
INSTITUTE OF APPLIED BIOSCIENCES
ΙΝΣΤΙΤΟΥΤΟ ΕΦΑΡΜΟΣΜΕΝΩΝ ΒΙΟΕΠΙΣΤΗΜΩΝ
CENTRE for RECEARCH and TECHNOLOGY-HELLAS

Institute of Applied Biosciences (INAB)
Centre for Research and Technology, Hellas (CERTH)

Real-World Data Management Systems

Development of **standardized** and **centralized** data repositories and management systems for **Real World Evidence**

enabling multi-center projects on clinical research studies

Web-based applications

Provide data availability and security in a user friendly way

Simplicity, data reliability, easiness in use and flexibility as necessary conditions

User – friendly, adjusted to user requirements

Real - world evidence (RWE) biomedical research is an increasingly important component of INAB | CERTH activities

RWD management systems

OBJECTIVES

A. collect and transform clinically relevant RWD into evidence

B. correlate data in order to provide accurate information about the diagnosis, prognostic assessment and management

C. facilitate translational and clinical research

Organize and analyze clinically relevant data from the **daily practice**, by gathering **homogenized high-quality datasets**, thus Improving the **quality and delivery of medical care**

RWD management systems

APPROACH

1. Standardization & Harmonization

Agreement on common policies and procedures with standards-based approaches, paving the way for standardized registration of RWD

2. Simplicity & Flexibility

Design in expandable modules allowing the rapid introduction of additional data categories and extension of data domain when and if needed

3. Data Availability & Usability

Unified access to clinically relevant structured data while fulfilling the data quality assurance requirements

RWD management systems

AIM

Combine information into a data integration framework

Data integrity, data accuracy and quality assessment

User-friendly integration of ontologies, terminologies and standards

Data correlations and statistical analysis

Data protection, security and availability

Data analytics, Data mining and knowledge discovery

RWD management systems

DATA PROTECTION

GDPR compliance

- Data are collected solely for specific and legitimate purposes
- Data are adequate and related to the purpose of collection
- Data are processed and treated lawfully and fairly in a transparent manner

ISO 27001:2013
for Information Security Management System

Data anonymization procedures

Pseudonymized identifier for each patient

The correspondence between original and anonymized records will be stored only at the local level.

Fully anonymized data used for statistical analysis. Aggregated results are reported.

ISO 22301:2019
for Business Continuity Management System

Appropriate technical and organizational measures
- data confidentiality, availability and integrity
- data protection and security

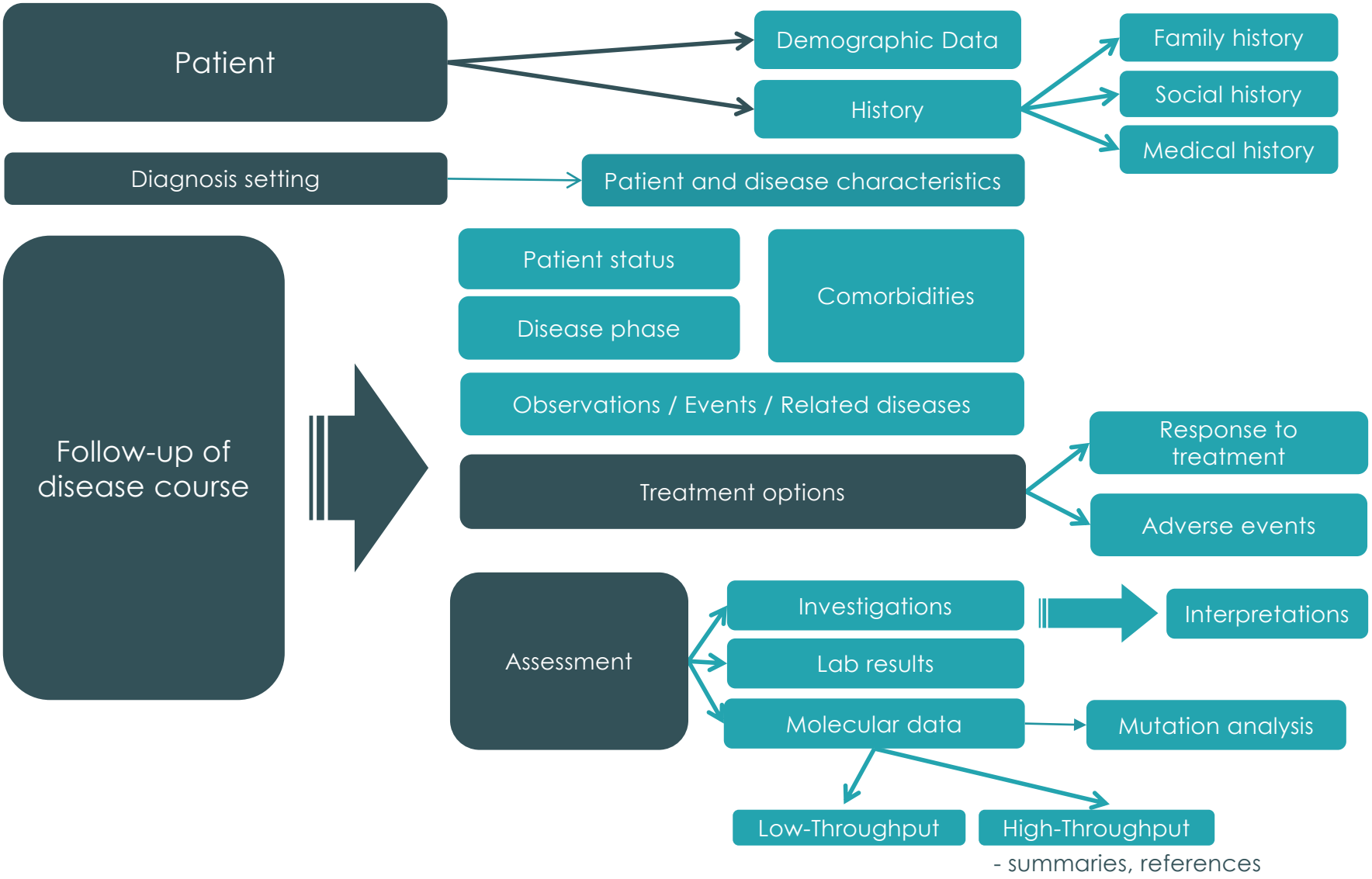
RWD management systems

DATA MODEL

Describe all the relevant information to form an accurate and complete representation of patient diagnosis and disease course

