

The logo consists of the letters 'INA3' in a bold, dark teal, sans-serif font. The 'I' is a simple vertical bar. The 'N' is formed by two diagonal strokes meeting at a point. The 'A' is formed by two diagonal strokes meeting at a point. The '3' is a stylized, rounded numeral.

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

# 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

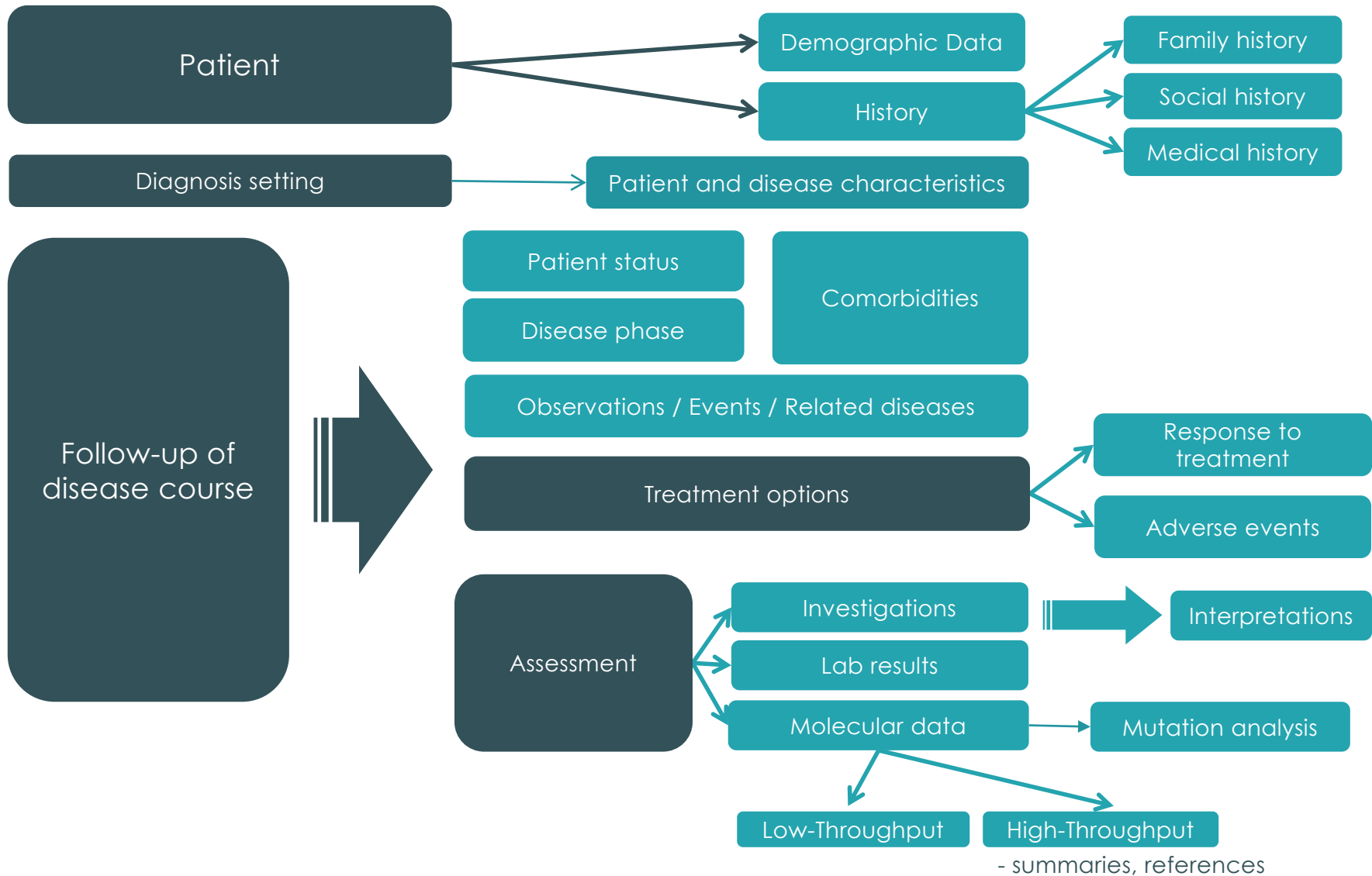
DATA MODEL

---

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

# RWD management systems

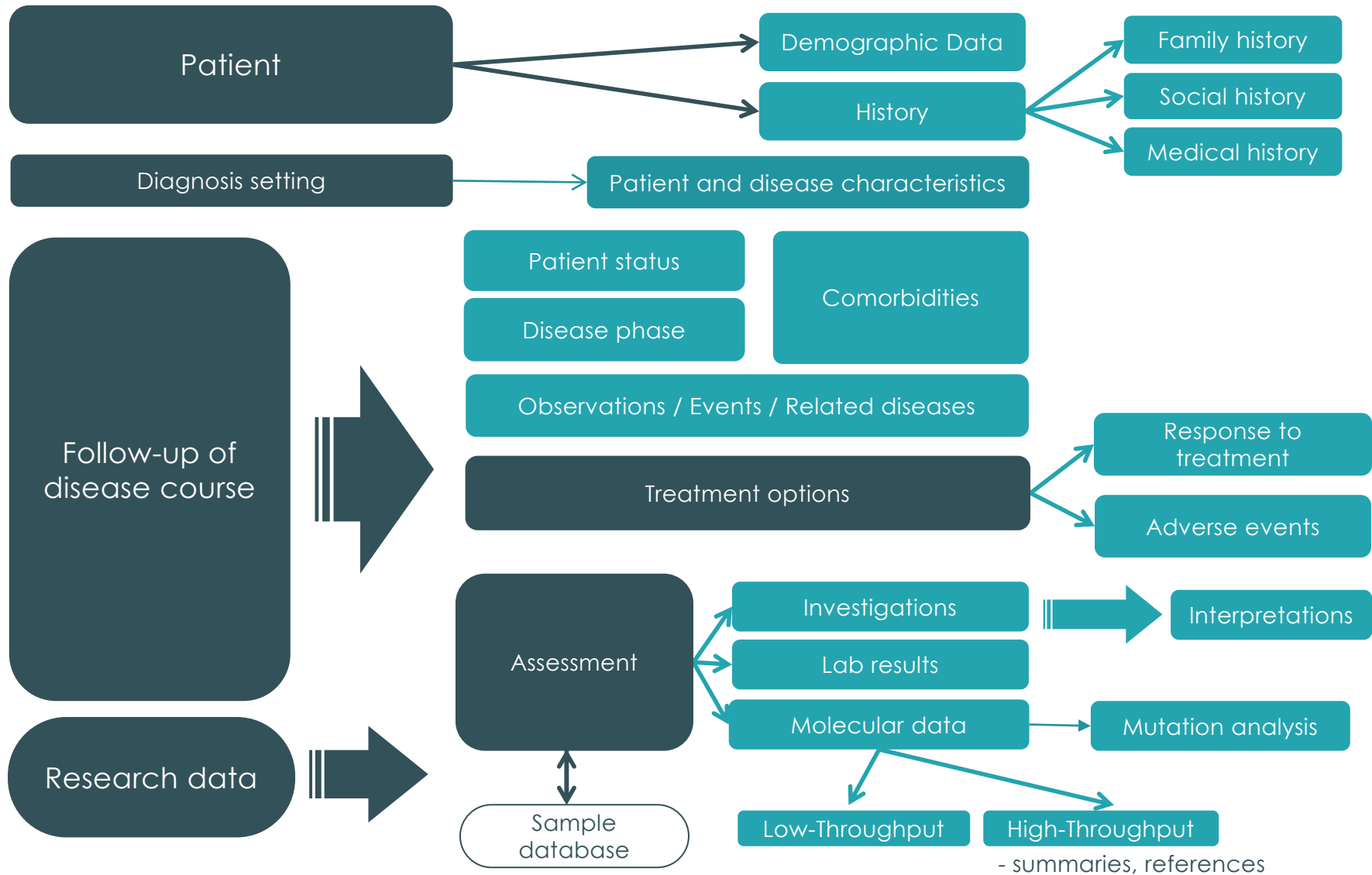
## DATA MODEL





# 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 <sup>9</sup> /l)	decimal	#0.00-1000.00	(x10 <sup>9</sup> /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

