
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 association 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 biological, 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

4. Data Protection

EU General Data Protection Regulation (GDPR)
Technical and Organizational measurements for data protection and data security

RWD management systems

APPROACH

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
- Quality control and curation mechanisms
- Data analytics, Data mining and knowledge discovery
- Data protection, security, privacy and availability

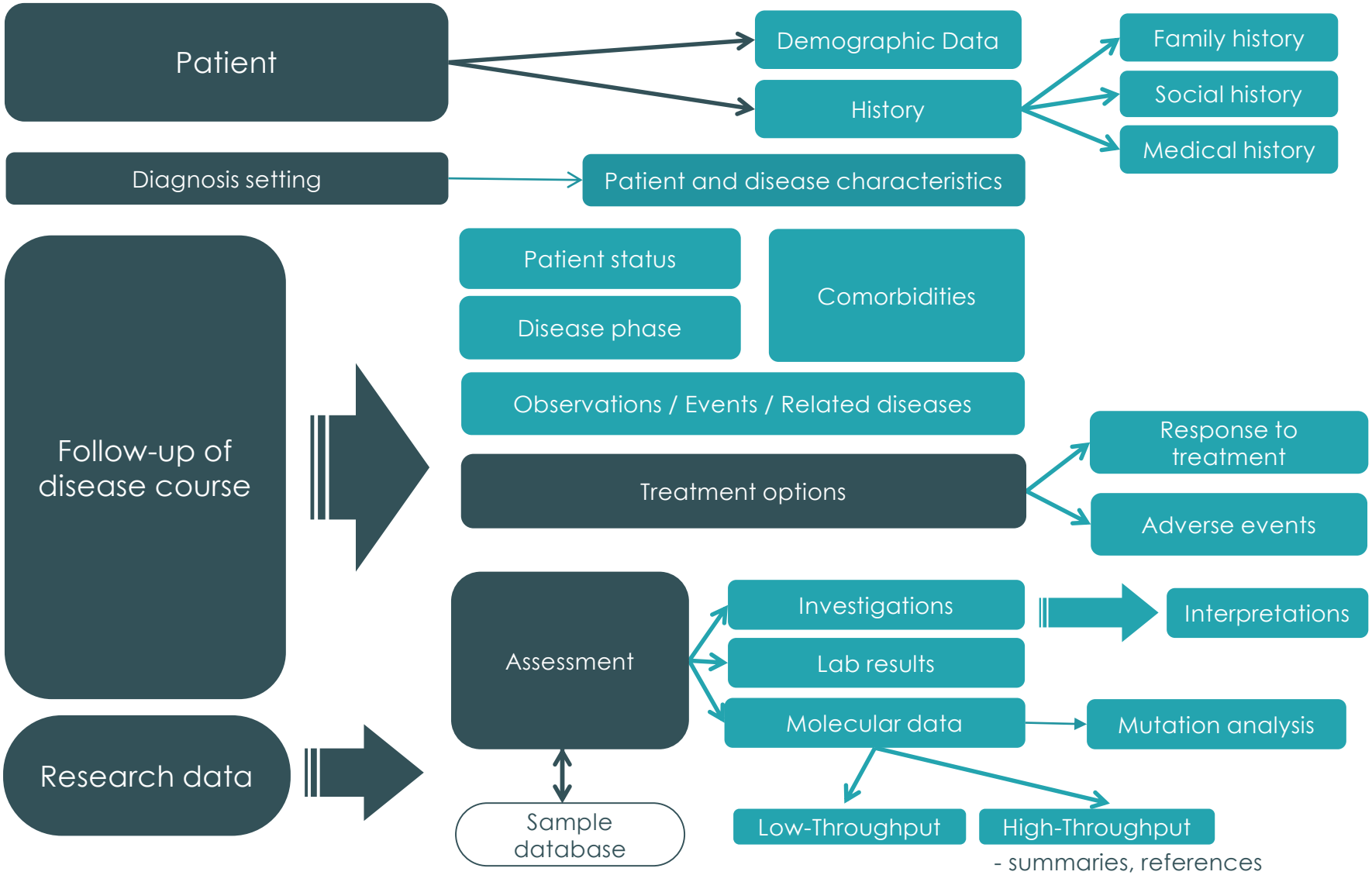
RWD management systems

DATA MODEL

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

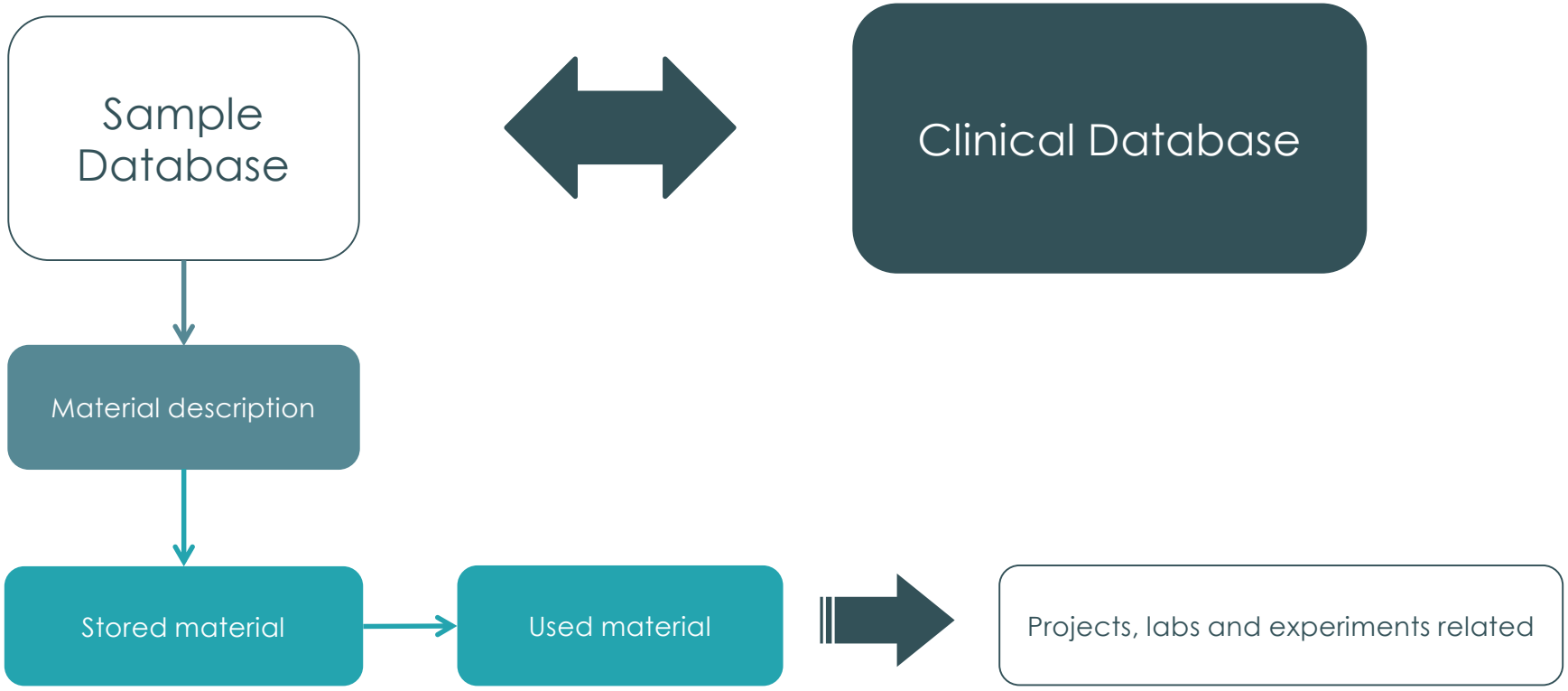
RWD management systems

DATA MODEL



RWD management systems

DATA MODEL



RWD management systems

DATA MODEL

Characteristics

- Meet the requirements for an accurate description of diagnosis, prognostic assessment and management of patients
- Adjusted to project-based requirements (extended / expanded)
- Expressing the complexity of disease

RWD management systems

DATA MODEL

Data Requirements

Collect all information about the data that will be stored in the database.