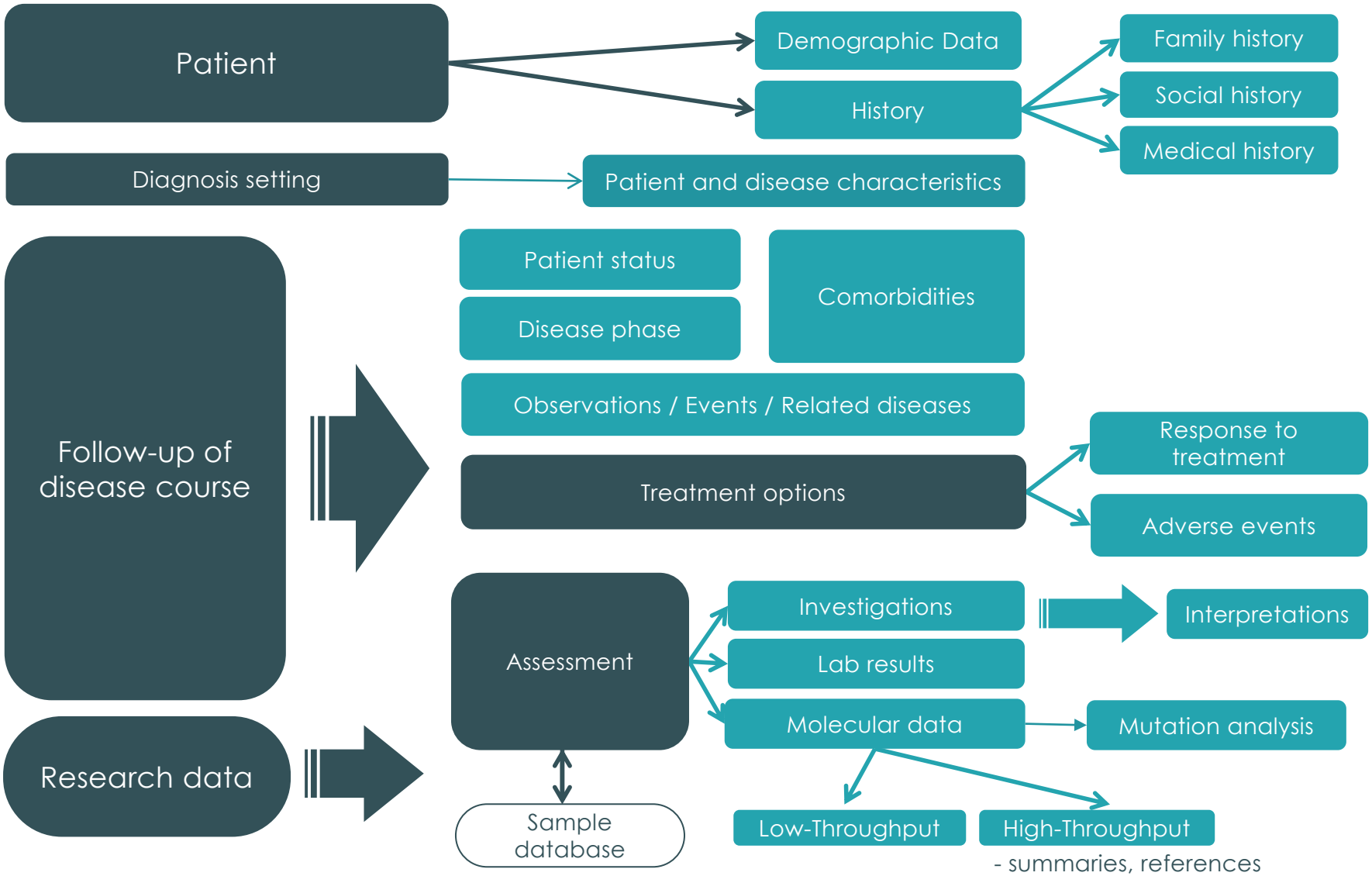
RWD management systems

DATA MODEL



RWD management systems

DATA MODEL



RWD management systems

DATA REQUIREMENTS

— Detailed description of data

Field	Data type	Data format	Unit	Value constraints
Diagnosis	list	Predefined		CLL,MBL,SLL
Date of diagnosis	date	DD/MM/YYYY		>1980
Rai stage at diagnosis	list	Predefined		0,I,II,III,IV
Binet stage at diagnosis	list	Predefined		A,B,C
Comorbidities at diagnosis	list	Predefined		YES,NO
Blood count date	date	DD/MM/YYYY		less than or equal to current date
Hb	decimal	#0.0-100.0	g/dL	0.0-100.0
WBC (x10^9/l)	decimal	#0.00-1000.00	(x10^9/l)	0.00-1000.00
Treatment status	list	Predefined		Treated,Untreated
Response to first treatment	list	Predefined		CR,CRi,Nodular PR,PR,PR with lymphocytosis,Progressive,Stable
Date of last follow-up	date	DD/MM/YYYY		less than or equal to current date
Survival status	list	Predefined		Alive,Dead

- ❑ Detailed list of data categories, entities, terminologies
- ❑ Definition of data types, data format and allowed values
- ❑ Configuration of data relationships and rules
- ❑ Consistency and integrity constraints

— User scenarios and workflows

RWD management systems

DATA REQUIREMENTS

— Consistency and integrity constraints

Required information

Field
Patient ID
Diagnosis
Date of diagnosis
Rai stage at diagnosis
Binet stage at diagnosis
Treatment status

Data **redundancy** control

Examples:

The Patient id must uniquely characterize one patient
A patient has only 1 RAI stage at diagnosis record
A patient cannot have 2 visits with the same date

Depended / Essential information

Examples:

If patient is reported DEAD, then Date of death is a required field
If patient is reported Treated, then Start date of treatment and treatment type are required fields

Additional constraints and rules for the **comparison of related fields**

Examples

Start date of treatment cannot be before Date of diagnosis
Treatment-related data are not applicable for Untreated patients
A patient with enlarged lymph nodes at time of diagnosis cannot be reported with RAI stage: 0

RWD management systems

USER MANAGEMENT

Center - based access

↓

Project - based access

↓

Lab-based access

Controlled access to data of the registration center

- ✓ **Role-based access**

  Role assignment: Restricted or Extended Access Privileges

RWD management systems

USER MANAGEMENT

User role examples of **center-based** access

Principal investigator of center (PI)

Overall access, review registration and management procedures

Data administrator (DA)

Overall read & write privileges

Data moderator (DM)

