

Open-Source MDR and SDR Managing Your Standards with the OpenStudyBuilder

Nicolas de SAINT JORRE, Novo Nordisk, Nantes, France
Mikkel TRAUN, Novo Nordisk, Copenhagen, Denmark

ABSTRACT

The OpenStudyBuilder (OSB) is an open-source project for clinical study specifications. This tool is a new approach for working with studies that once fully implemented will drive end-to-end consistency and more efficient processes - all the way from protocol development and CRF design - to creation of datasets, analysis, reporting, submission to health authorities and public disclosure of study information.

The OpenStudyBuilder contains:

- Standards and templates used for study specification
- Individual versioning of standard elements in the library
- Real-time team collaboration around study design by study team
- Central storage of study design data for multiple down-stream usages ('one source of truth')
- Versioning and full audit trail of study definitions

In the library section you can browse all kinds of standards, like for CDISC the SDTM, ADAM, SEND, CDASH and many more. Also, other dictionaries like SNOMED, LOINC, UCUM are available. MedDRA could be loaded as well of course. You can see different SDTM versions! As everything is FAIR, concepts can be created and maintained, so these can easily be reused. A core feature is the usage of syntax templates, for example for objectives and endpoints. Here general standard wordings and types are defined, which can then be instantiated with a concrete compound or comparator.

The OpenStudyBuilder, as an outcome of the CDISC 360° project and related to the TransCelerate Digital Data Flow project, is meant to be the core component for all study related activities and tools, managing and providing standards and study metadata. The solution comes along with tools, scripts and a graph database for industry and sponsor standards as well as study definitions.

We have close collaborations with other standard initiatives. The solution imports standards data through the CDISC library. For the TransCelerate DDF, we provide (and enhance) an API adapter to support the DDF API. And we are also working with CDISC on the COSMoS model and integration as well as being a CDISC Open Source Alliance (COSA) member.

INTRODUCTION

Why do we need to have every standard metadata in an efficient metadata system? As they are evolving over time and as they are driving the way we manage our Studies, having them in one tool ensures the accuracy in an efficient manner.

In the OpenStudyBuilder we can manage different types of standards:

- Controlled Terminologies
 - Code Lists. These can be defined by external organization (e.g., CDISC) or sponsor defined,
 - Dictionary Terminologies. These are a code submission value, an identifier and a definition defined by other standards organizations than CDISC (e.g., SNOMED, MED-RT, UNII, UCUM - MedDRA, LOINC etc.),
- Concept based standards. We refer to concept-based standards as data standards defined in more complex structures than code-value pairs. In the OpenStudyBuilder we have grouped the ones we currently cover into the following groups related to their scope:
 - Activities. These refer to our definition of Biomedical Concepts which cover clinical procedures and assessments with or without data collection,
 - Units. These refer to unit definitions including relationship to UCUM and CDISC CT as well as metadata supporting unit conversions,
 - CRFs. These generally refer to data collection instruments with reference to Activity Concepts as well as representation in CDISC ODM.XML format,

- Compounds. These refer to representations of medicinal products in the protocol as well as the data collection and submission datasets. This model is in alignment with the ISO IDMP standard.
- Syntax Templates. These are templates for textual representation of the main structured elements in the protocol e.g., study objectives, endpoints, time frames, criteria and activity instructions. The syntax template will support use of easily readable sentences with reference to template parameter values being linked to controlled terminologies and concept-based standards,
- CDISC Models and Implementation Guides with Sponsor metadata added to them.

OPENSTUDYBUILDER AS A METADATA REPOSITORY

In the library area of the frontend of the OpenStudyBuilder, you will be able to manage different types of standards. Metadata can be browsed, updated and retired. Additionally, new metadata can be created and maintained; they can easily be reused for example specially in studies managed inside the OpenStudyBuilder. This harmonized our metadata for better transversal management.

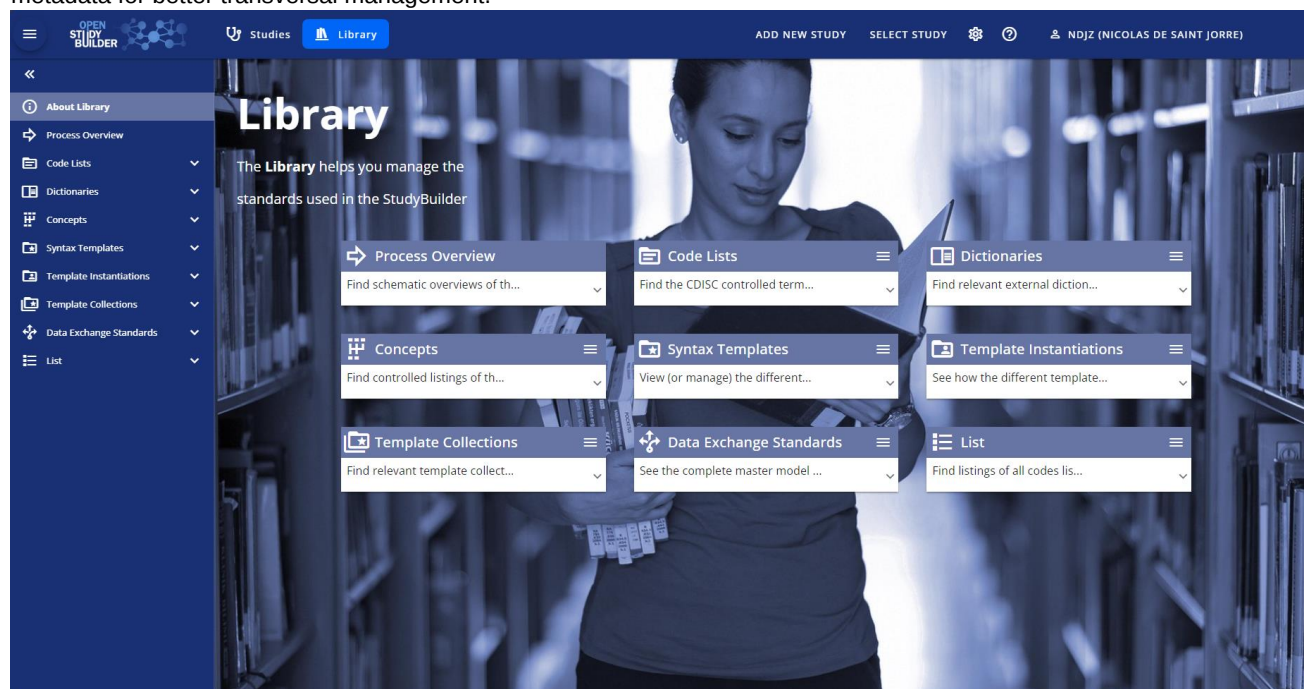


Figure 1: OpenStudyBuilder Library welcome page

LIBRARY	
CONTROLLED TERMINOLOGY	MEDICAL DICTIONARIES (e.g., MedDRA)
CONCEPTS (ACTIVITIES, UNITS, CRFs, COMPOUNDS)	TEMPLATES
DATA EXCHANGE STANDARDS	