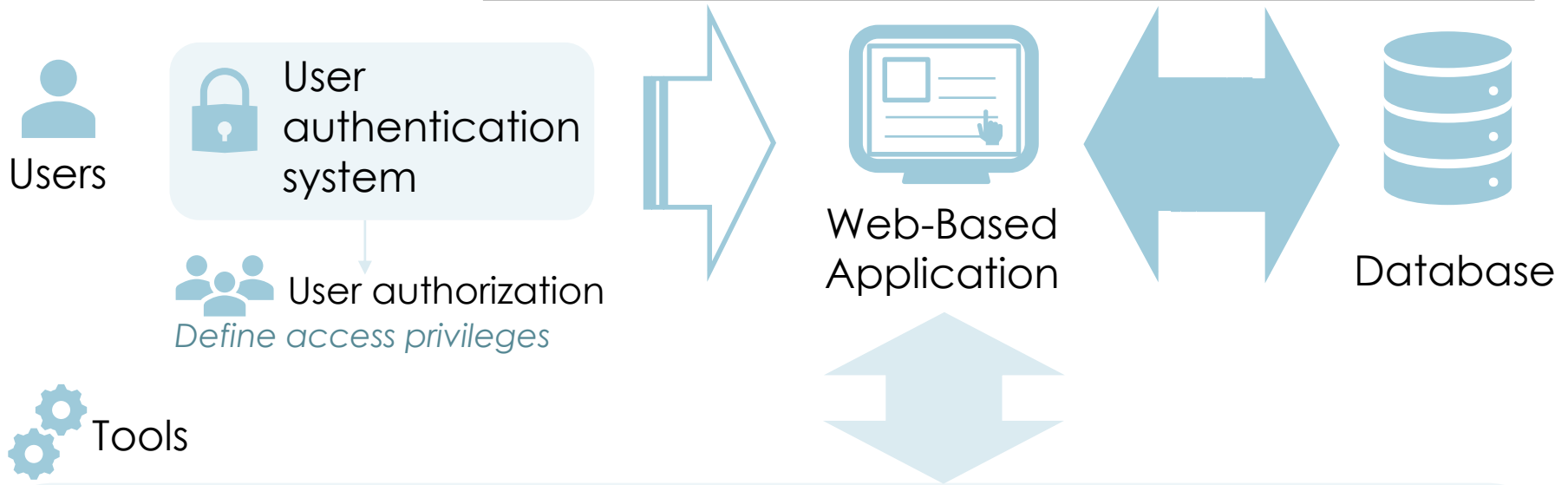
**User role** example of **project-based** access

### **General PI**

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





# RWD management systems

## SYSTEM DESIGN







## Tools

### 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

The screenshot shows the CLLdb web interface for registering a new GR01 case. The form is organized into several sections: Registration data (Patient ID, Referral hospital information, Referral physician), Demographic data (Gender, Year of birth, Age, Country, Ethnic origin), Diagnosis data (Diagnosis, Date of diagnosis, Age at diagnosis, Diagnosis comment), Treatment status (Treated/Untreated), and Current status (Survival status, Date of last known alive). A 'Save patient data' button is visible at the bottom right.

Treatment status  
Treatment status  Treated  Untreated



Treatment status  
Treatment status  Treated  Untreated  
Start date of treatment \*   
Front-line treatment type

- ✓ 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 Chronic Lymphocytic Leukemia						Strictly Confidential				
european research initiative on CLL		ERICLL DATABASE						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\_TretSeqAmsterdam

☑ 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 (0 rows)

3 / 8 processed rows with data have been validated.

▲ Number of rows with errors in sheet: 5

▲ There are warnings (2) 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

Data validation report  
Upload of Amsterdam dataset

Re-upload



Corrected dataset

Amsterdam data have been successfully uploaded!

File: ERIC CLLdb\_TretSeqAmsterdam  
✔ 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 (0 rows)

3 / 8 processed rows with data have been validated.

⚠️ Number of rows with errors in sheet: 5

⚠️ There are warnings (2) 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

Data validation report

Upload of Amsterdam dataset



Data curation

Data organization  
Data homogenization



Data import



PostgreSQL

Central

Database

Amsterdam data have been successfully uploaded!