- What are the types of data and the data categories of the different data entities?
- Which fields will be included in each data category?
- What is the data type and format of each field that will be stored in the database?
- Are there specific values allowed for the defined fields?
- What are the rules and relationships among different types of data?
- Which data are required?
- How will the data be assessed?

→ User scenarios and workflows

Detailed description of data

- ❑ 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

RWD management systems

DATA MODEL

1. Detailed list of data categories, entities, terminologies

Main data

Data category	Column name
Registration data	Patient id
	Registration center
	Hospital
	Physician
Demographic data	Gender
	Year of birth
	Age
Diagnosis data	Date of diagnosis
	Age at diagnosis
	Diagnosis
	PS at time of diagnosis

Treatment data

Treatment data	Date of first treatment
	First treatment type
	Response to first treatment

Lab data

Data category	Column name
Blood test	Blood test date
	Haemoglobin (g/dL)
	WBC
	ANC
	Platelets
	Serum albumin (g/dL)
	Serum creatinine
	Alkaline phosphatase
	Alkaline phosphatase (ULN)
	Serum urea
	Serum calcium
	Peripheral Blood plasma cells %
Bone marrow test	Date of bone marrow test
	Aspirate method
	BM infiltration in aspirate
	Trephine method
	BM infiltration in trephine
	BM Flow examination

RWD management systems

DATA MODEL

2. Definition of data types, data format and allowed values

Field	Data type	Data format	Unit	Value constraints
Diagnosis	text	List		CLL,MBL,SLL
Date of diagnosis	date	DD/MM/YYYY		>1980
Rai stage at diagnosis	text	List		0,I,II,III,IV
Binet stage at diagnosis	Text	List		A,B,C
Comorbidities at diagnosis	text	List		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	Text	List		Treated,Untreated
Response to first treatment	Text	List		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	Text	List		Alive,Dead

RWD management systems

DATA MODEL

3. Configuration of data relationships and rules

Request additional information

❑ Example: Treatment decsription

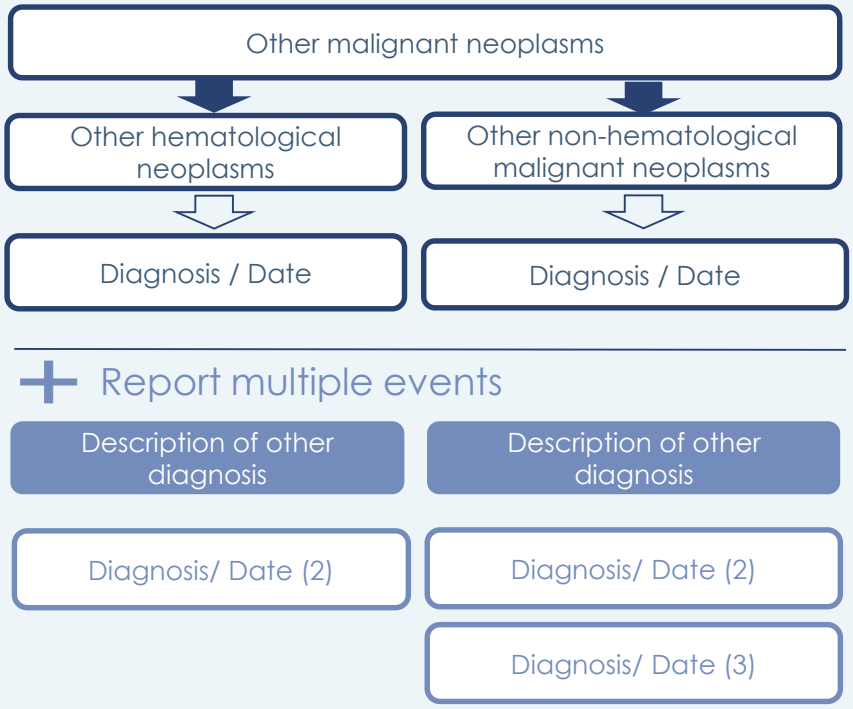
Treatment type list

Ibrutinib - single
Ibrutinib - combination
Acalabrutinib - single
Acalabrutinib - combination
Venetoclax - single
Venetoclax - combination
…
…
…
Other

→ Description of other type of first treatment (free text)

Request additional information

❑ Example: Report other malignancies



RWD management systems

DATA MODEL

4. 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 time
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

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 with original 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

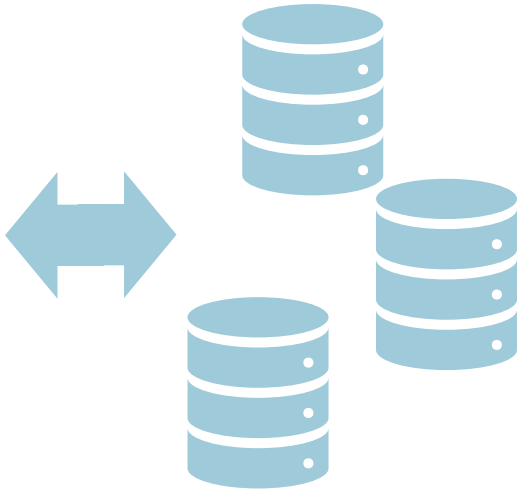
Appropriate technical and organizational measures
- data confidentiality, availability and integrity
- data protection and security