Figure 2: List of Metadata available in the library zone

The metadata coming from Standards like the CDISC organization or from dictionaries are loaded using the OpenStudyBuilder Application Programming Interface (hereafter called API) directly from JSON files (like for CDISC CT) or via a migration process with CSV files (see Figure 3).

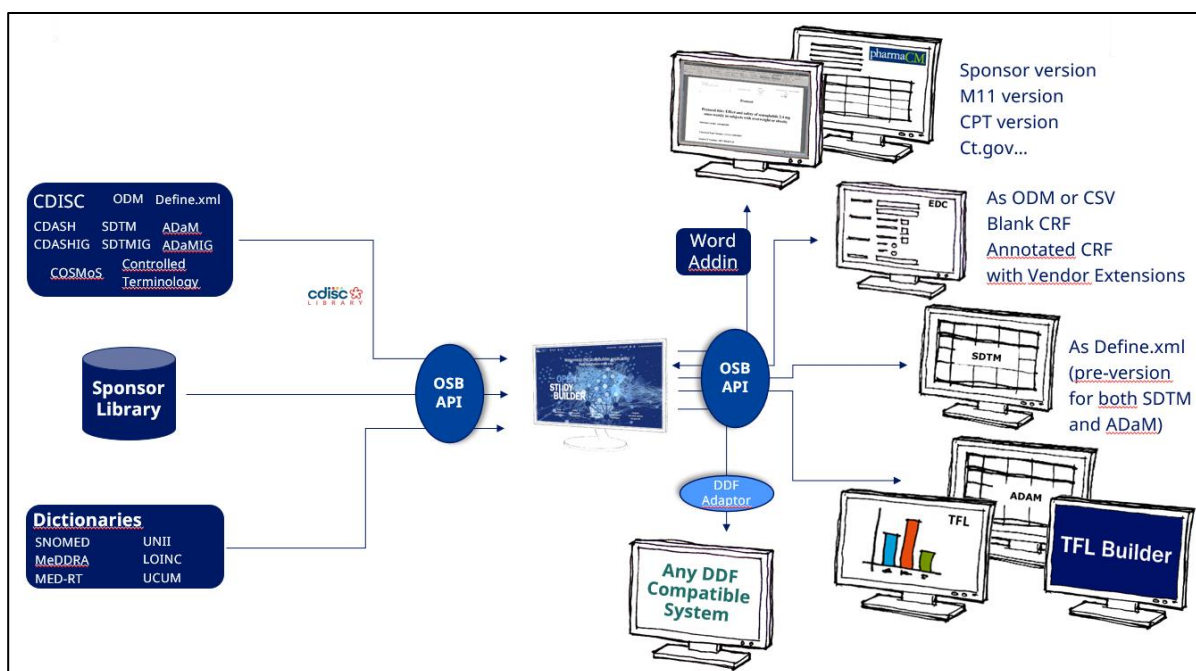


Figure 3: External workflow to and from OSB

This metadata is needed to create a complete and consistent set of concepts. This allows the creation of a protocol, for example following a sponsor's format, the ICH M11 guideline, the Common Protocol Template from TransCelerate, or others. This metadata is also used to create blank or annotated Case Report Form with or without vendor extension. We could also produce a pre-define.xml for both SDTM and ADaM, as TFL when needed! And we can already think of additional metadata usages.

It is already possible to convert some metadata from the OpenStudyBuilder into the Unified Study Definition Model (USDM) from TransCelerate via the OpenStudyBuilder API and a Digital Data Flow (DDF) Adaptor to produce standard metadata to be used by external compatible systems (as illustrated in Figure 3).

CONCEPTUAL MODEL OF THE OPENSTUDYBUILDER

The OpenStudyBuilder is defining the metadata repository with 5 layers of information at the library level (see Figure 4):

- The System Configuration where we have the Library Definition itself,
- The Industry Standards where we have the Controlled Terminology, the Foundational Data Standards from CDISC for example, the Conceptual Standards, the Therapeutic Area Standards and the external Dictionaries,
- The Sponsor Standards with the Sponsor Defined Terminologies, the Sponsor Defined Data Standards Extensions, the Sponsor Defined Conceptual Standards and the Therapeutic Area and the Project Standards,
- The Study Definitions where we are using the metadata managed by the two previous level to have the Study Definitions, the Study Designs and the Study Selections and Scheduling,
- And the Administrative Definition with the Projects, the Access groups, the Users and the System Roles.