Restricted read & write permissions to selected data categories

Read - only user

Restricted read permissions to selected features

Role-assignment example

Cytogenetics-lab User

Access only to cytogenetic data

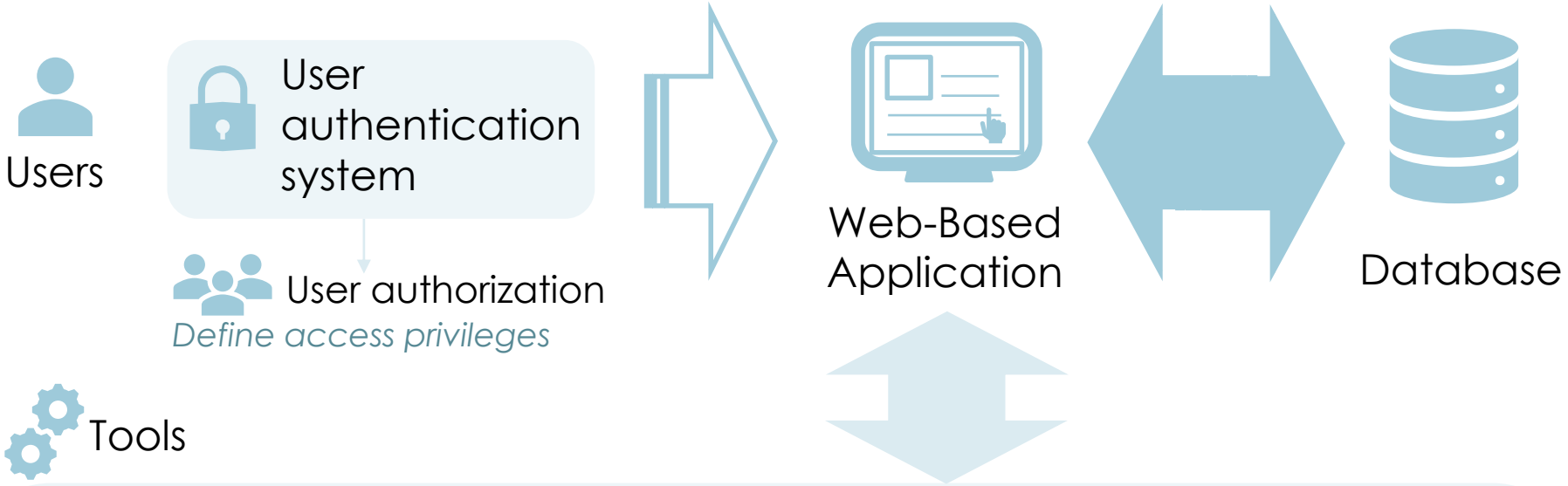
User role example of **project-based** access

General PI

Read-only access to data from all centers in a project

RWD management systems

SYSTEM DESIGN



Data management tools

- Data registration
  Data update
- Data import
- Data validation

Data retrieval tools

- Search
- Filter
- Preview
- Download

Data analysis tools

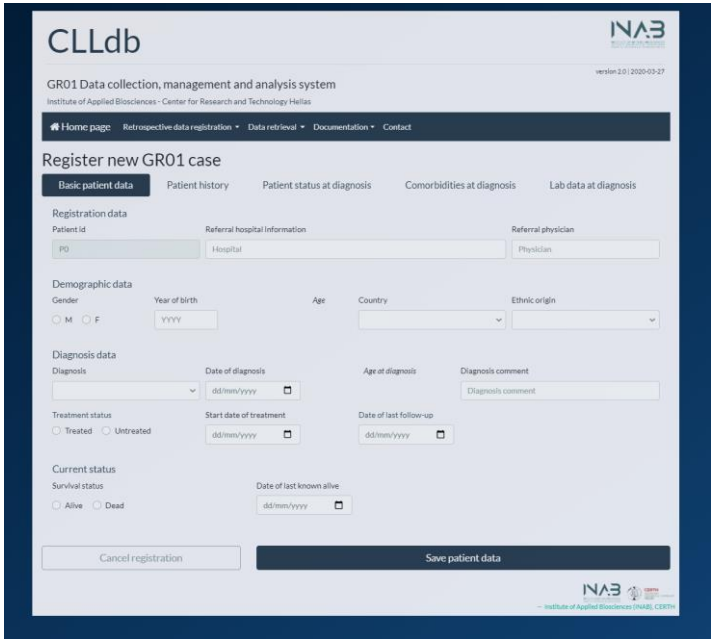
- Visualization and Export modules

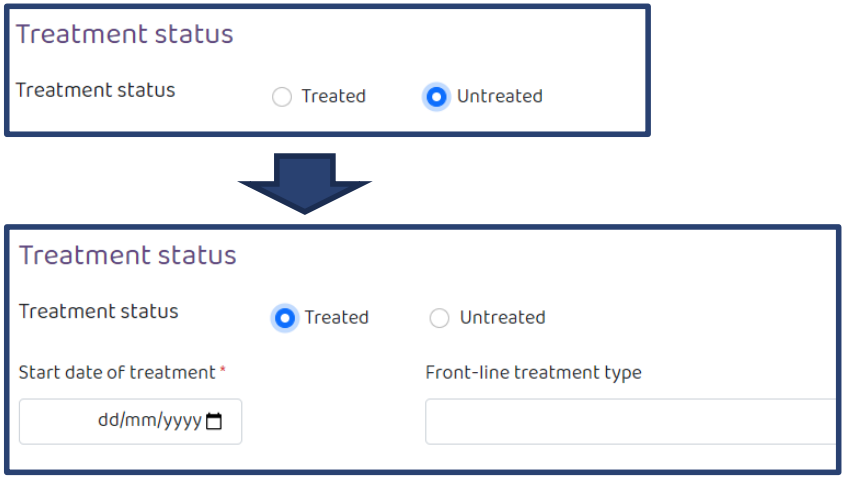
RWD management systems

DATA COLLECTION METHODS

Real-time registration systems

A. Online web forms for **prospective data** registration





✓ Organized and Dynamically adjusted according to selected criteria

RWD management systems

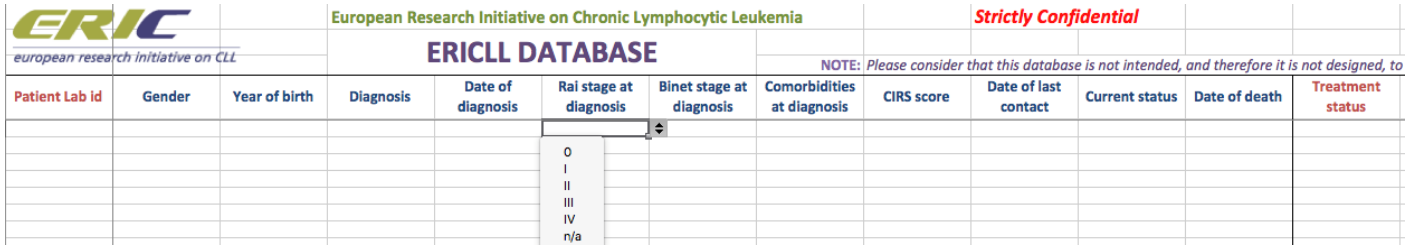
DATA COLLECTION METHODS

Real-time registration systems

facilitate data collection in a retrospective way

B. Retrospective data registration tools

specifically designed **template registration files** according to requirements


ERIC
european research initiative on CLL

European Research Initiative on Chronic Lymphocytic Leukemia

ERICLL DATABASE

Strictly Confidential

NOTE: Please consider that this database is not intended, and therefore it is not designed, to

Patient Lab id	Gender	Year of birth	Diagnosis	Date of diagnosis	Rai stage at diagnosis	Binet stage at diagnosis	Comorbidities at diagnosis	CIRS score	Date of last contact	Current status	Date of death	Treatment status

0
I
II
III
IV
n/a

** adjusted to local necessities and project-specific data collection*

✓ Protected sheets, including data validation
- *warnings, notes, validation rules and drop down lists*

Data upload, validation and import tool

integrate multi - originated data from different sources into a central repository

