



Olena Olishevskia
Senior Statistical Programmer
olena.olishevskia@estimates.fi

Co-authors:
Sami Virtanen, sami.virtanen@orionpharma.com
Mika Mäkinen, mika.makinen@estimates.fi

INTRODUCTION

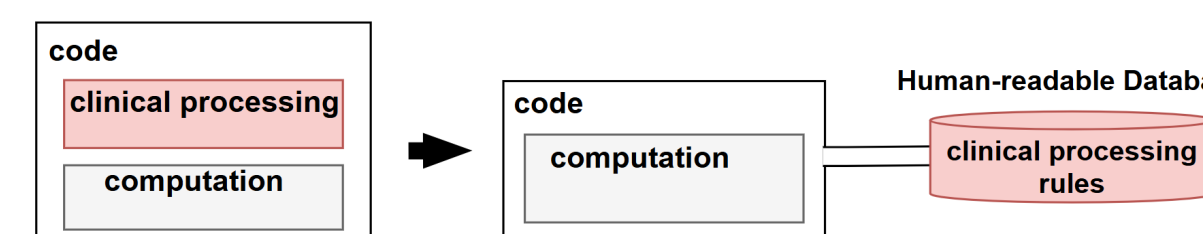
CHALLENGE

Standardization of raw data to SDTM standard can be challenging due to many reasons – poor data collection practices, varying data structures, inconsistent coding and many more.

Addressing these issues requires a framework of transparent and traceable **processing rules** to support consistent SDTM mapping.

APPROACH

We treat each **processing rule** as a distinct **decision** and store all decisions in a **centralized decision database**.



With a **centralized decision database**, we make the data transformation process more transparent, as we separate the clinical aspect of processing from the technical one.

BENEFIT

The **decision database** allows to:

- maintain consistent processing strategies across teams,
- evaluate data readiness for automatic SDTM domain creation.

It also allows decision-makers to evaluate the quality of data collection approaches, provide feedback upstream and improve the practices accordingly. The use of the **decision database** in data processing pipeline adheres to ICH E6 (R3) **emphasis on enhanced data governance**.