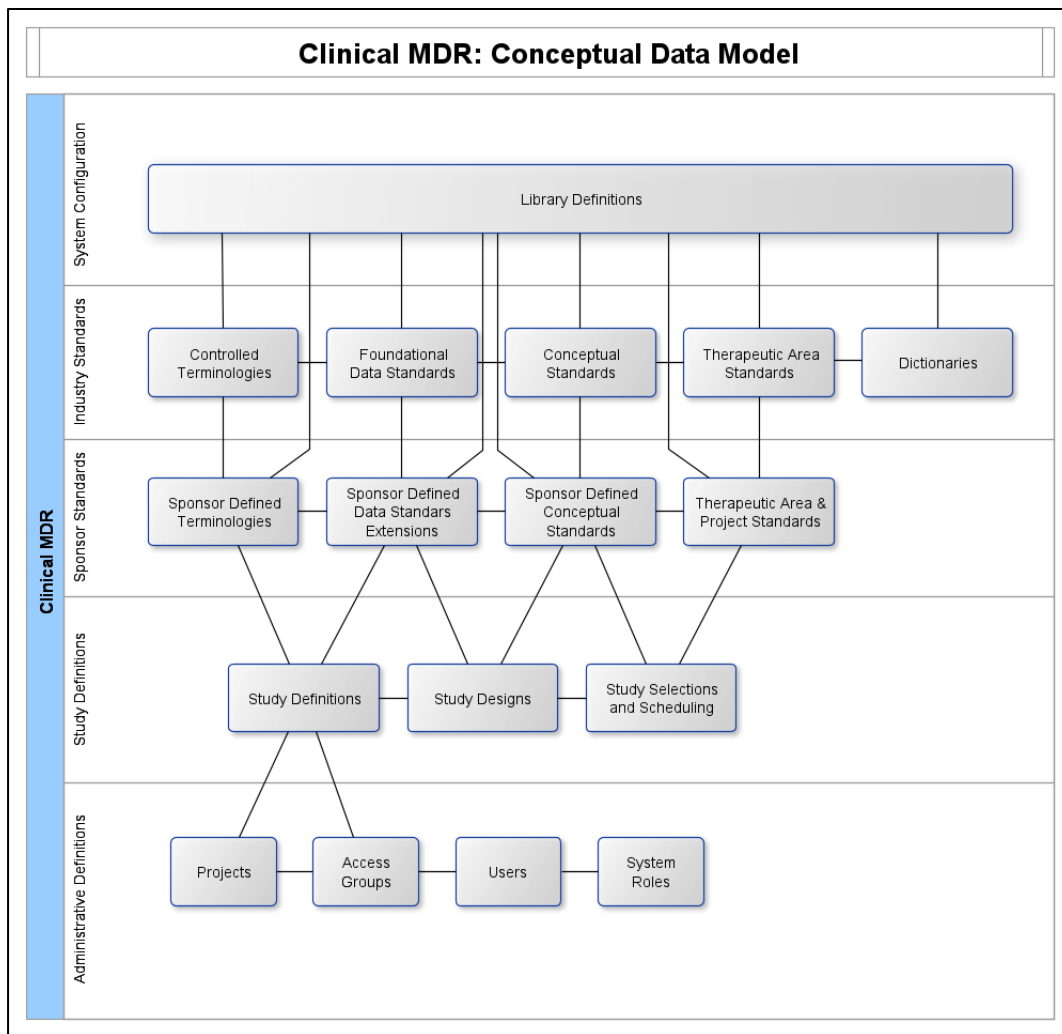


Figure 4: Conceptual Data Model

THE API OF THE OPENSTUDYBUILDER

The OpenStudyBuilder is using a Graph database to store every metadata.

What is a Graph Database? It is a database that uses graph structures for semantic queries with nodes, edges, and properties to represent and store data. A key concept of the system is the graph (or edge or relationship). The graph relates the data items in the store to a collection of nodes and edges, the edges representing the relationships between the nodes.

On top of this, the OpenStudyBuilder is having an interface (API) with multiples Endpoints (see Figure 5):

- The HTTP GET request is used to fetch a resource from the Graph database,
- The HTTP POST request is used to post data or create a resource in the Graph database,
- The PATCH request updates a resource but requires only the field(s) which is being updated in the payload,
- The DELETE request is used to delete a resource from the Graph database (here it is a soft delete always, we are not deleting any record).

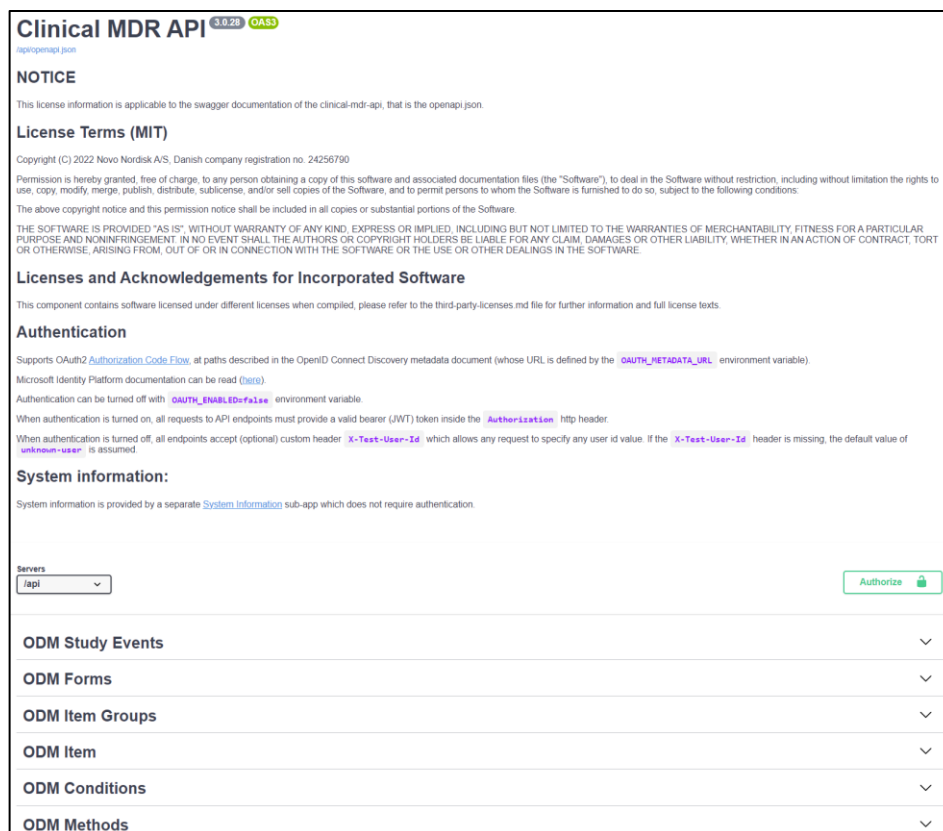


Figure 5: Swagger documentation of the OSB API (using Fast API)

VERSION CONTROL OVER OUR STANDARD METADATA

The OpenStudyBuilder is using a sophisticated version control solution to track the evolution of standards. For example, the Controlled Terminology can be updated 4 times a year, with addition, modification and even deletion – See Figure 6. This version Control is insuring the non-duplication of metadata. Here, for example the node "Unknown" is shared with every other codelist using it.