RWD management systems

SYSTEM DESIGN

Database development

Database system specifically designed for clinical and biological data



Interoperability

Main characteristics

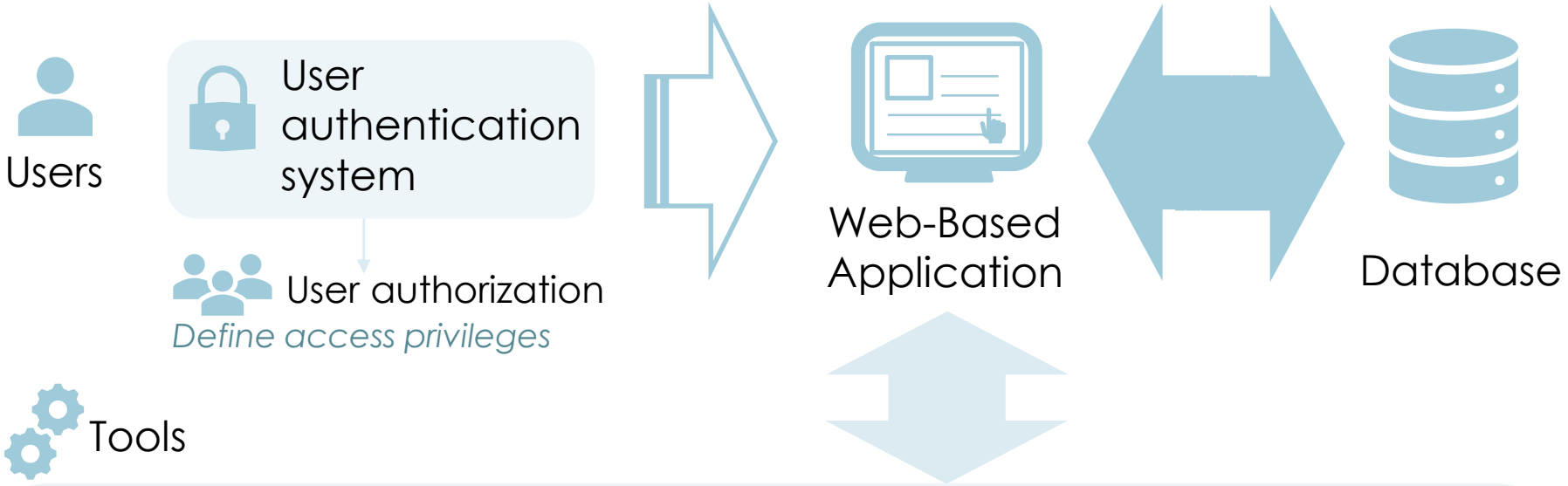
- Data Integrity
- Data Consistency
- Redundancy Control
- Data Availability
- Data Correlations

Data security

- ❑ User management
- ❑ Project Management
- ❑ Activity Logs
- ❑ Back up mechanisms
- ❑ Maintenance procedures
- ❑ Reporting system

