Paper SI03

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

LIBRA	RY
CONTROLLED TERMINOLOGY	MEDICAL DICTIONARIES (e.g., MedDRA)
CONCEPTS (ACTIVITIES, UNITS, CRFs, COMPOUNDS)	TEMPLATES
DATA EXCHANGE	E 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):

- <u>The System Configuration</u> where we have the Library Definition itself,
- <u>The Industry Standards</u> 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,
- <u>The Sponsor Standards</u> with the Sponsor Defined Terminologies, the Sponsor Defined Data Standards Extensions, the Sponsor Defined Conceptual Standards and the Therapeutic Area and the Project Standards,
- <u>The Study Definitions</u> 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,
- <u>And the Administrative Definition</u> 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).

Clinical MDR API 0020 0050
ngkropingti Jion
NOTICE
This license information is applicable to the swagger documentation of the clinical-mdr-api, that is the openapi json.
License Terms (MIT)
Copyright (C) 2022 Novo Nordisk A/S, Danish company registration no. 24256790
Permission is hereby granted, fee of charge, to any person obtaining a copy of this software and associated documentation fles (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:
The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.
THE SOFTWARE IS PROVIDED 'AS IS' WITHOUT WARRANTY OF ANY KIND, EXPRESS OR MIPLED, INCLUDING BUT NOT UNTED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND GONNERINGEMENT IN O EVENT SHALL THE AUTHORS OF COPYRIGHT LODERS BE LIBLE FOR ANY CLAIM, DAMAGES OR OTHER LIBULITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARUSING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.
Licenses and Acknowledgements for Incorporated Software
This component contains software licensed under different licenses when compiled, please refer to the third-party-licenses md file for further information and full license texts.
Authentication
Supports OAuth2 Authorization Code Flow, at paths described in the OpenID Connect Discovery metadata document (whose URL is defined by the OAUTH_NETADATA_URL environment variable).
Microsoft Identity Platform documentation can be read (here).
Authentication can be turned off with OAUTH_ENABLED=fslse environment variable.
When authentication is turned on, all requests to API endpoints must provide a valid bearer (JWT) token inside the Authorization http header.
When authentication is turned off, all endpoints accept (optional) custom header X-Test-User-Id which allows any request to specify any user id value. If the X-Test-User-Id header is missing, the default value of unknown-user is assumed.
System information:
System information is provided by a separate System Information sub-app which does not require authentication.
Servers
lapi 🗸
ODM Study Events V
ODM Forms V
ODM Item Groups
ODM Item V
ODM Conditions ~
ODM Methods ~

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 shifts 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.