RWD management systems

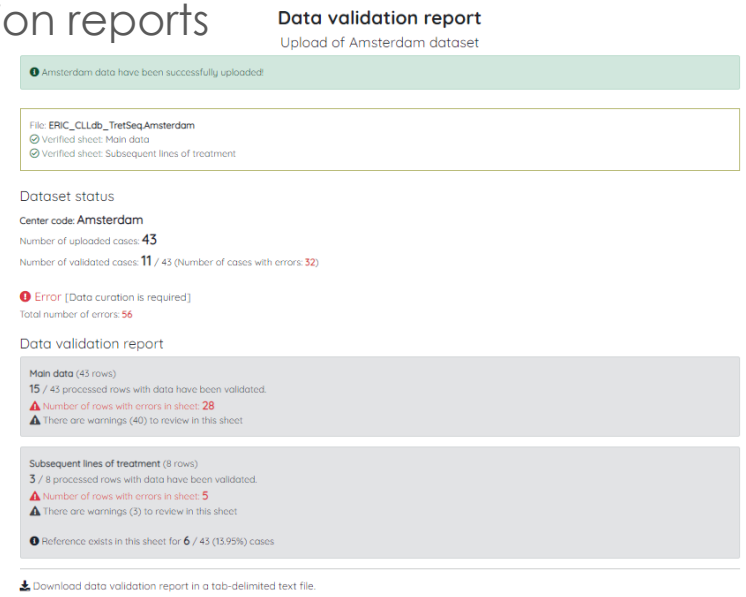
DATA COLLECTION METHODS

Data Validation and Integration

- **translation** of terms / **conversion** of data
- **validation** of data formats
- detection of **data redundancy**
- additional constraints and rules for the **comparison of related fields**

Retrospective data registration

Data validation procedure → Data validation reports



Data validation report

Upload of Amsterdam dataset

Amsterdam data have been successfully uploaded!

File: ERIC_CLLdb_TretSeq.Amsterdam
Verified sheet: Main data
Verified sheet: Subsequent lines of treatment

Dataset status

Center code: Amsterdam

Number of uploaded cases: 43

Number of validated cases: 11 / 43 (Number of cases with errors: 32)

Error [Data curation is required]

Total number of errors: 56

Data validation report

Main data (43 rows)
15 / 43 processed rows with data have been validated.
Number of rows with errors in sheet: 28
There are warnings (40) to review in this sheet

Subsequent lines of treatment (8 rows)
3 / 8 processed rows with data have been validated.
Number of rows with errors in sheet: 5
There are warnings (3) to review in this sheet

Reference exists in this sheet for 6 / 43 (13.95%) cases

Download data validation report in a tab-delimited text file.

RWD management systems

DATA COLLECTION METHODS

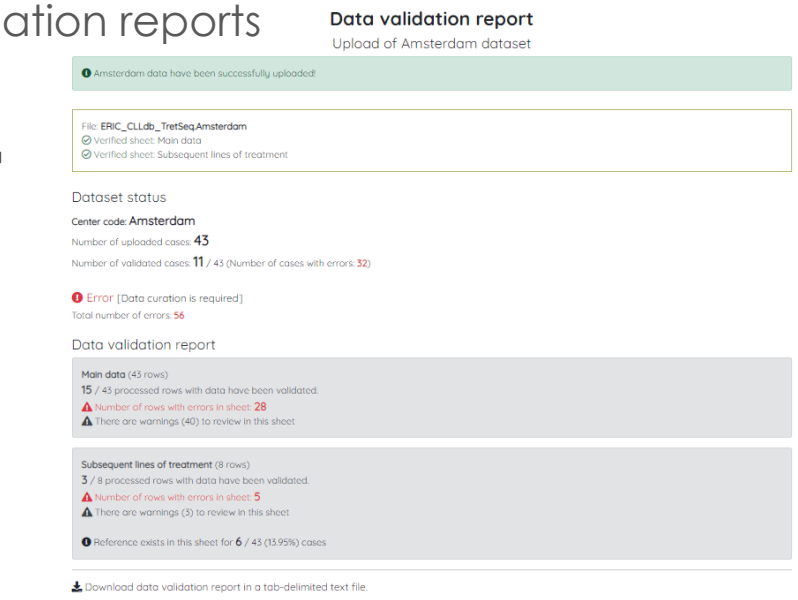
Data Validation and Integration

- **translation** of terms / **conversion** of data
- **validation** of data formats
- detection of **data redundancy**
- additional constraints and rules for the **comparison of related fields**

Retrospective data registration

Data validation procedure → Data validation reports

Corrected dataset → **Re-upload**



RWD management systems

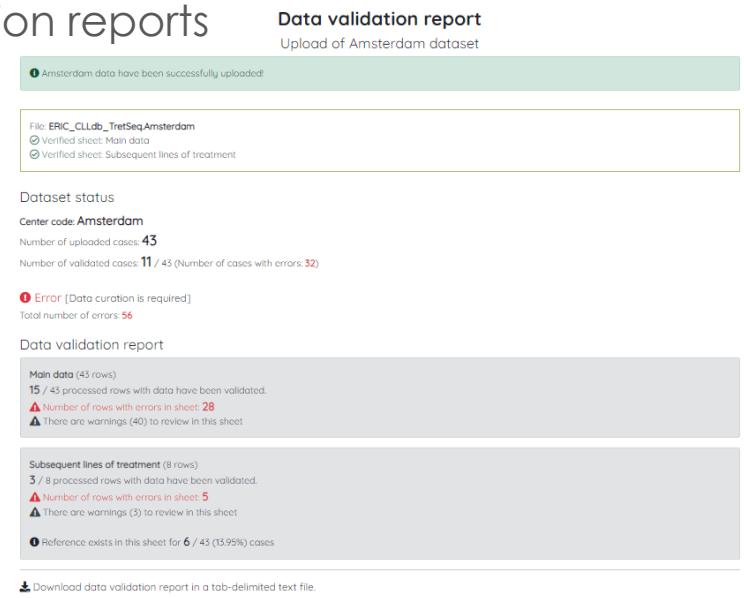
DATA COLLECTION METHODS

Data Validation and Integration

- **translation** of terms / **conversion** of data
- **validation** of data formats
- detection of **data redundancy**
- additional constraints and rules for the **comparison of related fields**

Retrospective data registration

Data validation procedure → Data validation reports



↓ Data curation

Data organization
Data homogenization

↓

Data import

PostgreSQL Central Database

RWD management systems

DATA RETRIEVAL AND ANALYSIS

1. data selection
   - **Query tools** supporting dynamic definition of selection filters.

