535251362092
21. Kessler JA, Johnson TM, Henderson MK. The development of the biobank collections valuator as an automated tool to determine the noneconomic value of sample collections. Biopreservation Biobanking 2025; doi: 10.1177/19475535251374854
22. Rush A, Byrne JA, Watson PH. Guideline on valuation of research biospecimen collections. Biopreserv Biobank 2025; doi: 10.1089/bio.2024.0159
23. American Cancer Society and American Society of Clinical Oncology (ASCO). Stages, prognostic markers, and risk groups for neuroblastoma. 2025. Available from: https://www.cancer.org/cancer/types/neuroblastoma/detection-diagnosis-staging/staging.html [Last accessed: December 31, 2025].
24. Code de la santé publique – Article L. L1243-3. 2021. Available from: https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000043896103 [Last accessed: February 9, 2026].

Address correspondence to:
Laure Sanvee-Edoh, MS
Centre Léon Bérard – Plateforme de Gestion des Echantillons Biologiques
Cheney B (RDC)
28 Rue Laënnec
69008 Lyon
France

E-mail: laure.sanvee@gmail.com