Select rows						00		00	Select rows							••••
n :	Q,						D	Subject)		٩,						
SNOWED ID	Preferret synonym	Preferred synorym (lower case)	Abbraviation	Definition	Status	Version	Modil		MED RT ID	Class name	Class name (lower case)	Abbreviation	Definition	Stature	Version	Modified
64572001	Doease	doese		Disease Idisorder)	(Hind)	1.8	344.2 M		N0000022185	ORAL HYPOGLYCEMIC	teal hypoglycenic agents		ORAL HYPOGLYCEMIC	(final)	1.0	Jul A. 2023, 3 PM
362965005	Disorder of body system	alsorder of body system		Disorder of body system ideoxiter1	(final)	1.0	34 4, 2 PM		N0000131730	PCK9 intellars	pold intimus		PCR9 inhaburs	Freed	1.8	jui 4, 2023, 1 PM
609964302	Pre-exciting type T diabetes melit	pro-existing type 1 diabetes mails		Pre-mining type 1 diabetes	(fise)	10	347.2 PM			zavennas	pitnesimuti		Zilvesmub	(114)	1.8	jul 4, 2023. PW
446331000	inset failure with control	head failure with normal election -	wints	metter_		10		1	N0000175945	STRycolaite	Letyroxine		i-thyroxine	(Fruit)	1.0	200 A, 2023, 294
	ejection			with monmail spectical	(198		P54	1	N0000175848	Tubuln inholding	tubule initiating agent		Tubulet insideing	(FAAL)	1.0	Jul 4, 2023,
442685003	Nonatohonic steatohepatitis	nonacoholic sonaturiepattila.	NASH	Nonacoholic solatotepatito	trut	1.0	544,2 PM		N0000180190	Thiazolidireditire	thispolidirectore		Thiatolidinesione	(Fruit)	1.0	Jul A. 2023.
441190003	Severe hereditary factor is anficie	severe hereditary factor of defice		Severe hereditary tacks	(find)	u	24.4.2 PM	.t.	NOODO175408	Sulfonyturea	ustonyluma		Suttoryturea	Final	1.0	jui 4, 2023. PM
				Di beficie					N0000175880	Sulfonamide	sufforarride		Sultanaride	(final)	1.0	3014, 2023
Dotomenes / Unique Inj	gredient Identifier	r) for Active Substances (D	. See	per page. <u>11 *</u>	1043		UCUI	Difference / 10	com ode for Units	of Measure) 🗇			Boos pre page	<u>8.7</u>	totalat c
Determine / (Unique In; Select rows	gredient Identifier	r) for Active Substances (D	-	per page. <u>0) *</u>	111428	0 3		Determenter / D A (Unified C Select rows	cum ode for Units	of Measure) 🗇			Bass per page	<u>a *</u>	Cost 31 C
Determiner / Unique In; adect rows	gredient Identifier Q	r) for Active Substances (Ð		pe page <u>0</u> *	00		Literary UCUI	Definitionerse 2 10 A (Unified C Select rows	ode for Units	of Measure) 🗇			Roos per page	<u>a *</u>	(1994)30 (C
Schmanes / Inique Inj lect rows	ne gredient identifier Q. Substance name	for Active Substances (Substance name flower cone)	3 Abtensistion	tum Satur	verpage <u>0</u> *	Nodiee	0 0	Linay UCUI	Definition 2 1 0 A (Unified C Select rows	cum ode for Units Q ucum	of Measure) ⑦	Status	Ver	Root per page	u *	139432 ×
Internets (Inique In; Inique In; Uner D: Otherange	ne gredient Identifier Q, Substance name sevonmkosse) for Active Substances (Substance name (source name) wompense	D Abbreviation	Status (rear)	Version 1.0	10023		Lineary UCUI Second 1	Difference 2 D A (Unified C Select rows UCOUR code gree	oue ode for Units o Q uccas gamp	of Measure) (2) I description er weit	Status (real)	- 200 - 200 - 20 - 200 - 20 - 2	Story	0 * Moon pi 4.3	4394/30 C
otomeree (nique Inj et rows uner to gatecrange unerotage	Q Q Substance name Labotimicouse NovelCouchy) for Active Substances (Substance name (sources) watersone namesaum	D Abbenväsion	Settin Settin (rear)	Perpape <u>10</u> * Werston 1.8 1.8	100239			Definition of a local A (Unified C Select rows UCUM costs gras scionuus)	cont ode for Units o Q, ucost grave grave grave	of Measure) (*) Indescription arr week	Status (nar) (nar)	Ver ta ta	nicu gani ka kata	9 * Modi 14 4 3 pr 4 4 3	4004/30 C
Determines (Unique Inj lect rows uner ito generocanes unerrows seeserver:	Q Q Sidetano name Labomicout Norisconov Questot) for Active Substances (Solutators name (source case) Wathyrane machanas genera	Abtenvision	Satur Rea Trans Trans	итаре: <u>()</u> * 	110429		Lineary UCUI Sector	Diffusione 2 U A (Unified C Select rows UCUM code gree sc.osu.ti (rsr)	COM Ode for Units of O, UCUM gravit p growth gravit do	of Measure) () Indescription Fund Fund Fund Fund Fund Fund Fund Fun	Status (nos) (nos) (nos)	19 19	Rome per page	9 * Mogi 1/4 2 1/4 3	1334/32 C