2. data retrieval
   - Including data visualization modules

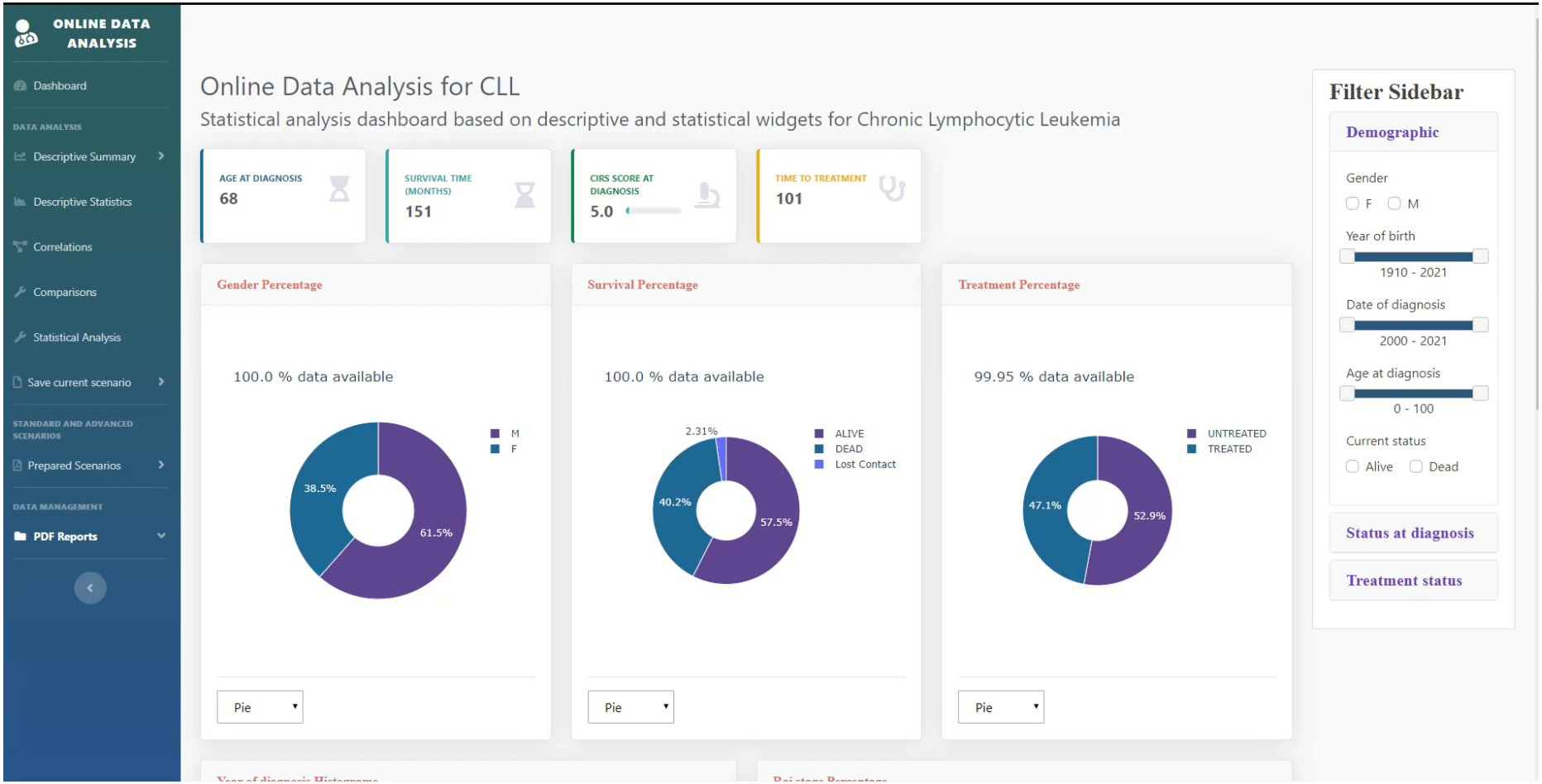
3. data export options

RWD management systems

DATA ANALYTICS

A web-based application for online statistical analysis

✓ Project-based configuration

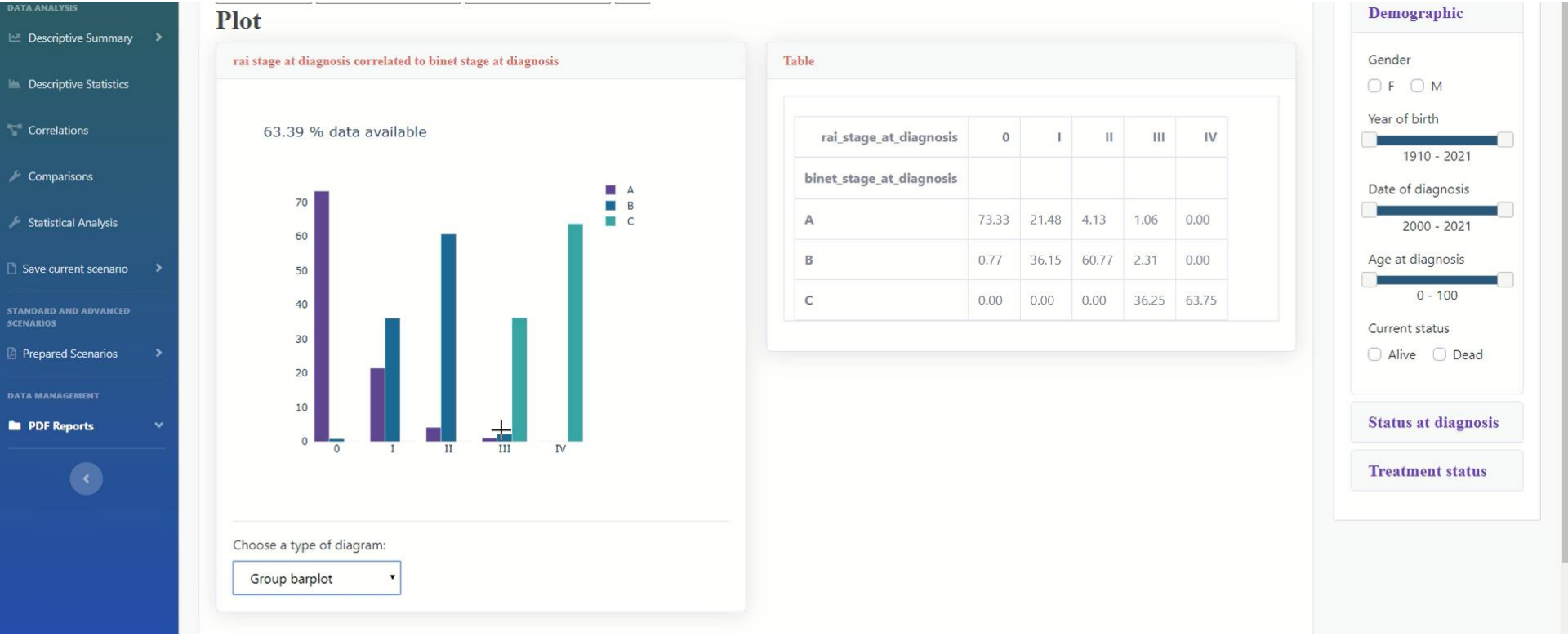


RWD management systems

DATA ANALYTICS

A web-based application for online statistical analysis

Data correlations



RWD management systems

ERIC LL database

Collection of prospective and retrospective clinical and biological data
from patients with **Chronic Lymphocytic Leukemia**
at the time of diagnosis and follow-up on a **project basis**