File: ERIC CLLdb\_TretSeqAmsterdam

- ☑ 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 (0 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 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

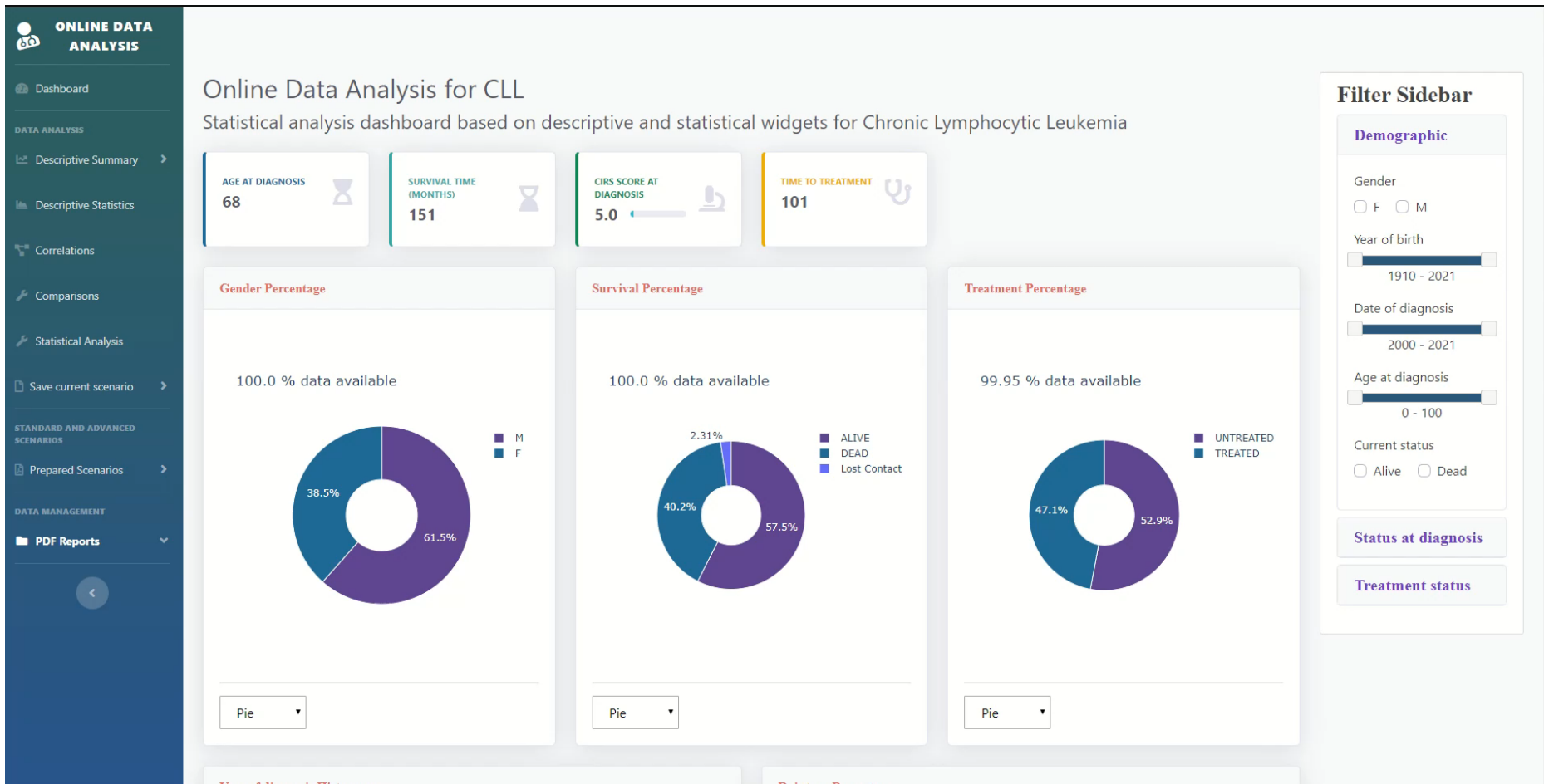


# RWD management systems

## 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

DATA ANALYSIS

- Descriptive Summary >
- Descriptive Statistics
- Correlations
- Comparisons
- Statistical Analysis
- Save current scenario >