ettroven / nique inj ettroves userse gereorande usersolde seasstwe offenoviz	Q gredient Identifier Q Statsnonner Livomkowe worstower Gutsurot Statsrot Stat) for Active Substances (Substance name (source case) substance mechanics gamme supproces	D Abtensistion	Satur Tour Tour Tour Tour	Prop. 0 * Verser 18 18 18 18	111113 Modflet 244,2023		Linuy y UCUI Seet 1 1 1 1 1 1 1	Distances 7 2 2 4 (Unified C Select rows UCUM cells gm9 40.080.01 pt.87 pt.87 pt.87	COM Ode for Units of UCUM grave p grave p grave p states and ucum	of Measure) (2) Inferencies In	Status Teat Teat Teat	Ver 18 18 18		9 * Motil 344,3 344,3 344,3 344,3	 (394/30) (394/30) (4)
ett rows ett rows user rows csreactorize seekstrwic orteorioize seekstrwic orteorioize writichise	Q Section Identified Q Solution name LiberineCond Source) for Active Substances (Substance name lower cost) instrumer mechanism gamme stagen. substanter boreterer	Automation		Praye. 0 *	1 11 11 21 2	D C C C C C C C C C C C C C C C C C C C	Linuy y UCUI Seet 1 1 1 3 3 7	Distances / J J A (Unified C Select rows UCUM code gwe dCARMAC (UCUM code (UC	COM ode for Units Q UCOM gran gran gran gran gran gran gran gran	of Measure) (2) Identrytion er weik er weik	States (B) (B) (B) (B) (B) (B) (B) (B) (B) (B)	Ver 64 64 16 16	Rom yer piger	19 * Modi 204.3 204.3 204.3 204.3 204.3	 4 10 mil 20 4 20 mil 20 4 20 mil 20 4 20 mil 20 4 2 mil 20
Intervene of Inique Ing Inique Ing Inique Ing Intervene Unitervene Optimiser Seasestwick Optimiser Seasetwick Optimiser Seasetwick Optimiser Seasetwick Se	Control Identified C) for Active Substances (Solutions name Inser case) Insertymore methodes stagion stagion paratrand paratrand method	Automation	Salan Nation Nat	Werson 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0	1 11 11 21 2	2 2 2 2 2 2 2 2 2 2 2 2 2 2	Linuy, y UCUI Seet 1 1 1 3 3 7 1 3 3 7 1 3	Distancies 2 D A (Unified C Select rows UCUUI com gwe scionuli (scar) (s	COM Odde for Units - UCOM grave p graves	of Measure) (2)	Status (non) (non) (non) (non) (non) (non) (non) (non)	9000 100 100 100 100 100 100 100	Rom pre pope	19 * Mooti 104.3 104.3 104.3 104.3 104.3 104.3 104.3	4100 ef 30 C C C C C C C C C C C C C C C C C C
ett rows ett rows uner rows un	Cursting and a second se second second sec) for Active Substances (Soldstance name (source) wortpusse mechanism gaptime stappin. pastime pastime mathema adaptime	Automation	XAMM Nat Nat Nat Nat Nat Nat Nat Nat Nat Nat	Verson 18 18 18 18 18 18 18 18 18 18 18 18 18	1 111121		Liney UCUI Seet 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Distancies 2 D A (Unified C Select rows UCUM costs gwe sk0ext81 gwe sk0ext81 gways gways sk0ext81 sk0ext81 sk0ex	COM ode for Units - Q UCOM grave p graves	of Measure) ②	Status (nos)	1997 100 100 100 100 100 100 100 100 100 10		10 * Moon pr4,3 pr4,3 pr4,3 pr4,3 pr4,3 pr4,3 pr4,3 pr4,3 pr4,3	4 10 of 30 4 4 10 of 30 7
Detrovers (Jaique Inj Interrows Utell ID Q5180c3058 Utell ID Q5180c3058 SeasSetwer Q4Febriov72 H113C1558 SeasSetwer Q4Febriov72 H113C1558 SeasSetwer S150c3312N H1259H2055 Kriedrissop) for Active Substances (Substance name (nower cast) boarywore mentaum gowne substance gowne substance autorum method method method method method method	3 Abbreviation		Versen 16 16 16 16 16 16 16 16 16 16 16 16 16	1 111121		Larry UCUI I I I I I I I I I I I I I I I I I I I	Continuentes 2 / 0 4 (Unified C Select roops UCUM costs grow 4 (COST) 4 (COST)	COLU ode for Units of q UCDA grosp g g g g grosp g grosp g g g g g g g g g g g g g g	of Measure) (2) Identryston sund sund super per resulter super per resulter super per resulter storg und storg und	5000 (Re) (Re) (Re) (Re) (Re) (Re) (Re) (Re)			<u>мо</u> мол ри 4,3 ри 4,3	COURTER C C C C C C C C C C C C C C C <tr< td=""></tr<>