A project of **ERIC, the European Research Initiative on CLL**, a Scientific Working Group (SWG) of the **European Hematology Association (EHA)** aimed at improved management of CLL through collaborative research

http://www.ericll.org/

ERIC
european research initiative on CLL

Supervised by INAB

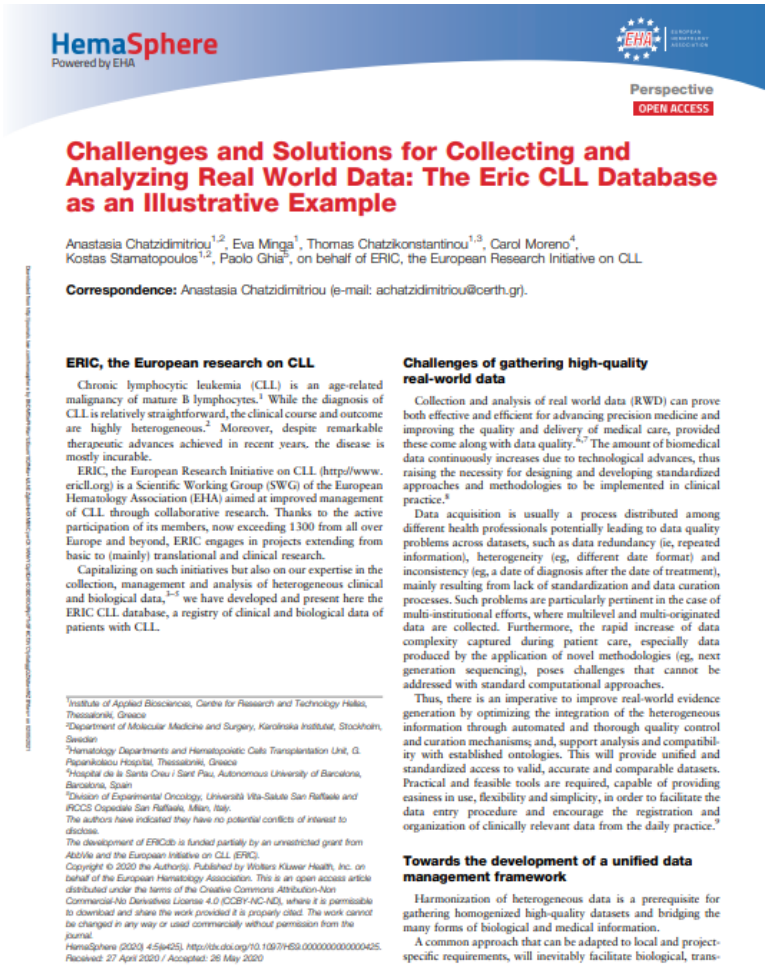
RWD management systems

european research initiative on CLL

THE ERIC CLL DATABASE

a large-scale initiative aimed at addressing the outstanding basic, translational and clinical research questions in CLL

Current status



HemaSphere
Powered by EHA

Perspective
OPEN ACCESS

Challenges and Solutions for Collecting and Analyzing Real World Data: The Eric CLL Database as an Illustrative Example

Anastasia Chatzidimitriou[1,2], Eva Minga[1], Thomas Chatzikonstantinou[1,3], Carol Moreno[4], Kostas Stamatopoulos[1,2], Paolo Ghia[5], on behalf of ERIC, the European Research Initiative on CLL

Correspondence: Anastasia Chatzidimitriou (e-mail: achatzidimitriou@certh.gr).

ERIC, the European research on CLL

Chronic lymphocytic leukemia (CLL) is an age-related malignancy of mature B lymphocytes.[1] While the diagnosis of CLL is relatively straightforward, the clinical course and outcome are highly heterogeneous.[2] Moreover, despite remarkable therapeutic advances achieved in recent years, the disease is mostly incurable.

ERIC, the European Research Initiative on CLL (http://www.ericll.org) is a Scientific Working Group (SWG) of the European Hematology Association (EHA) aimed at improved management of CLL through collaborative research. Thanks to the active participation of its members, now exceeding 1300 from all over Europe and beyond, ERIC engages in projects extending from basic to (mainly) translational and clinical research.

Capitalizing on such initiatives but also on our expertise in the collection, management and analysis of heterogeneous clinical and biological data,[3–5] we have developed and present here the ERIC CLL database, a registry of clinical and biological data of patients with CLL.

[1]Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece
[2]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden
[3]Hematology Departments and Hematopoietic Cells Transplantation Unit, G. Papanikolaou Hospital, Thessaloniki, Greece
[4]Hospital de la Santa Creu i Sant Pau, Autonomous University of Barcelona, Barcelona, Spain
[5]Division of Experimental Oncology, Università Vita-Salute San Raffaele and IRCCS Ospedale San Raffaele, Milan, Italy.
The authors have indicated they have no potential conflicts of interest to disclose.
The development of ERICdb is funded partially by an unrestricted grant from AbbVie and the European Initiative on CLL (ERIC).

HemaSphere (2020) 4:5(e425). http://dx.doi.org/10.1097/HS9.0000000000000425.
Received: 27 April 2020 / Accepted: 26 May 2020

Challenges of gathering high-quality real-world data

Collection and analysis of real world data (RWD) can prove both effective and efficient for advancing precision medicine and improving the quality and delivery of medical care, provided these come along with data quality.[6,7] The amount of biomedical data continuously increases due to technological advances, thus raising the necessity for designing and developing standardized approaches and methodologies to be implemented in clinical practice.[8]

Data acquisition is usually a process distributed among different health professionals potentially leading to data quality problems across datasets, such as data redundancy (ie, repeated information), heterogeneity (eg, different date format) and inconsistency (eg, a date of diagnosis after the date of treatment), mainly resulting from lack of standardization and data curation processes. Such problems are particularly pertinent in the case of multi-institutional efforts, where multilevel and multi-originated data are collected. Furthermore, the rapid increase of data complexity captured during patient care, especially data produced by the application of novel methodologies (eg, next generation sequencing), poses challenges that cannot be addressed with standard computational approaches.

Thus, there is an imperative to improve real-world evidence generation by optimizing the integration of the heterogeneous information through automated and thorough quality control and curation mechanisms; and, support analysis and compatibility with established ontologies. This will provide unified and standardized access to valid, accurate and comparable datasets. Practical and feasible tools are required, capable of providing easiness in use, flexibility and simplicity, in order to facilitate the data entry procedure and encourage the registration and organization of clinically relevant data from the daily practice.[9]

Towards the development of a unified data management framework

Harmonization of heterogeneous data is a prerequisite for gathering homogenized high-quality datasets and bridging the many forms of biological and medical information.

A common approach that can be adapted to local and project-specific requirements, will inevitably facilitate biological, trans-

Chatzidimitriou A, Minga E, Chatzikonstantinou T, Moreno C, Stamatopoulos K, Ghia P, on behalf of ERIC, the European Research Initiative on CLL. **Challenges and solutions for collecting and analyzing Real World Data: The ERIC CLL Database as an illustrative example.** HemaSphere: October 2020 - Volume 4 - Issue 5 - p e425 doi: 10.1097/HS9.0000000000000425.