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

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

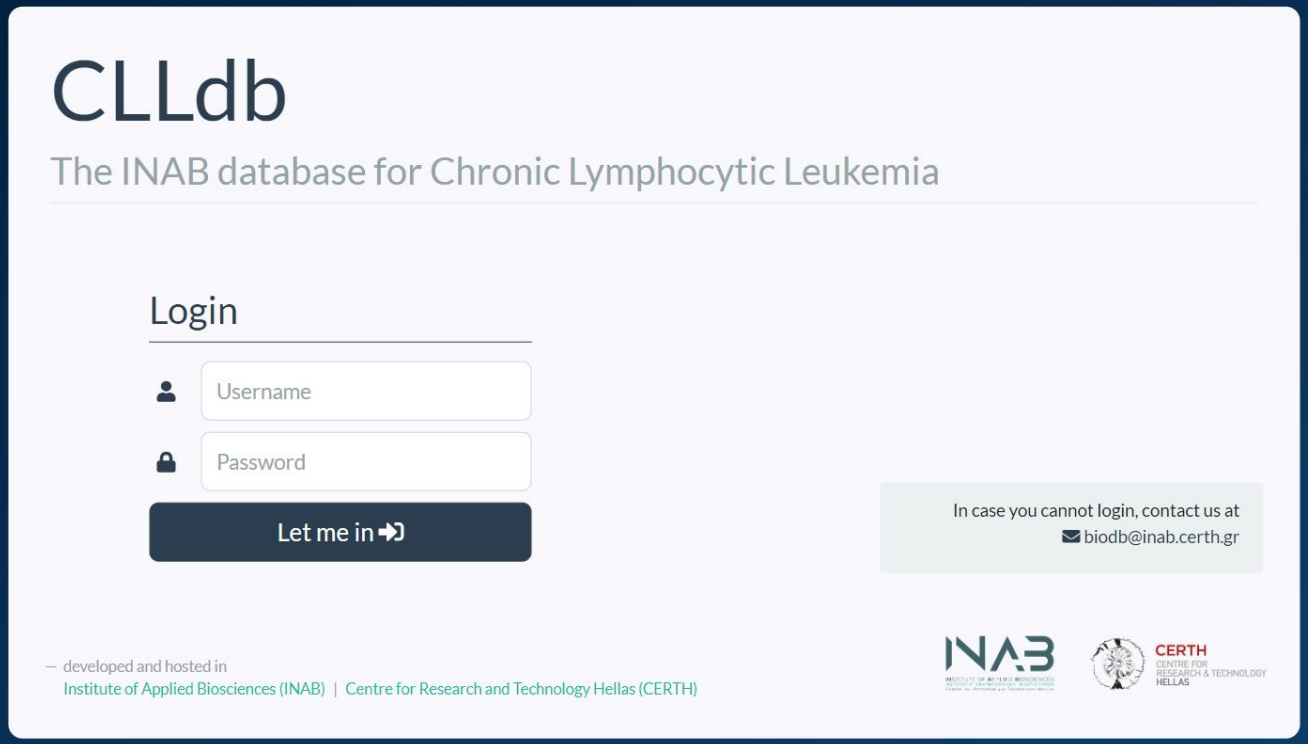
General PI

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

RWD management systems

USER MANAGEMENT

User Authentication



Strictly personal user accounts - credentials

Login Page

RWD management systems

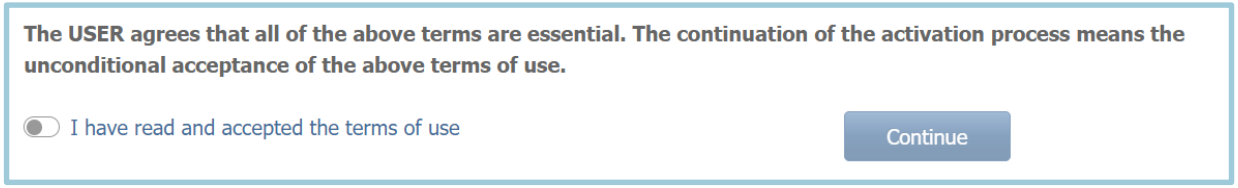
USER MANAGEMENT

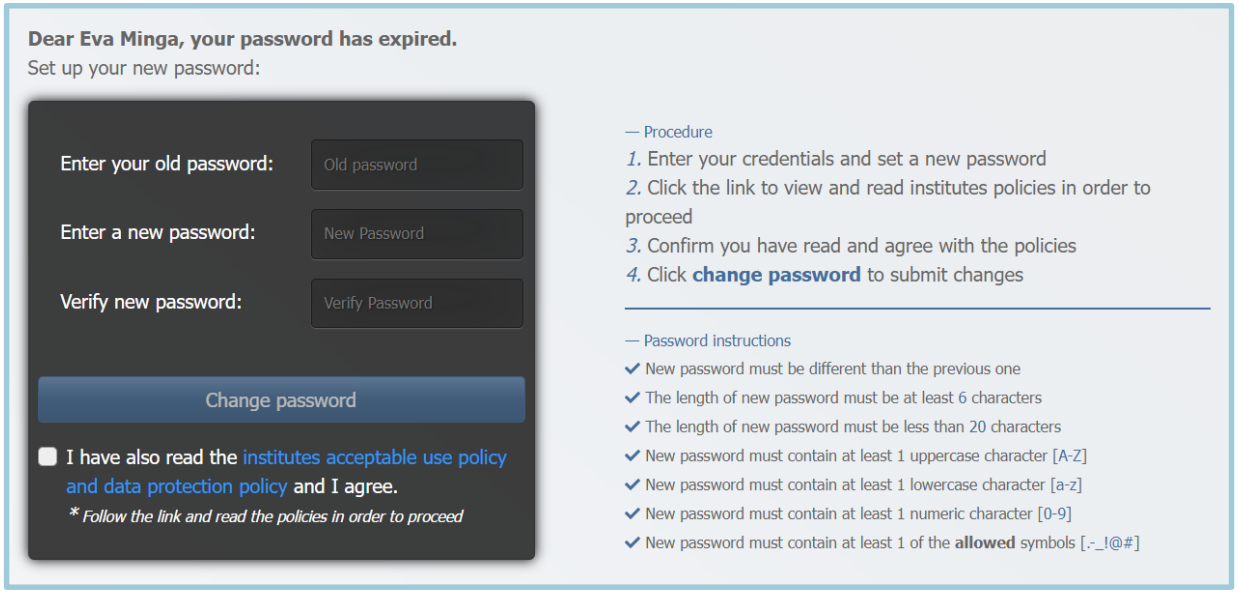
User Account Activation

Login Credentials
Username
Temporary Password

- ✓ Terms and Policies agreement
- ✓ Set up new password

Confirmation email





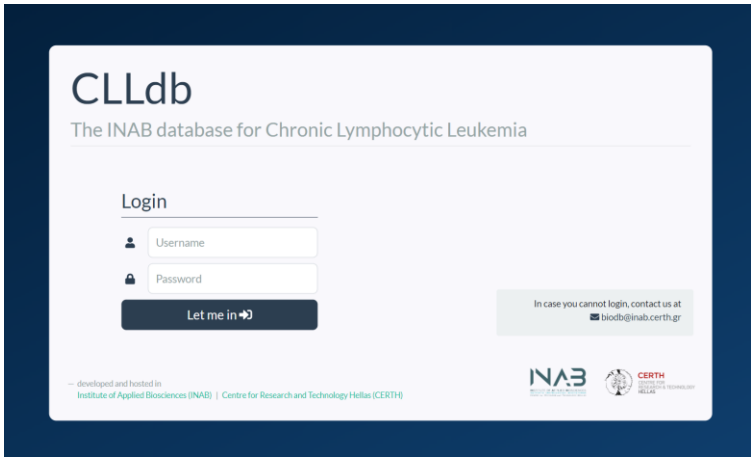
Set up Password

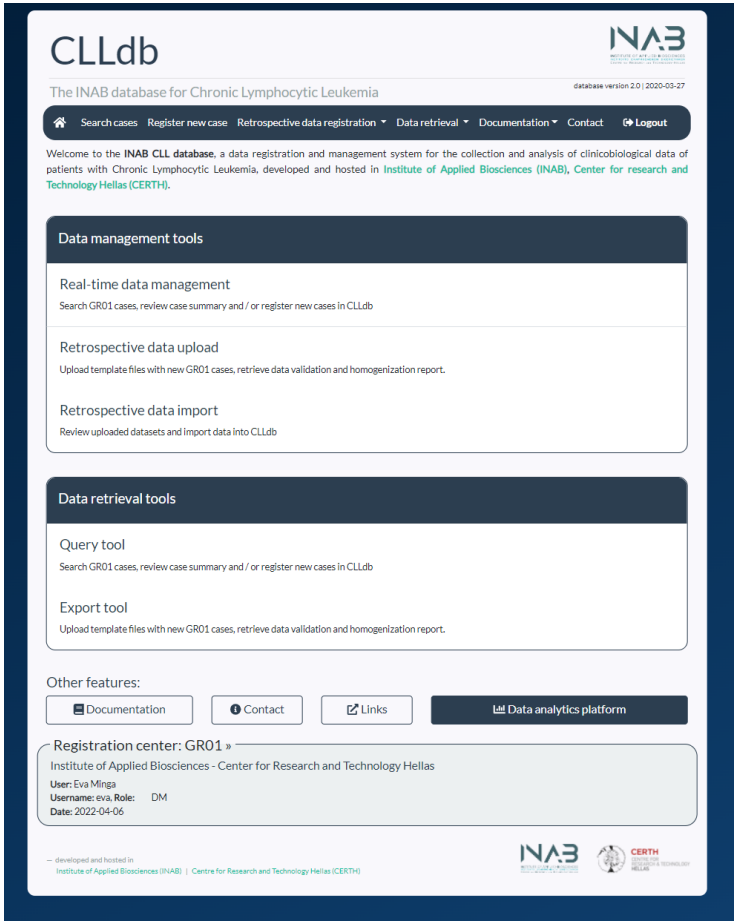
RWD management systems

USER MANAGEMENT

User Account Activation

Confirm and login to activate account:





Change of password
is periodically required

RWD management systems

DATA COLLECTION METHODS

Real-time registration systems

Prospective data registration

Online web forms

- ✓ Register new record
- ✓ Search for existing record
- ✓ Overview record data
- ✓ Update record

Retrospective data registration

Data registration tools

- ✓ Massive import of patient retrospective data
- ✓ Transform and upload data exports from other sources
- ✓ Data homogenization and integration

✓ Available for edit online

Integration and validation mechanisms

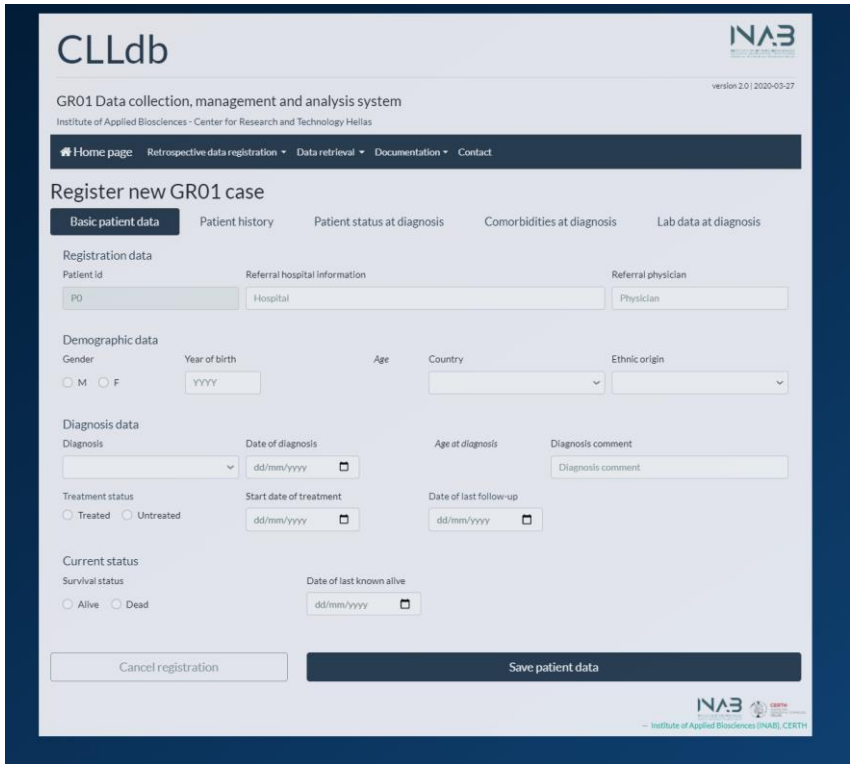
elimination of data inconsistency and redundancy
enhanced data management and organization