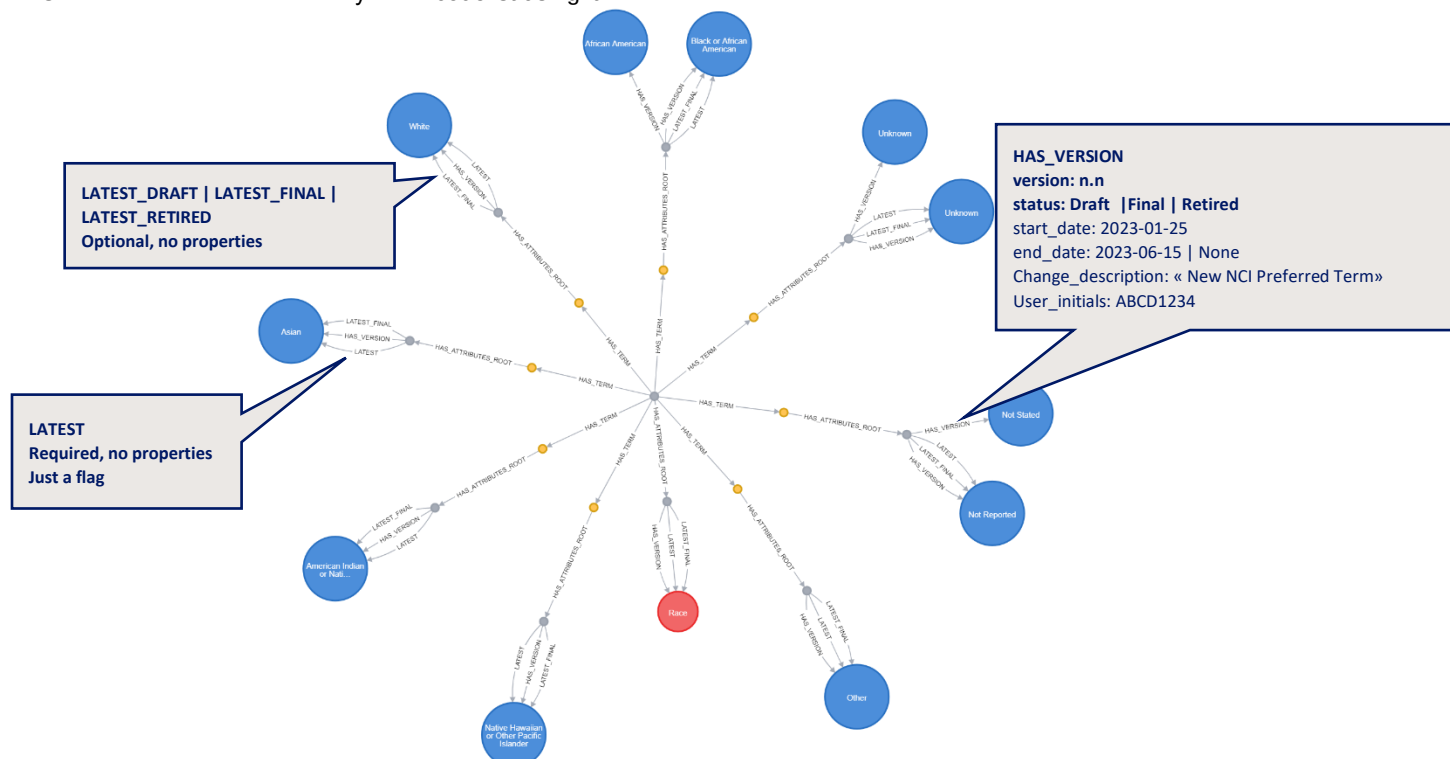


Figure 6: Version Control on the Codelist RACE

THE CONTROLLED TERMINOLOGY IN OPENSTUDYBUILDER

One of the first data loaded into the OpenStudyBuilder is directly coming from the CDISC Organization: The API is used to download JSON files from the CDISC Library and then injected into an intermediate database where we remove duplicates. Then the API is moving that metadata into the OpenStudyBuilder database, where this is flagged as 'CDISC' in the library. It is possible to browse the Codelists by catalogues (model) or by packages (date) and catalogues (model).

As soon as this metadata is available, it is possible to add sponsor information like a dedicated name. It is also possible to add sponsor codelists and terms, but again, here the API is re-using the existing metadata, preventing duplication – see figure 7 and 8.