STANDARD AND ADVANCED SCENARIOS

- Prepared Scenarios >

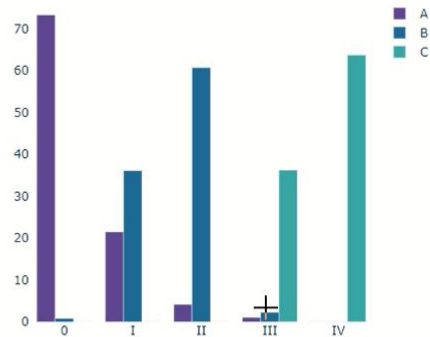
DATA MANAGEMENT

- PDF Reports ▾

#### Plot

rai stage at diagnosis correlated to binet stage at diagnosis

63.39 % data available



Choose a type of diagram:

Group barplot ▾

#### Table

rai_stage_at_diagnosis	0	I	II	III	IV
binet_stage_at_diagnosis					
A	73.33	21.48	4.13	1.06	0.00
B	0.77	36.15	60.77	2.31	0.00
C	0.00	0.00	0.00	36.25	63.75

#### Demographic

Gender

F  M

Year of birth

1910 - 2021

Date of diagnosis

2000 - 2021

Age at diagnosis

0 - 100

Current status

Alive  Dead

Status at diagnosis

Treatment status

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**

## THE ERIC CLL DATABASE

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

## Current status



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

Anastasia Chatzidimitriou<sup>1,2</sup>, Eva Minga<sup>1</sup>, Thomas Chatzikonstantinou<sup>1,3</sup>, Carol Moreno<sup>4</sup>, Kostas Stamatopoulos<sup>1,2</sup>, Paolo Ghia<sup>5</sup>, 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.<sup>1</sup> While the diagnosis of CLL is relatively straightforward, the clinical course and outcome are highly heterogeneous.<sup>2</sup> 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,<sup>3-5</sup> we have developed and present here the ERIC CLL database, a registry of clinical and biological data of patients with CLL.

### 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.<sup>6,7</sup> 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.<sup>8</sup>

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 data 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 ease of 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.<sup>9</sup>

### 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-

<sup>1</sup>Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece

<sup>2</sup>Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden

<sup>3</sup>Hematology Departments and Hematopoietic Cells Transplantation Unit, G. Papanikolaou Hospital, Thessaloniki, Greece

<sup>4</sup>Hospital de la Santa Creu i Sant Pau, Autonomous University of Barcelona, Barcelona, Spain

<sup>5</sup>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).

Copyright © 2020 the Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the European Hematology Association. This is an open access article distributed under the terms of the Creative Commons Attribution-Non

Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

HemaSphere (2020) 4:5(e425). <http://dx.doi.org/10.1097/HS9.0000000000000425>.

Received: 27 April 2020 / Accepted: 28 May 2020



# RWD management systems

## THE ERIC CLL DATABASE

---

Online platform for project data management and analysis



# ERIC<sup>CLL</sup> 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](http://www.ericll.org)


Welcome to the **ERIC data management system**  
for the collection of prospective and  
retrospective data on a project basis

- You should be an authorized user to login.

 **Login**

@ Email

 Password

**Log In** 

Need help with your credentials? Contact us at  
[biodb.inab@certh.gr](mailto:biodb.inab@certh.gr)

# RWD management systems

## THE ERIC CLL DATABASE

Online platform for project data management and analysis

### Project-based data collection

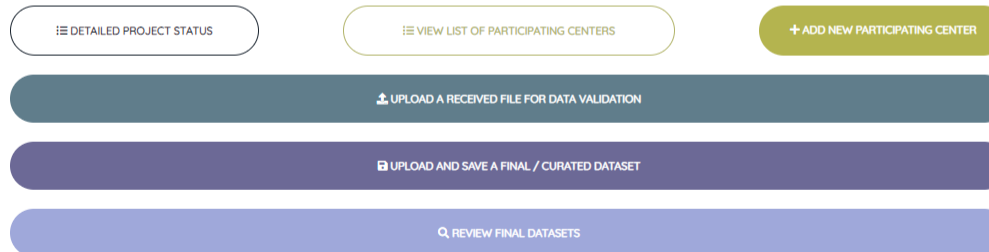
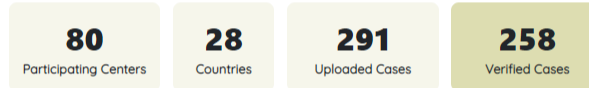


### ERICLL database projects

#### Treatment Sequencing

Real-World Evidence on Therapeutic Strategies and Treatment-Sequencing in Patients with Chronic Lymphocytic Leukemia: An International Study of ERIC, the European Research Initiative on CLL

Current registration status



# RWD management systems

## THE ERIC CLL DATABASE

Online platform for project data management and analysis

### Upload tool

#### Upload form

Template Name\*

Participating Center\*

Received at\*

File\*

The filename should be in the proper syntax ex: "Center.TemplateName.xlsx" and up to 2Mb. NOTE: Upload anonymized data only. Do not include personal information in the template.