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 sematic 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 and activity instances with activity items as in Figure 11.

	De				100 100 100		• • •	NUT INCOME DE LAN	a 10888)											
- TREADER	~ ~		20		200 HEW 310			Hoja percocki be sere	() posses)							Difference / Activities /				
n Densiberr	Ubrar	y / Concepts / Ar														Billrubin Orine 🕤				
Process Overview	ACC	ivicies ()														Overview COSMoS YAML				
		t of Activities A	ctivities by Grouping	Activities instances	Requested Activitie	5										Name	Bilirubin Urine			
	0	B Select rows							00	0						Sentence Case Name	bilirubin urine			
Cancepts A	N	initial	٩													Definition				
Units																Activity instance class	CategoricFindin	5		
		Library	Activity group	Activity subgroup	Activity Name	Sentence case name	Abbreviation	Modified	Status							Abbreviation		Library		Sponsor
		Sponsor	Contraceptive Counselling	Contraceptive Counselling	Bilateral Tubal Occlusion	bilateral tubal occlusion		341 4, 2023, 2:54 PM	Real							ADaM parameter code	BILJU2	Topic code		BILIRUBIN_URINE
yntax Templates 🗸 🗸		Sponsor	Laboratory Assessments	Biomarkers	BleAdd	bie acid		34 4, 2023, 2:54 PM	final							Activity groupings	Activity grou	P	Activity subj	proup
Template Collections	1	Sponsor	Laboratory Assessments	Biochemistry	Blinden	bilinubin		3al 4, 2023, 2:54 PM	Final								Laboratory Asse	somenta	Biochemistry	
	1	Sponsor	AE Requiring	Pancreatitis	Distant Correct	dilated common bile duct		3al 4, 2023, 2:55	Final						1	Activity	Name	Definition		Library
		Sponsor	Laboratory	Biochemistry	Direct Minubio	direct bilirubin		34 4, 2023, 2:54	(Final)								Bilirubin			Sponsor
		Saursor	Assessments Elability Criteria	Elaibility Orberia	Electrony Collection	themative contentia meet	U	orary / Concepts /	Activities /	List of Activiti	ies / Bilirubin									
					Mat	1	E	lilirubin 💿								Activity items	Name	CT term name	Unit name	Activity item
	1	Sponsor	Additional Data	Galibadder Disease	Common Bile Duct	common bile duc		Overview COSM	IOS YAML								BILI	Total Bilirubin Measurement		test_name_cod
	< <u>-</u>	Sponsor	At Requiring	Gallbladder Disease	Galistone in Bile	galistore à bie duct											LD	Laboratory Data Domain		domain
								Name	0	arubin							URINALYSIS	Urinalysis		finding_categor
								Sentence Case Nar	me b	ilirubin							URINE	Urine		specimen
						```		Definition											-	
							$\mathbf{i}$	Abbreviation				Library	Sponsor							
								Activity groupings		Activity gro	up		Activity subgroup							
							×			Laboratory As	sessments		Biochemistry							
								Activity instances		Name	Definition	Activity	Topic code	ADaM parameter						
												class		code						
										Bilirubin (N)		NumericFinding	BIURUBIN_N_URINE	BILIU3						
										Bilirubin Urine		CategoricFinding	BILIRUBIN_URINE	BILIU2						
										Bilirubin, AE		NumericFinding	BILI	BILI						
										Requiring Additional										
										Data										
										Bilirubin,		NumericFinding	BILIRUBIN_SERUM	BILIS3						

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.