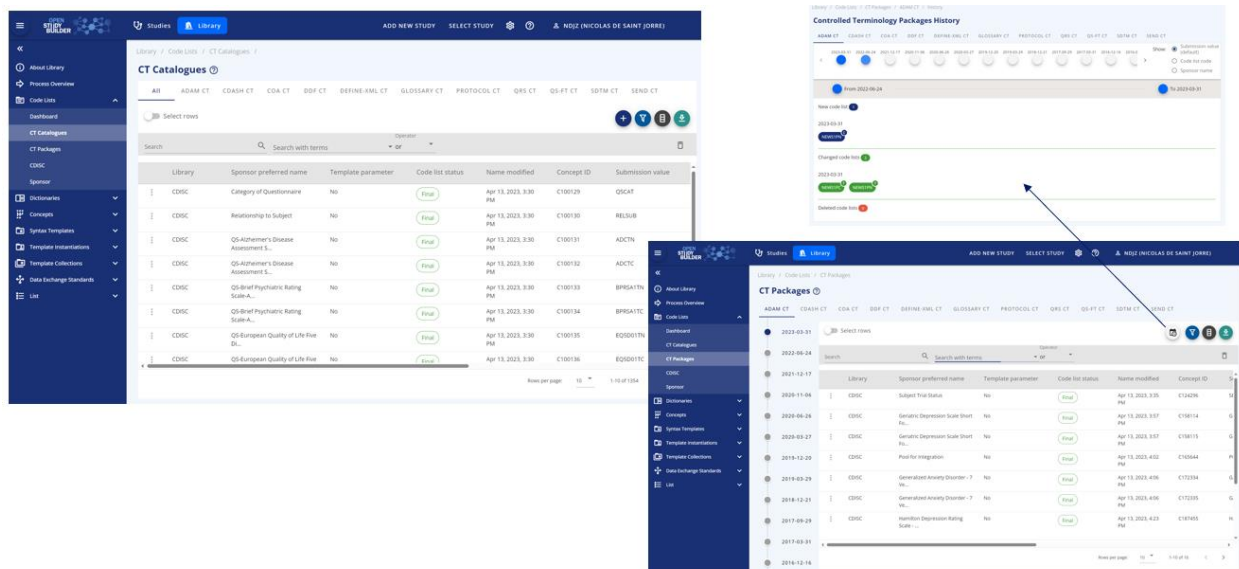


Figure 7: Controlled Terminology with metadata history

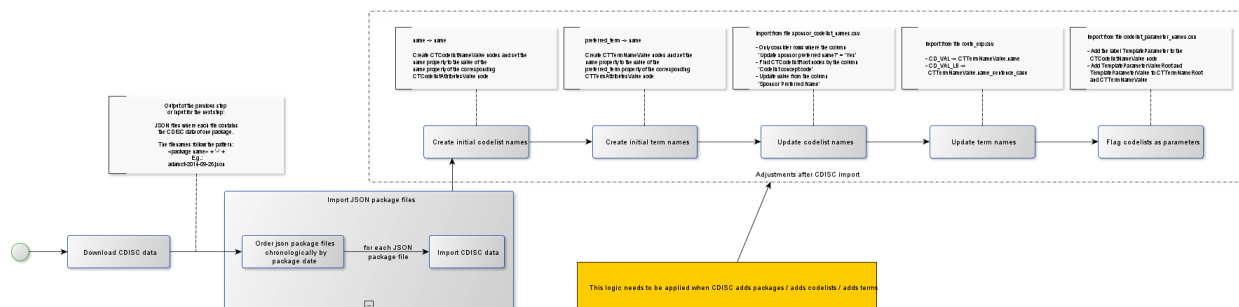


Figure 8: Controlled Terminology upload workflow

THE DICTIONARIES IN THE OPENSTUDYBUILDER

In the Library of the OpenStudyBuilder, you have access to dictionaries. You will find here the following metadata (see Figure 9):

- The SNOMED CT - Systematized Nomenclature of Medicine - Clinical Terms,
- The MED-RT - Medication Reference Terminology for Pharmacologic Class (PCLASS)
- The UNII - Unique Ingredient Identifier for Active Substances
- UCUM - Unified Code for Units of Measure

Here also, the metadata are loaded with a version control track. And for each dictionary, it is possible for the end user to add terms inside the sponsor library.

Library / Dictionaries / SNOMED						
SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms) for Diseases and Disorders						
Search	Q					
SNOMED ID	Preferred synonym	Preferred synonym (lower case)	Abbreviation	Definition	Status	Version
1	44572001	Diabetes	diabetes	Diabetes (disease)	Final	2023.1.0
2	36286005	Disorder of body system	disorder of body system	Disorder of body system (disorder)	Final	2023.1.0
3	605964002	Pre-existing type 1 diabetes mellitus	pre-existing type 1 diabetes mellitus	Pre-existing type 1 diabetes mellitus	Final	2023.1.0
4	444225005	Heart failure with normal ejection	heart failure with normal ejection	Heart failure with normal ejection	Final	2023.1.0
5	44265003	Nonalcoholic steatohepatitis	nonalcoholic steatohepatitis	Nonalcoholic steatohepatitis (disease)	Final	2023.1.0
6	44170003	Severe hereditary factor in diabetes	severe hereditary factor in diabetes	Severe hereditary factor in diabetes	Final	2023.1.0
7	44070008	Severe hereditary factor	severe hereditary factor	Severe hereditary factor	Final	2023.1.0
Rows per page: 10 1-10 of 20						
Library / Dictionaries / UCUM						
UCUM (Unique Ingredient Identifier) for Active Substances						
Search	Q					
UNI ID	Substance name	Substance name (lower case)	Abbreviation	Status	Version	Modified
1	Q51804364	LIVOSTERONE	livostosterone	Final	2023.1.0	2023.1.0
2	U78805429	MOXIFLOXACIN	moxifloxacin	Final	2023.1.0	2023.1.0
3	546450700C	GLYBURIDE	glyburide	Final	2023.1.0	2023.1.0
4	Q99912072	STAGOLIN	stagolin	Final	2023.1.0	2023.1.0
5	Y1107058	INSULIN HUMAN	insulin human	Final	2023.1.0	2023.1.0
6	362007500	PARACETAMOL	paracetamol	Final	2023.1.0	2023.1.0
7	Y11053224	METFORMIN	metformin	Final	2023.1.0	2023.1.0
8	W7729505	DULAGLUTIDE	dulaglutide	Final	2023.1.0	2023.1.0
9	X76676900C	GLIPIZIDE	glipizide	Final	2023.1.0	2023.1.0
10	440071405	PROGLUTERONE	progluteron	Final	2023.1.0	2023.1.0
Rows per page: 10 1-10 of 20						
Library / Dictionaries / MED-RT						
MED-RT (Medication Reference Terminology) for Pharmacologic Class (PCLASS)						
Search	Q					
MED-RT ID	Class name	Class name (lower case)	Abbreviation	Definition	Status	Version
1	N000020185	ORAL ANTIDIABETIC AGENTS	oral hypoglycemic agents	ORAL ANTIDIABETIC AGENTS	Final	2023.1.0
2	N000019730	PCSK9 INHIBITORS	pcsk9 inhibitors	PCSK9 INHIBITORS	Final	2023.1.0
3		ZILVERMAB	zilvermab	ZILVERMAB	Final	2023.1.0
4	N000017045	L-THYRONE	l-thyroxine	L-THYRONE	Final	2023.1.0
5	N000017046	Tubulin Inhibiting Agent	tubulin inhibiting agent	Tubulin Inhibiting Agent	Final	2023.1.0
6	N000018010	Thiazolidinedione	thiazolidinedione	Thiazolidinedione	Final	2023.1.0
7	N000017048	Sulfaphenazole	sulfaphenazole	Sulfaphenazole	Final	2023.1.0
8	N000017049	Sulfonamide	sulfonamide	Sulfonamide	Final	2023.1.0
Rows per page: 10 1-10 of 20						
Library / Dictionaries / UCUM						
UCUM (Unique Code for Units of Measure)						
Search	Q					
UCUM code	UCUM description	Status	Version	Modified		
1	gww	gram per week	Final	2023.1.0	2023.1.0	2023.1.0
2	gwwww	gram per week	Final	2023.1.0	2023.1.0	2023.1.0
3	(gww)	gram per week unit	Final	2023.1.0	2023.1.0	2023.1.0
4	(gww)	gram per week unit	Final	2023.1.0	2023.1.0	2023.1.0
5	(gww)	gram per week unit	Final	2023.1.0	2023.1.0	2023.1.0
6	(gww)	gram per week unit	Final	2023.1.0	2023.1.0	2023.1.0
7	(gww)	gram per week unit	Final	2023.1.0	2023.1.0	2023.1.0
8	(gww)	gram per week unit	Final	2023.1.0	2023.1.0	2023.1.0
9	(gww)	gram per week unit	Final	2023.1.0	2023.1.0	2023.1.0
10	(gww)	gram per week unit	Final	2023.1.0	2023.1.0	2023.1.0
Rows per page: 10 1-10 of 20						