RWD management systems

DATA COLLECTION METHODS

Real-time registration systems

A. Online web forms for prospective data registration



- ✓ Organize data in order to facilitate data registration and improve data management
  - separate categories
  - multiple timepoints
- ✓ Create different types of forms for controlled access to data management according to user roles and access privileges
- ✓ Design dynamically adjusted forms based on selected criteria

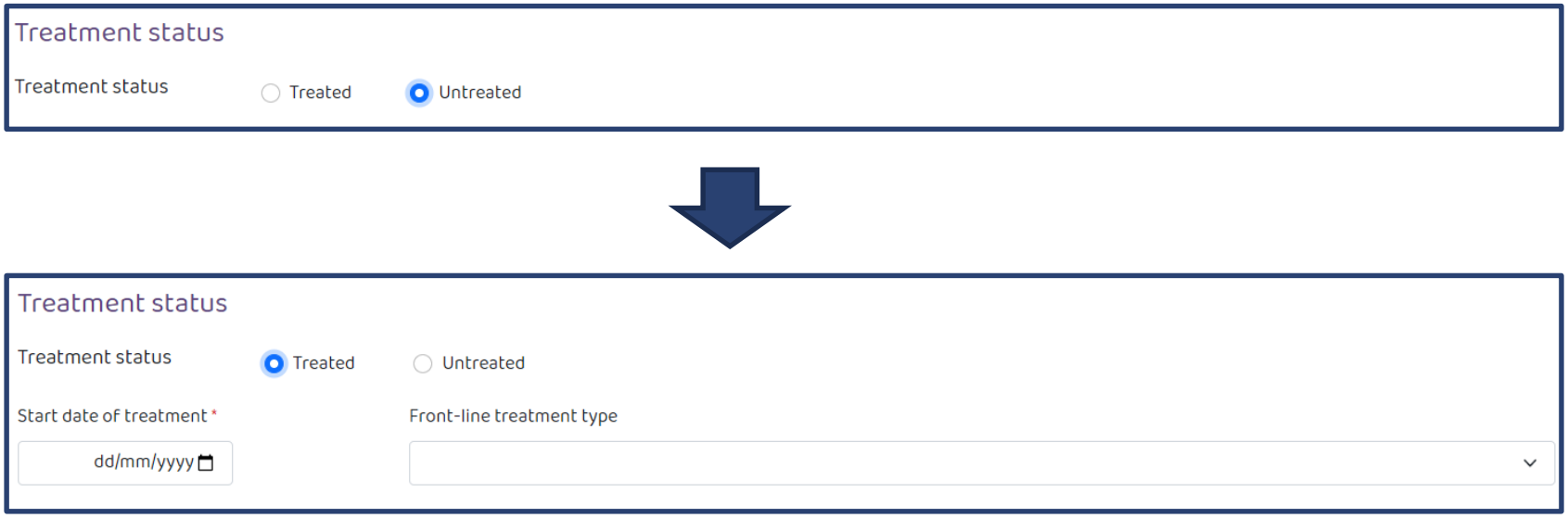
RWD management systems

DATA COLLECTION METHODS

Real-time registration systems

A. Online web forms for prospective data registration

Data that are requested only under specific conditions:



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**

RWD management systems

DATA COLLECTION METHODS

Data Validation and Integration

Examples of reported errors

Data format errors

- del11q date: '2014., 2017.04.12.', error in date - NOTE: date must be in 'DD/MM/YYYY' format
- Hg: 'Low' is invalid for numeric field Hg.
- b2-microglobulin:'130', value violates numeric limits (0.1-100)

Data redundancy errors

- Patient id: X2 already exists
- Date of blood test: 2016-04-03', There is another blood test for this patient with the same date.

Predefined values

- Type of first treatment: 'FC', value is not included in the list of predefined values.
- Response to first treatment: 'CR, MRD negative', value is not included in the list of predefined values.

RWD management systems

DATA COLLECTION METHODS

Data Validation and Integration

Required Data

- Missing patient data: Start date of treatment is a required field for treated patients.
- Date of diagnosis of non-hematological neoplasm: Information is required when Other non-hematological malignant neoplasms=YES

Data inconsistency

- Rai stage at diagnosis: 'II' is not a valid value for Diagnosis 'SLL'
- Description of hematological adverse events at 1st-line treatment is not applied when depended information is missing: Adverse events at 1st-line treatment is empty

Date compare

- Date of diagnosis: '30/04/2015' cannot be more than 15 days greater than Date of first treatment: '05/02/2015'
- Start date of 3rd-line treatment: '10/01/2016' cannot be earlier than Start date of treatment – 2nd line of treatment: '11/02/2016'
- Date of death:'19/02/1997' cannot be less than del17p date: '13/11/2002'

RWD management systems

DATA COLLECTION METHODS

Data Validation and Integration

Examples of reported warnings

Warnings on missing data

- Further treatment at relapse / progression is essential for analysis (missing data)

Warnings on missing data for depended fields

- Missing ULN for b2-microglobulin is required for b2-microglobulin='3.49'

Warnings on data transformation

- Date of best response to treatment='27/04/20' has been auto-converted into '2020', please correct file if necessary and upload again (required format: 'DD/MM/YYYY')
- Tissue type: BM has been replaced with 'Bone Marrow'

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**

Prospective data registration → Data rules and constraints are implemented in the forms

- ✓ Form cannot be saved if required fields are empty
- ✓ Form input fields are restricted to the defined data format.
- ✓ Multiple choice or single choice options and drop down lists ensure data consistency
  - restricting input to the list of allowed values -
- ✓ Fields are dynamically displayed in order to avoid data inconsistency
  - only if applicable based on user's input -
- ✓ Required Fields are dynamically added based on user's input
- ✓ Duplicate records cannot be saved