Comment\*

**⚠ Please consider that this database is not intended, and therefore it is not designed, to include any personal data, with respect to personal data protection.**

I have reviewed the template before upload and it contains anonymized data only. Personal information is not included in this file. Furthermore, I have read and agreed with the terms and conditions of the ERIC CLL database.

#### Data validation report

##### Data validation report

Upload of Amsterdam dataset

Amsterdam data have been successfully uploaded!

File: ERIC\_CLLdb\_TretSeqAmsterdam  
✔ 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 (5) 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

## THE ERIC CLL DATABASE

Online platform for project data management and analysis

### Project overview



ERIC CLL database

Projects

Tools

Documentation

#### Project Information

Name of project:	Treatment Sequencing
Description:	Real-World Evidence on Therapeutic Strategies and Treatment-Sequencing in Patients with Chronic Lymphocytic Leukemia: An International Study of ERIC, the European Research Initiative on CLL
Participating centers:	80

#### Project status overview

#	Center	Department	City	Country	Uploaded Cases	Curated cases	Date
1	Ahepa	Aristotle University of Thessaloniki	Thessaloniki	Greece	0	0	Curat
2	Alexandroupoli	University Hospital of Alexandroupolis	Alexandroupoli	Greece	None	None	F
3	Amsterdam	Amsterdam Medical Center	Amsterdam	Netherlands	None	None	F
4	APHP	Saint-Louis Hospital, Assistance Publique-Hopitaux de Paris (APHP)	Paris	France	None	None	F
5	Argentina	Hospital Italiano La Plata	Buenos Aires	Argentina	None	None	F
6	Athens	University of Athens	Athens	Greece	None	None	F
7	Australia	University of Sydney	Sydney	Australia	None	None	F
8	Barcelona	Hospital del Mar	Barcelona	Spain	None	None	F
9	Belfast	Belfast City Hospital	Belfast	United Kingdom	None	None	F
10	Belgrade	University of Belgrade	Belgrade	Serbia	None	None	F

« 1 2 3 4 5 6 7 8 Next »

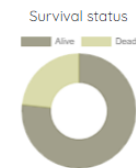
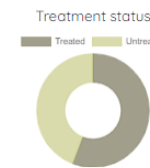
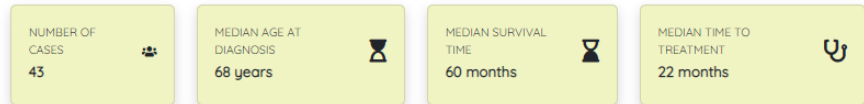
Download project status in CSV

← BACK TO PROJECT PAGE

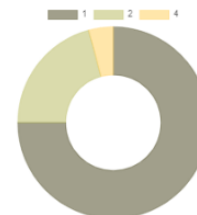
VIEW HISTORY OF YOUR UPLOADS

#### Data overview

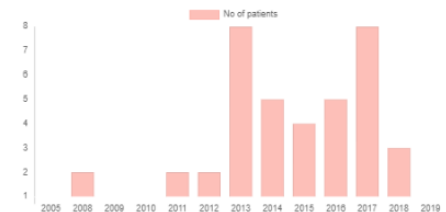
Center: Amsterdam



Total lines of treatment Percentage



No of patients diagnosed per year



← BACK TO THE UPLOAD

← BACK TO PROJECT PAGE



INSTITUTE OF APPLIED BIOSCIENCES  
ΙΝΣΤΙΤΟΥΤΟ ΕΦΑΡΜΟΣΜΕΝΩΝ ΒΙΟΠΙΣΤΗΜΩΝ  
CENTRE for RESEARCH and TECHNOLOGY-HELLAS