Figure 9: Dictionaries available in the OpenStudyBuilder

CONCEPTS: ACTIVITIES AND ACTIVITY INSTANCES

In OpenStudyBuilder we use the term ActivityConcept for the implementation of a Biomedical Concept with the following definition:

We differentiate between two different types of activity concepts: a semantic logical observation and a semantic specific activity. If the activity is related to data collection, then the activity corresponds to a logical data observation; otherwise, it's a semantic specific activity.

The activity related to a logical data observation is defined at a level of granularity where it will match an observation: an example can be Systolic blood pressure measured sitting, in unit of mmHg, at a specific point in time.

The semantic specific activity not related to data collection can e.g., be dosing instructions for the subject.

The OpenStudyBuilder has functionality to create an electronic "protocol". As a user you can define the study design, the visits and what is to be performed/collected at each visit (called Schedule of Activities (SoA) in OpenStudyBuilder, others may call it Flowchart). The Activity Concepts are the activities in the protocol Schedule of Activities.

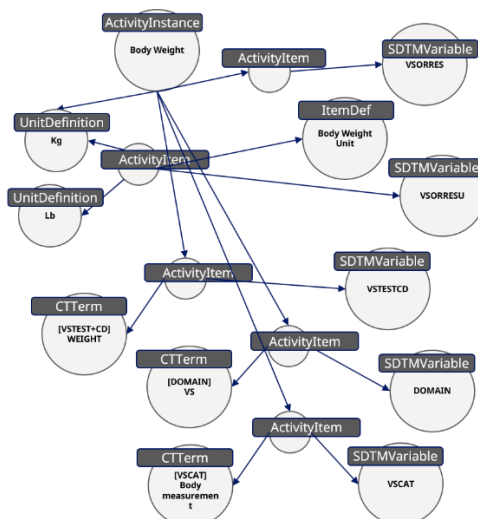


Figure 10: Body Weight Example Activity Concept

Browsing the Library will allow the end user to manage Activities but also Activity Instances with Activity Items as in

the Figure 11.

The figure illustrates the workflow from a general activity to a specific activity instance and its associated items. It consists of three main panels:

- Activities Panel:** A table listing various activities. The 'Bilirubin' activity is highlighted.
- Bilirubin Activity Details Panel:** A detailed view of the 'Bilirubin' activity, showing its definition, abbreviation, and activity groupings.
- Bilirubin Urine Activity Instance Details Panel:** A detailed view of the 'Bilirubin Urine' activity instance, showing its definition, abbreviation, and activity groupings.

Arrows indicate the navigation path from the 'Activities' list to the 'Bilirubin' activity details, and then to the 'Bilirubin Urine' activity instance details.

Figure 11: From Activity to Activity Instance to Activity Items...

CONCEPTS: UNITS WITH CONVERSION RULES

Next to the Activities, the OpenStudyBuilder is equipped with 374 units that are linked, when applicable, to the UCUM and the Controlled Terminology Codelist UNIT of CDISC.

The figure shows the 'Units' table in the OpenStudyBuilder interface. The table lists various units with columns for Library, Name, Master unit, Display unit, Unit subsets, UCUM unit, CT Unit terms, Convertible unit, SI unit, and US conversion. Two callout boxes highlight the 'UCUM unit' and 'CT Unit terms' columns, with text indicating their relationship to UCUM and CT units.

Relationship to UCUM Units

And to CT Units

Figure 12: Unit listing with UCUM and CT mapping

CONCEPTS: CASE REPORT FORM

In addition to the Library, we offer a comprehensive Case Report Form (CRF) management system that adheres to the CDISC Operational Data Model (ODM) for terminology, structure, and import/export format. This industry-standard approach ensures optimal connectivity and interoperability.

Our CRF management system empowers end-users to create Forms, Item Groups, and Items, which can be bound within a template. This template can be validated and reused, and its metadata can be linked to the activities available in the Library. This integration enables seamless usage by the protocol (refer to Figure 13).