RWD management systems

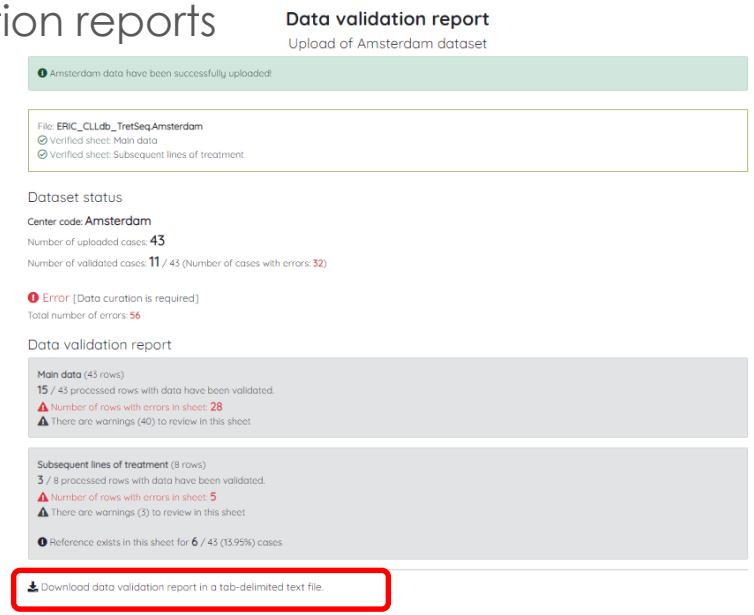
DATA COLLECTION METHODS

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



RWD management systems

DATA COLLECTION METHODS

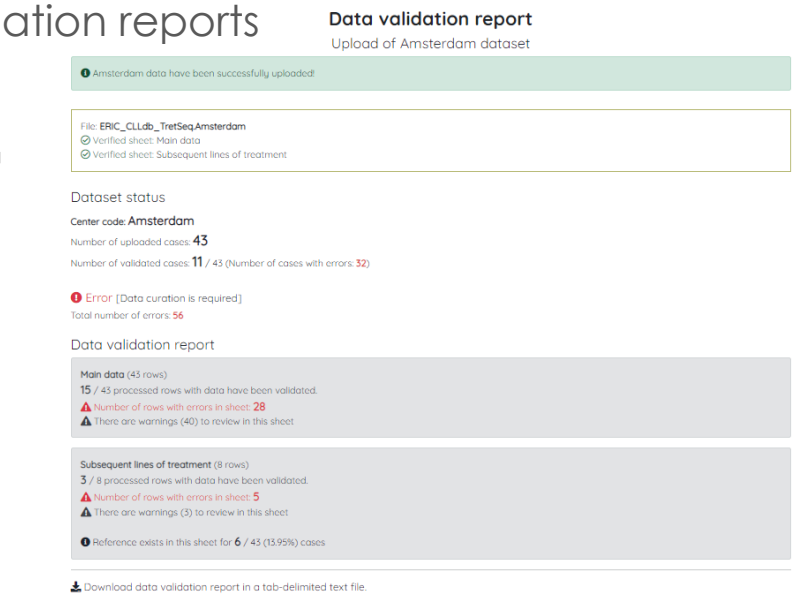
Data Validation and Integration

- **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

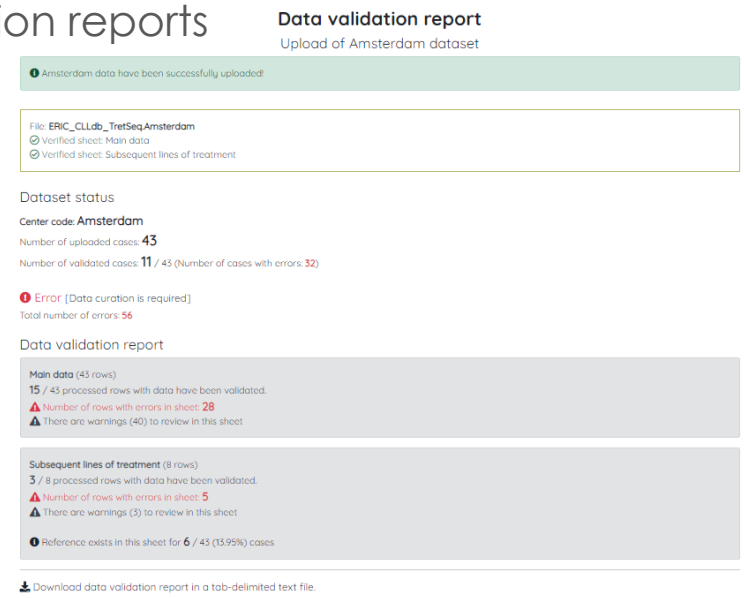
DATA COLLECTION METHODS

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

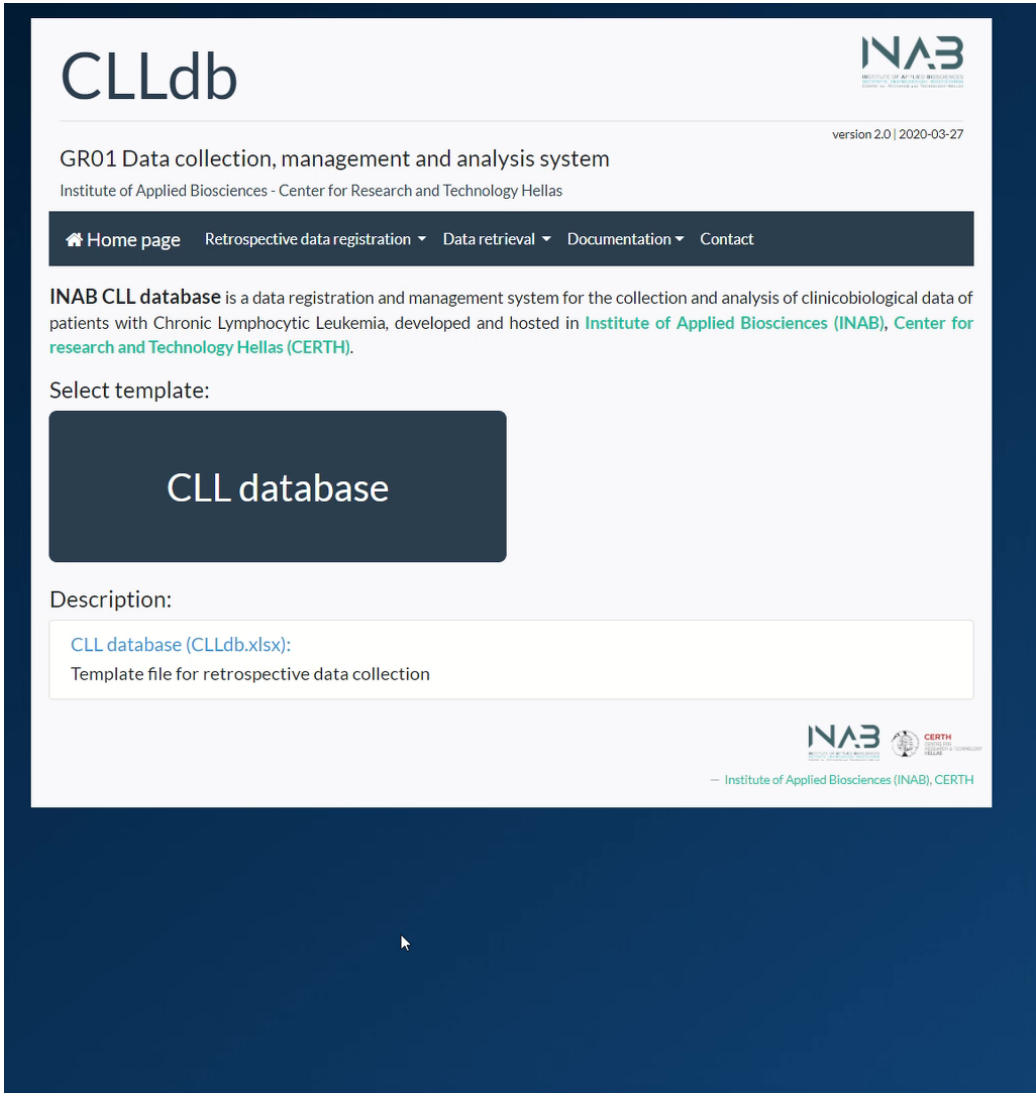
RWD management systems

DATA COLLECTION METHODS

Example of retrospective data registration system for the INAB CLL database:

Upload of new dataset

- Upload template file
- Wait for the system to process excel data:
  - Organize information
  - Validate data types
  - Check rules and constraints
- Retrieve data validation report
  - *Errors in dataset*
  - *Review warnings*