RWD management systems

THE ERIC CLL DATABASE

Online platform for project data management and analysis

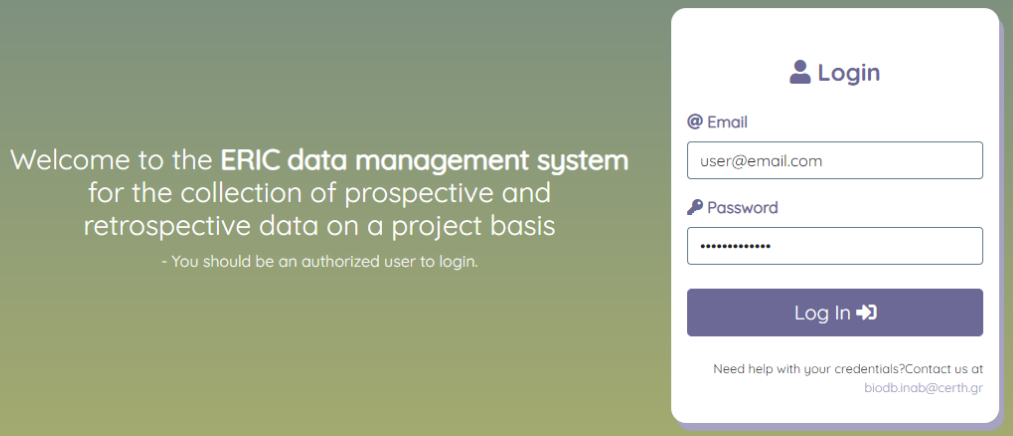
ERIC european research initiative on CLL

ERIC LL database

Registry of clinical and biological data of patients with Chronic Lymphocytic Leukemia

- A project of ERIC, the European Research Initiative on CLL -

www.ericll.org



— Developed and hosted by Institute of Applied Biosciences (INAB) | Centre for Research and Technology Hellas (CERTH)

INAB

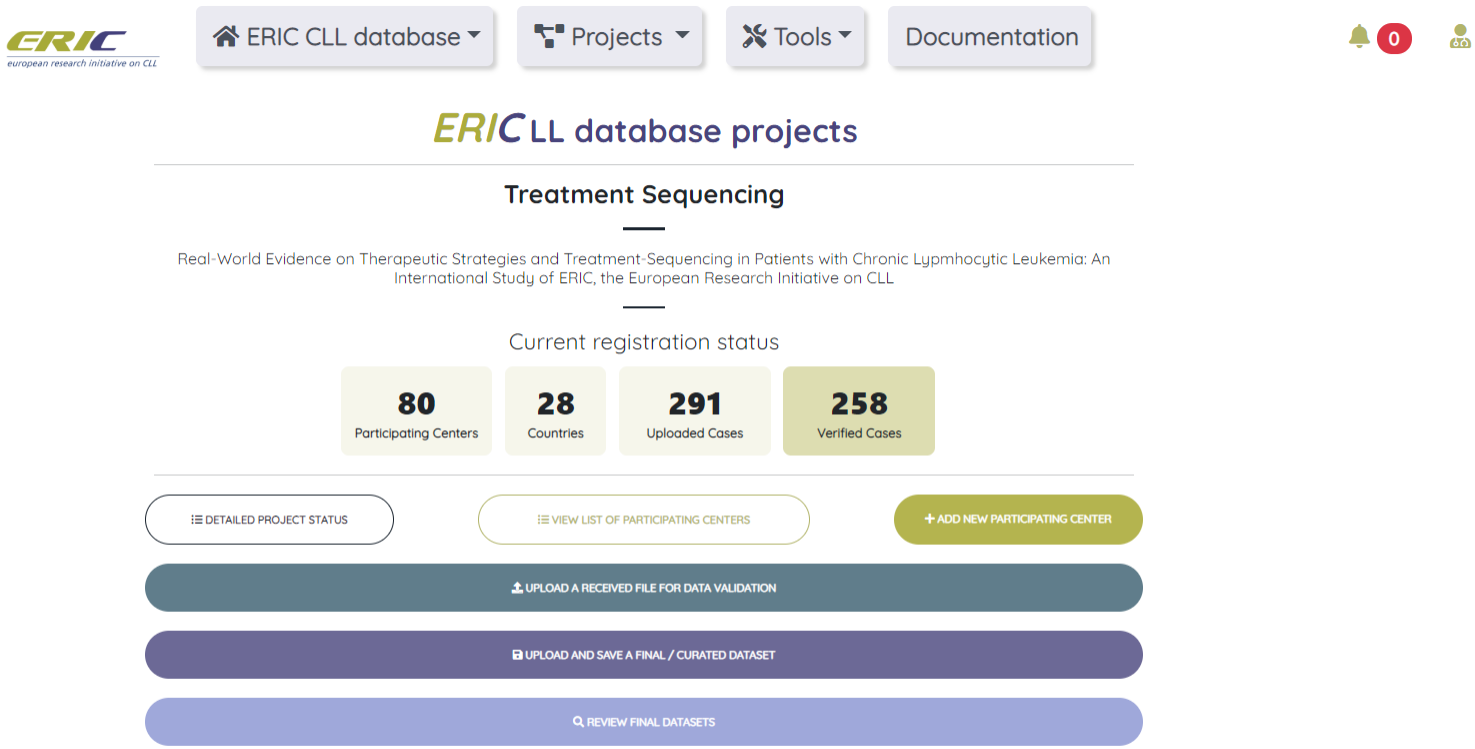
CERTH CENTRE FOR RESEARCH & TECHNOLOGY HELLAS

RWD management systems

THE ERIC CLL DATABASE

Online platform for project data management and analysis

Project-based data collection



— Developed and hosted by Institute of Applied Biosciences (INAB) | Centre for Research and Technology Hellas (CERTH)