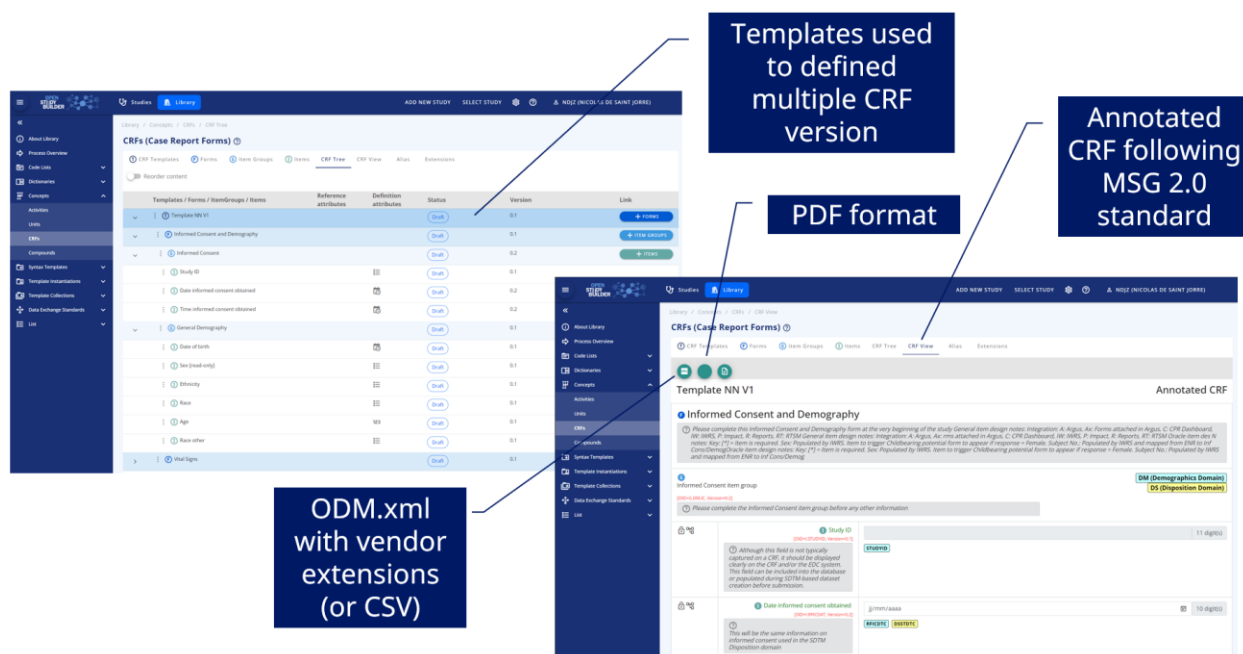


Figure 13: Case Report Forms with Annotated version

CONCEPTS: COMPOUNDS

These refer to the representations of medicinal products in the protocol, as well as the datasets used for data collection and submission. Our model aligns with the ISO Identification of Medicinal Products (IDMP) standard, ensuring consistency and interoperability.

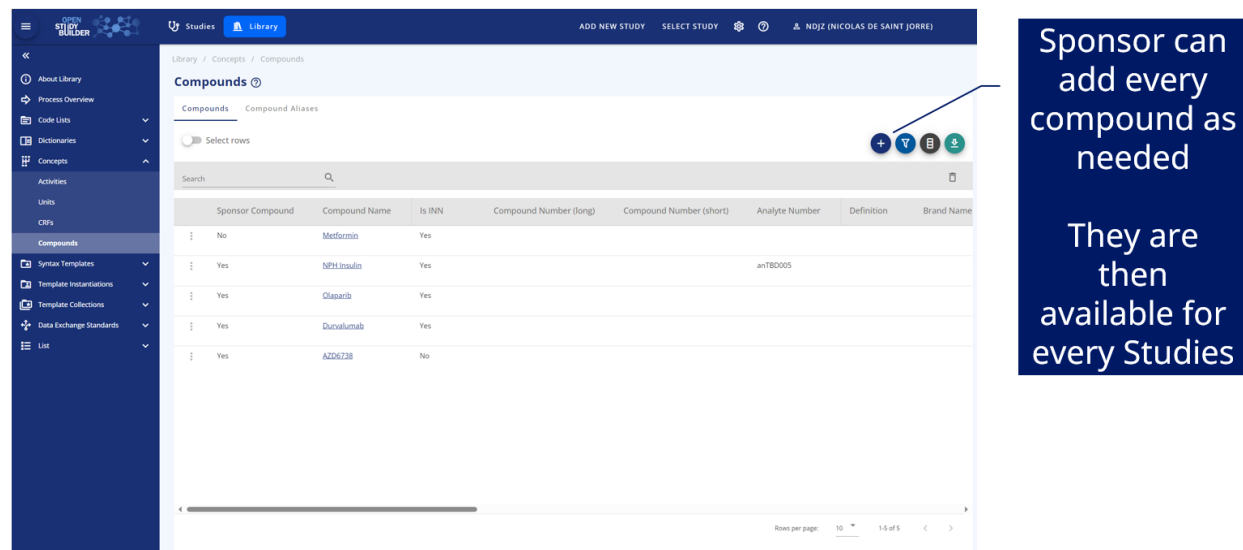


Figure 14: Compounds in the Library

SYNTAX TEMPLATE: OBJECTIVES, ENDPOINTS, TIME FRAME, CRITERIA, ACTIVITY, FOOTNOTES

Our syntax templates provide a textual representation of the main structured elements in the protocol, such as study objectives, endpoints, time frames, criteria, and activity instructions. These templates use easily readable sentences that reference template parameter values linked to controlled terminologies and concept-based standards. References to template parameters are displayed between square brackets, as shown in Figure 14.

Objective Templates

Sequence number	Indication or disorder	Objective category	Confirmatory testing	Parent template	Modified	Status	Version
09	Heart failure	Not Applicable	Yes	To compare the effect of Compound relative to Comparator on Activity/Instance when something	Jul 4, 2023, 3:08 PM	Final	1.0
08	Not Applicable	Not Applicable	Yes	Time to first occurrence of MACE, a composite endpoint consisting of: CV death, nonfatal MI, nonfatal stroke, or hospitalization for unstable angina	Jul 4, 2023, 3:08 PM	Final	1.0

Add objective template pre-instantiation

Set template parameter values

To compare the effect of metformin relative to rgh insulin on diabetic retinopathy - cataract when something

Unformatted plain text representation

To compare the effect of metformin relative to rgh insulin on diabetic retinopathy - cataract when something

Templates are using 'parameters' (in Green) that will be replaced when used at Study level

Figure 14: Example of Objective Template with reference to template parameter

DATA EXCHANGE STANDARDS – CDISC MODELS

In our latest release, the Library now manages Models and Implementation Guides for the Study Data Tabulation Model (SDTM/SDTMIG), complete with version control similar to that of the Controlled Terminology. This set of metadata offers internal connections, enabling end-users to browse the Models and Implementation Guides by packages and navigate to the Controlled Terminology (refer to Figure 15, accessible via the link in the Codelist column).