RWD management systems

DATA COLLECTION METHODS

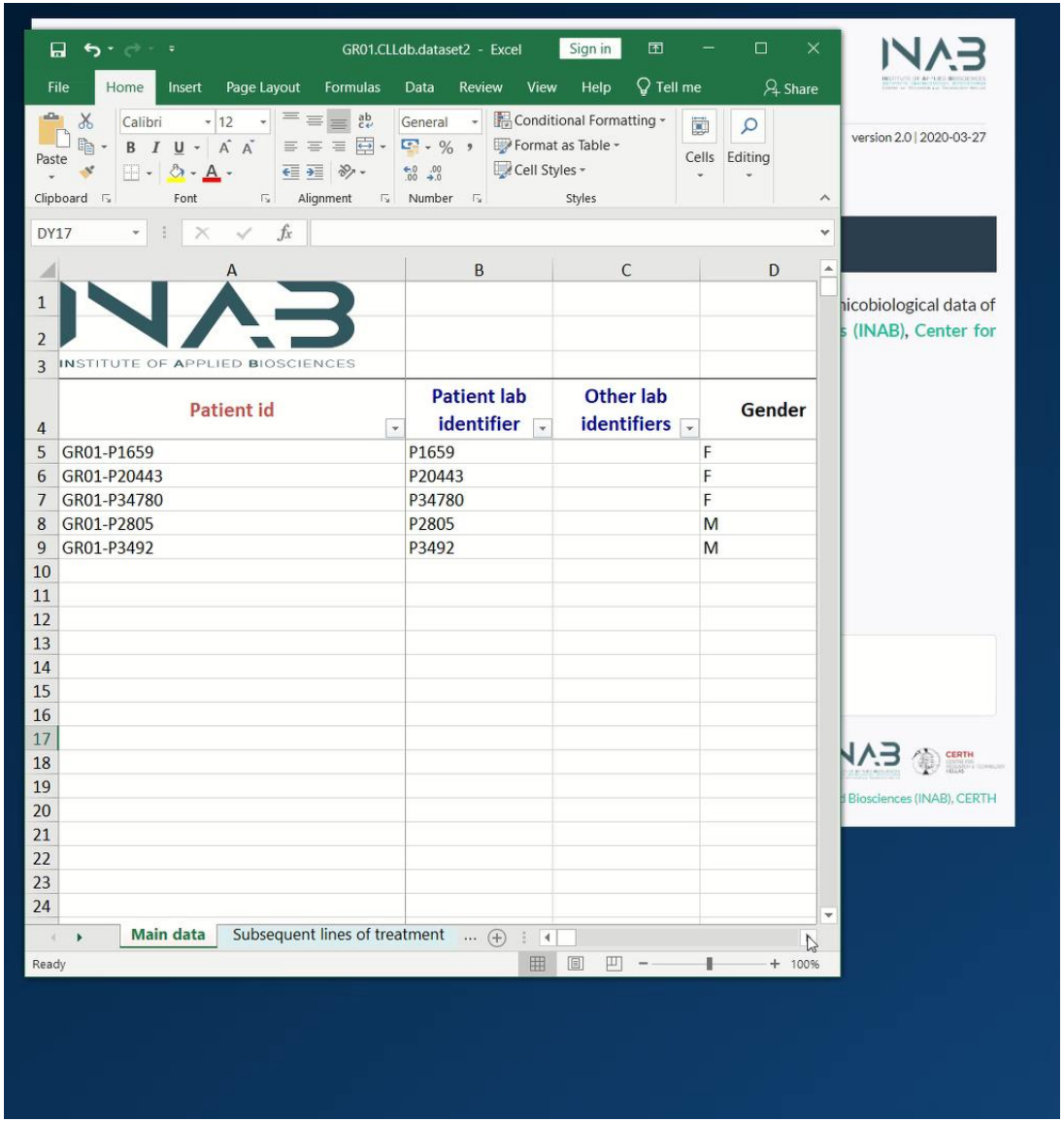
Example of retrospective data registration system for the INAB CLL database:

Upload of new dataset

- Upload template file
- Wait for the system to process excel data:
  - Organize information
  - Validate data types
  - Check rules and constraints
- Retrieve data validation report
  - *Errors in dataset*
  - *Review warnings*

Data curation process

- Update data in the template file according to the report
- Upload the updated version



RWD management systems

DATA RETRIEVAL AND ANALYSIS

Tools for data retrieval and analysis, based on user access privileges

Query tools

- allowing for dynamic definition of selection filters in order to retrieve data

Data retrieval and overview

Export modules

- report generation (patient summary)
- selected data download (authorized - anonymized)
- retrieve anonymized data

Download data options

Statistic and Visualization tools

- statistical analysis
- visualization (diagrams)

RWD management systems

DATA RETRIEVAL AND ANALYSIS

1. data selection
   - *Query / Search tools*

2. data retrieval

   Preview results for selected set of cases - *Including data visualization modules*