DECISION DATABASE

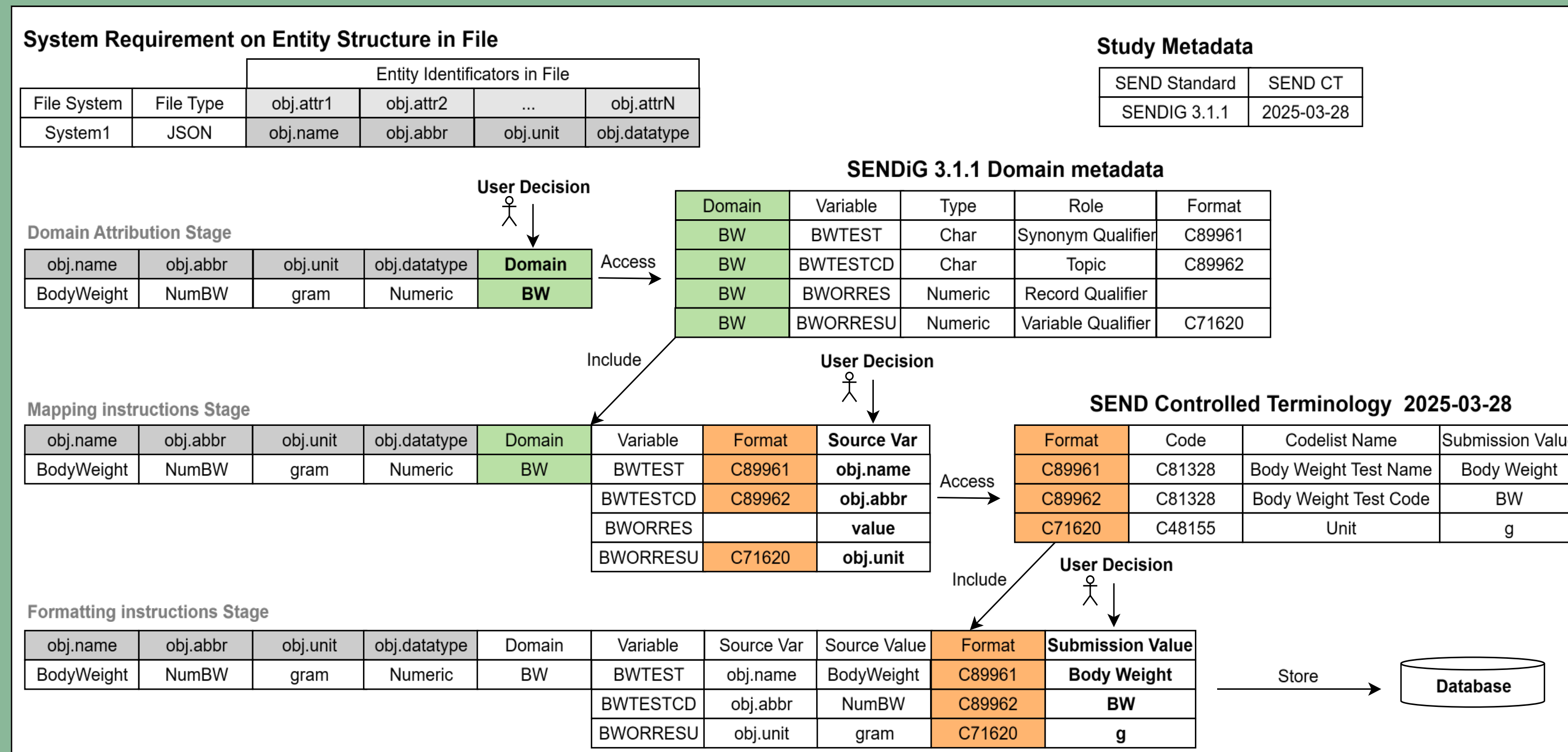
MAIN FOCUS

Entity is defined by combination of values in the identification variables and describes unique measurements, procedures, finding, events etc.

Entity-wise (cross-domain) decisions include rules on harmonizing data from different systems.

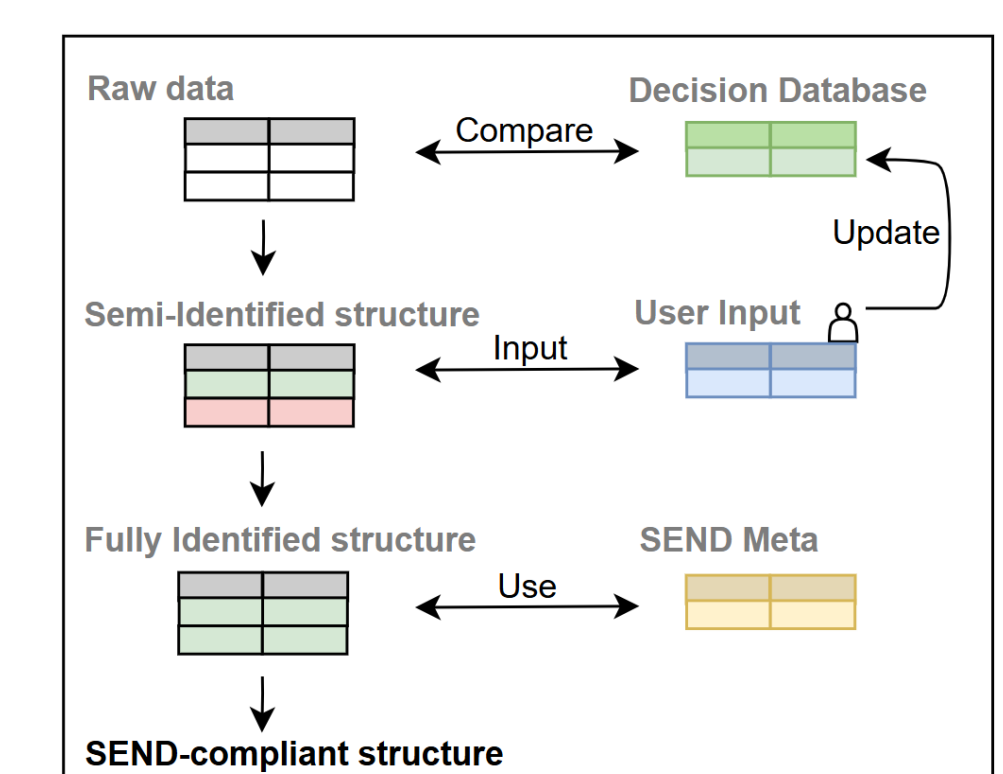
Within-entity (in-domain) decisions include rules on how variables and values within entity should be transformed to fit SDTM and SEND standards like:

- origin (user input, raw value, derivation)
- derivation method (direct or formatted copy, use of macros).



PRACTICAL USE

The **decision database** is used to build the **automated data transformation pipeline**.



By auto-applying confirmed rules from database to new datafiles we achieve automated creation of SEND-compliant domains with minimized manual effort.