SDTM/SDTMIG

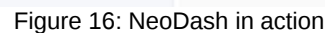
Name	Ordinal	Label	Data Type	Role	Core	Codelist	Described Value Domain	Implements	Value List	Description
STUDID	1	Study Identifier	Char	Identifier	Req	CDISC	Unique identifier for a study	CDISC	CDISC	Unique identifier for a study
DOMAIN	2	Domain Abbreviation	Char	Identifier	Req	CDISC	Two-character abbreviation for the domain	CDISC	CDISC	Two-character abbreviation for the domain
SUBJID	3	Unique Subject Identifier	Char	Identifier	Req	CDISC	Identifier used to uniquely identify a subject across all studies for all applications for submissions involving the product	CDISC	CDISC	Identifier used to uniquely identify a subject across all studies for all applications for submissions involving the product
SEQID	4	Sequence Number	Num	Identifier	Req	CDISC	Sequence number given to ensure uniqueness of related records within a domain. May be any valid number	CDISC	CDISC	Sequence number given to ensure uniqueness of related records within a domain. May be any valid number
GRPID	5	Group ID	Char	Identifier	Perm	CDISC	Used to tie together a block of related records in a single domain for a subject	CDISC	CDISC	Used to tie together a block of related records in a single domain for a subject
SPECID	6	Specimen ID	Char	Identifier	Perm	CDISC	Internal or external specimen identifier. Examples: specimen ID	CDISC	CDISC	Internal or external specimen identifier. Examples: specimen ID
SPEID	7	Specimen Defined Identifier	Char	Identifier	Perm	CDISC	Specimen-defined reference number. May be prepended on the CDISC or a regional identifier as defined in the sponsor's operational database. Examples: Line number on the lab log	CDISC	CDISC	Specimen-defined reference number. May be prepended on the CDISC or a regional identifier as defined in the sponsor's operational database. Examples: Line number on the lab log
LABTESTCD	8	Lab Test or Examination Short Name	Char	Tag	Req	CDISC	Short name of the measurement, test, or examination described in LABTEST. It can be used as a column name when converting a dataset from a protocol to a horizontal format. The value in LABTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g., "TEST1" is not valid). LABTESTCD cannot contain characters other than letters, numbers, or underscores. Examples: "ALT", "GLUC"	CDISC	CDISC	Short name of the measurement, test, or examination described in LABTEST. It can be used as a column name when converting a dataset from a protocol to a horizontal format. The value in LABTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g., "TEST1" is not valid). LABTESTCD cannot contain characters other than letters, numbers, or underscores. Examples: "ALT", "GLUC"

Importing Models from the CDISC Library with Version Control + Sponsor additional metadata

CDASH and ADaM Models will be added soon...

Figure 15: SDTM/SDTMIG in the OpenStudyBuilder

NeoDash is a low-code, open-source Dashboard Builder designed for Neo4j, the Graph database used by the OpenStudyBuilder. With NeoDash, users can easily create interactive dashboards featuring tables, graphs, bar charts, line charts, maps, and more, using the data available in the database (refer to Figure 16).



Finally, the OpenStudyBuilder aims to become a reference implementation for Digital Data Flow (DDF), offering a range of features to support this goal. As such, any tools that support DDF will be able to work seamlessly with the OpenStudyBuilder.

One of the main advantages of DDF is its standard API, which enables the connection of diverse up- and downstream systems. As the OpenStudyBuilder supports the same APIs, all tools that support the DDF project through their standard APIs should also be usable with the OpenStudyBuilder.



CONCLUSION

Currently, the OpenStudyBuilder provides a wealth of metadata in the Library section, complete with a full version management workflow. End-users can consume this data in every study, ensuring harmonization of information and increasing interoperability between the protocol, CRF, and SDTM Define.xml specifications. These metadata can be cleverly connected, creating a concrete platform that heavily utilizes biomedical concepts to enhance clinical trials.

ACKNOWLEDGMENTS

We would like to thank Novo Nordisk for sharing this project as an open-source solution. Additionally, we would like to thank the complete OpenStudyBuilder team which is a squad of around 30 highly skilled professionals who work together in a dynamic environment to deliver and enhance the amazing OpenStudyBuilder solution.

We are looking forward to collaborations, so if you are interested in joining the OpenStudyBuilder project, please reach out to us.



Figure 18: OpenStudyBuilder team

RECOMMENDED READING

1. Project Homepage, detailed project information, guides and more is available on the OpenStudyBuilder homepage: <https://openstudybuilder.com/>
2. Project Resources, if you would like to see various information resources available, you can check out the resource-list: https://openstudybuilder.com/info_resources/
3. OpenStudyBuilder demonstration video, we have a 30' demonstration video available showing the functionality of the OpenStudyBuilder in version 0.4: <https://www.youtube.com/watch?v=dL5CY0BwfEs>
4. GitLab repository, the complete solution is available in GitLab. A docker setup is available as well: <https://gitlab.com/Novo-Nordisk/nn-public/openstudybuilder/OpenStudyBuilder-Solution/>
5. Newsletter, our newsletter is getting out through linkedIn: <https://www.linkedin.com/newsletters/openstudybuilder-6990328054849916928/>
6. Slack channel, we do have a slack space available where you can ask any question: https://join.slack.com/t/openstudybuilder/shared_invite/zt-19mtauzic-Jvrhtmy7hGstgviVb1Wsw

CONTACT INFORMATION

(In case a reader wants to get in touch with you, please put your contact information at the end of the paper.)

Your comments and questions are valued and encouraged. Contact the author at:

Nicolas DE SAINT JORRE / Mikkel TRAUN

Novo Nordisk

7 Rue Général Lanrezac

Nantes / 44000

Work Phone: +33 6 20 88 38 10

Email: ndjz@novonordisk.com

OpenStudyBuilder Mail: OpenStudyBuilder@gmail.com

OpenStudyBuilder Slack: https://join.slack.com/t/openstudybuilder/shared_invite/zt-19mtauzic-Jvrhtmy7hGstgyilvB1Wsw

Brand and product names are trademarks of their respective companies.