3. data analysis

   Load data to statistical analysis tool

4. data export options

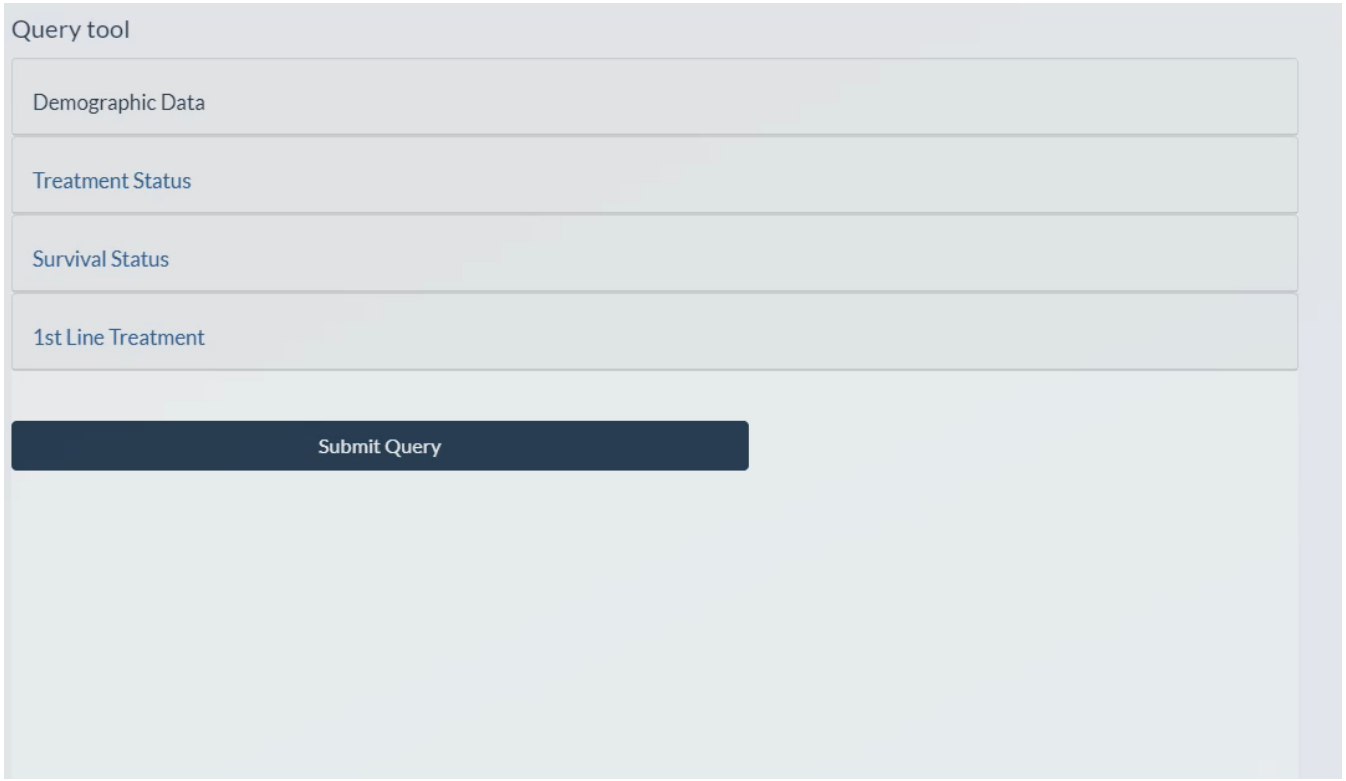
RWD management systems

DATA RETRIEVAL AND ANALYSIS

Data Retrieval

Data filtering

Query tool supporting dynamic definition of selection filters.



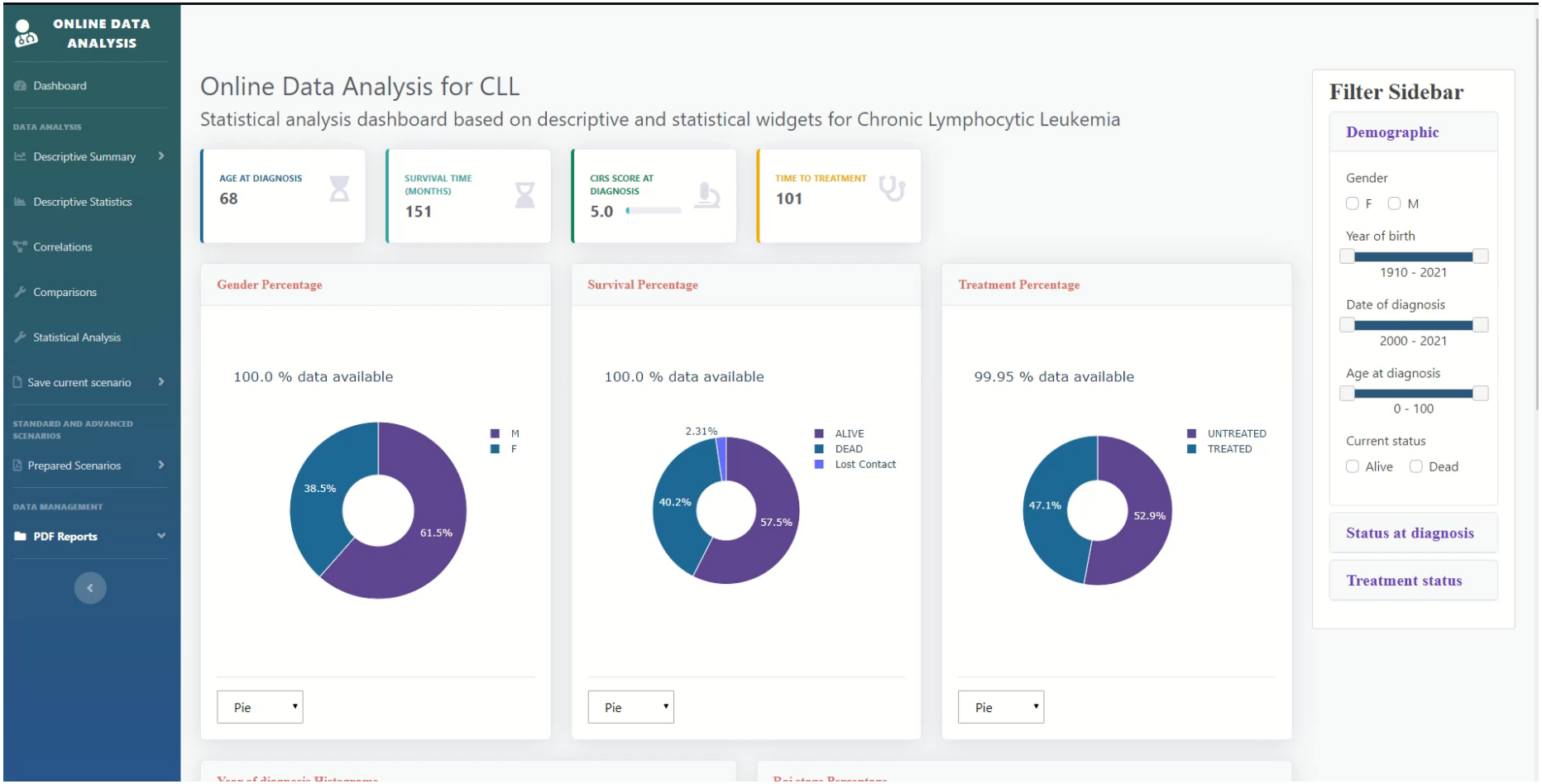
Query Tool

RWD management systems

DATA RETRIEVAL AND ANALYSIS

A web-based application for online statistical analysis

✓ Project-based configuration



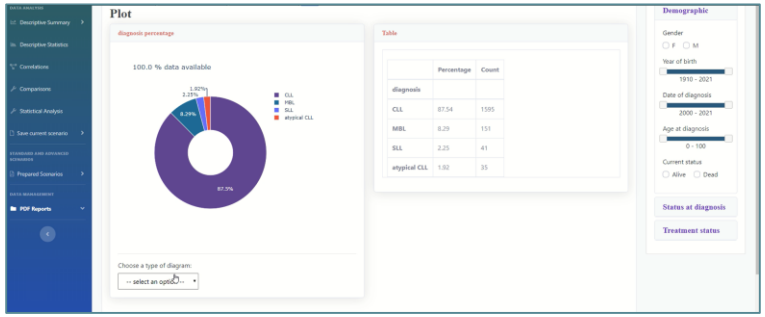
RWD management systems

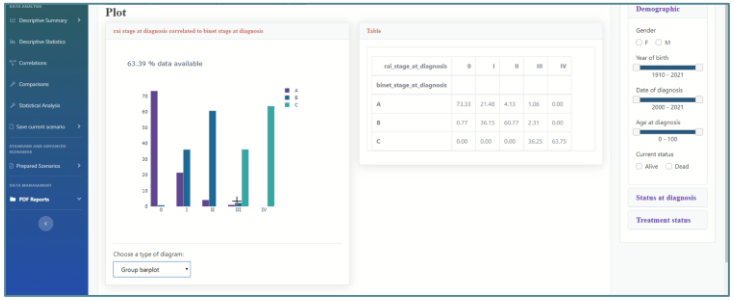
DATA RETRIEVAL AND ANALYSIS

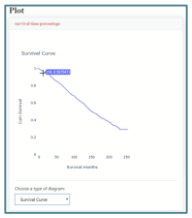
A web-based application for online statistical analysis

Data descriptive statistics and correlations

1. **Dynamic preview** and **edit** in real-time
2. **Correlations** between variables
3. **Comparisons** in different sub-groups and datasets
4. **Advanced statistical** analysis based on personalized prediction model







Prepared scenarios for statistical analysis
Automatically generated **pdf report**

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