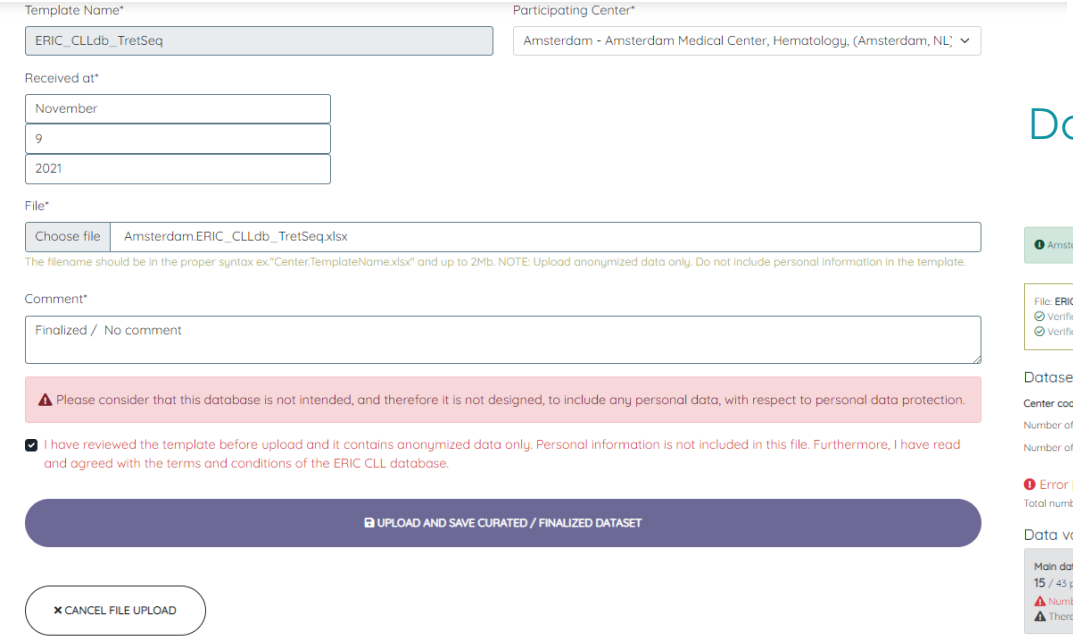
RWD management systems

THE ERIC CLL DATABASE

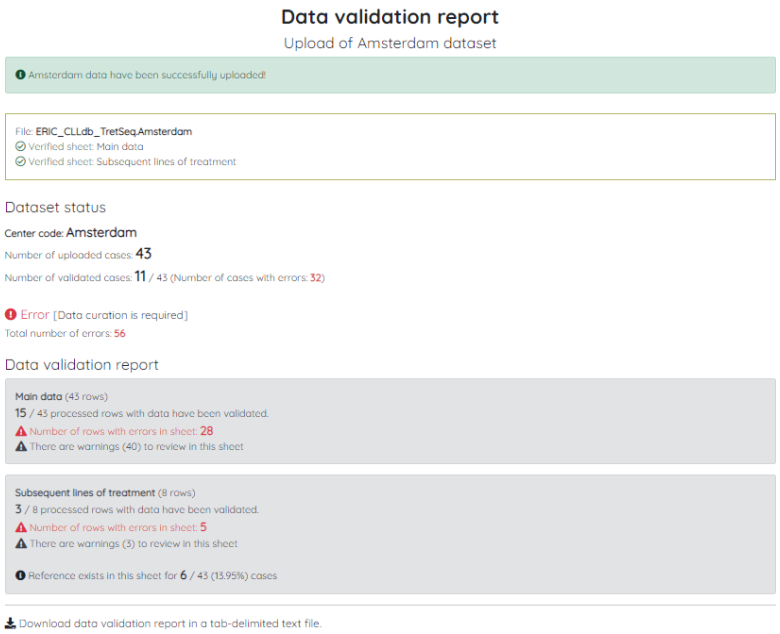
Online platform for project data management and analysis

Upload tool

Upload form



Data validation report

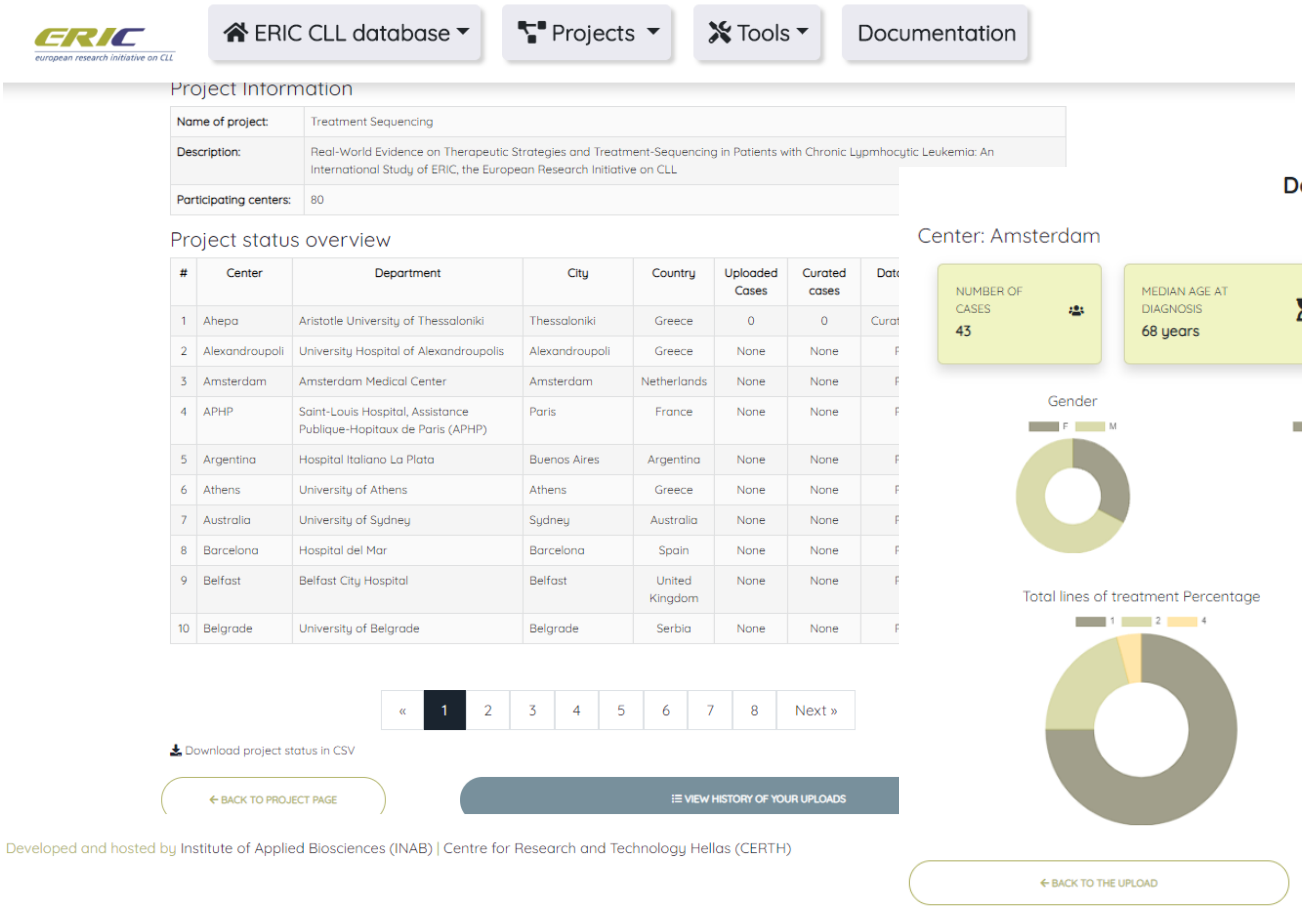


RWD management systems

THE ERIC CLL DATABASE

Online platform for project data management and analysis

Project overview



INSTITUTE OF APPLIED BIOSCIENCES
ΙΝΣΤΙΤΟΥΤΟ ΕΦΑΡΜΟΣΜΕΝΩΝ ΒΙΟΕΠΙΣΤΗΜΩΝ
CENTRE for RECEARCH and TECHNOLOGY-HELLAS