MODULAR SOLUTION

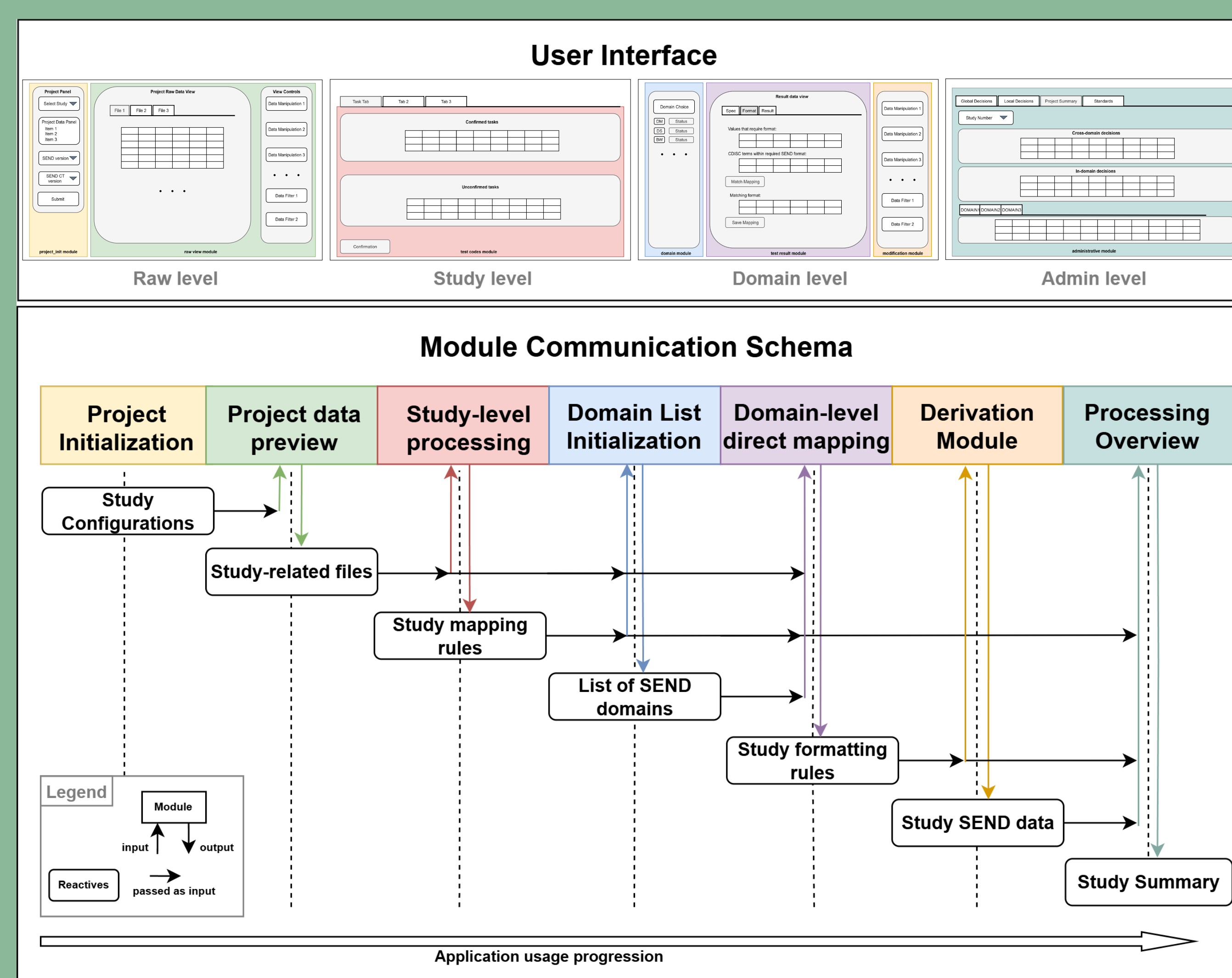
CONCEPTUAL DIVISION

Processing pipeline is divided into 4 main parts:

- **Raw level** (view of all study-related raw data)
- **Study level** (harmonization between datafiles, data division into domains etc.)
- **Domain level** (value formatting to CT or custom formats etc.)
- **Admin level** (project-wise view of all decision and result data).

Within each page, we divide processing pipeline into **block** with distinct encapsulated functionality. As we use **Shiny(R)** framework for the application, the conceptual **blocks** are translated into **Shiny modules**.

Such division is a basis for incorporating **User Access Level**, which allows decision-making and monitoring across teams.



FUNCTIONALITY

Data-driven approach and use of **decision database** allows to incorporate many useful features:

- Customized advanced search through study-related files
- Automatic detection of new data entities or changes in structure of new datafiles
- Entity-based domain attribution
- Metadata-driven creation of SEND domains

STUDY SUMMARY

As the processing of each study may involve tens, if not hundreds decisions, it is important to provide users with ability to cross-check and compare approaches across projects, teams and departments.

Administrative page provides overview of all the documents, decisions, standards and custom rules, that contributed to domain creation, as well as final domains themselves.

CONCLUSIONS

COMPLIANCE

SEND domains are following CDISC guidelines by:

- minimizing human error by limiting user actions in UI
- enforcing “confirmed” processing decisions over user-input
- supporting versioning of SENDIG and CT
- storing only validated processing rules in decision database

TRANSPARENCY

Process of domain creation is fully traceable by incorporating:

- full log of user actions
- database of processing decisions in human-readable format
- use of documented and validated functions for data transformations

USE CASE: APPLICATION PROTOTYPE

Based on our approach we developed an application prototype and successfully tested it on 48 studies animal studies and:

- gathered hundreds of processing rules to decision database
- developed a set of validated macros for metadata-driven SEND domains
- automatically created sets of SEND domains in .XPT format
- created audit-trail level log files