														Relationshi to UCUM Units	р
	9	Uy s	itudies	Library					ADD NEW STUDY	SELECT STUDY	🕸 🕜 🔺 andja	Z (NICOLAS DE SAIN	T JORRE)		
«		Librar	iry / (	Concepts / Units									_	And to CT	
About Library		Un	nits (	0										Units	
Process Overview				last route						/			000		
Code Lists	~	_	,	rectrows							/				
Dictionaries	ř	Sei	earch		Q				/	/ /			ō		
Activities	Â			Library 1	Name	Master unit	Display unit	Unit subsets	UCUM unit	CT Unit terms	Convertible unit	SI unit	US convention		
Units				Sponsor	keV	No	Yes			keV.	Yes	No	Yes		
CHFs															
Compounds				Sponsor	RATIO	No	Yes		(Ratio)	RATIO	Yes	No	Yes		
Syntax Templates	~			Sponsor	iU/kg	No	Yes	Dose Unit	[IU]/kg	IU/kg	Yes	No	Yes		
Template Instantiations	×			Sponsor	PA	No	Yes		/a	PA	Yes	No	Yes		
Template Collections	č														
Units Exchange Standards	Ĵ.			Sponsor	LB	No	Yes		[lb_av]	LB	Yes	No	Yes		
<b>1 0 0</b>				Sponsor	weeks	No	Yes	Age Unit, Study Time, Time Unit	wk	WEEKS	Yes	No	Yes		
				Sponsor	week	No	Yes	Age Unit, Study Preferred Time Unit, Study Time, Time Unit	wk	WEEKS	Yes	No	Yes		
				Sponsor	years	No	Yes	Age Unit, Study Time, Time Unit	a	YEARS	No	No	No		
				Sponsor	months	No	Yes	Age Unit	mo	MONTHS	No	No	No		
			:	Sponsor	hours	No	Yes	Age Unit, Time	h	HOURS	Yes	Yes	Yes 👻		
											Rows per page:	10 × 1-10 of 374	$\langle \rangle$		

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.

OPEN - OPEN					-		
	Ut Studies		ADD NE	W STUDY SELECT STUDY 😂	3 🕜 🚨 NDJZ (NICO	DLAS DE SAINT JORRE)	Sponsor can
«	Library / Concepts / Compounds						Sponsor can
About Library	Compounds 🔊						add every
Process Overview	Compounds Compound Alia:	505					
🔄 Code Lists 🗸 🗸							compound as
Dictionaries V	Select rows					🛨 🔽 🖪 🕙	, , , , , , , , , , , , , , , , , , ,
E Concepts		0					needed
Activities	Search						
Units	Sponsor Compound	Compound Name Is INN	Compound Number (long)	Compound Number (short)	Analyte Number	Definition Brand Name	
CRFs	. No	Metformin Yes					They are
Compounds							incy are
Syntax remplates	Yes	NPH.Insulin Yes			anTBD005		then
Template Collections	: Yes	<u>Olaparib</u> Yes					11 1 1 6
+ Data Exchange Standards	; Yes	Durvalumab Yes					available for
i≣ List ✓							avery Studies
-	Yes	AZD6738 No					every Studies
			•			•	
					Rows per page: 10	▼ 1-5 of 5 < >	

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.

	🗘 Studies 🚺 Library	ADD NEW	STUDY SELECT STUDY 🏟 🕜 🔺 N	IDJZ (NICOLAS DE SAINT JORRE)	
About Library	Library / Syntax Templates / Objective Templates / Parent <b>Objective Templates ③</b>				Templates
Process Overview     Code Lists     Dictionaries	Parent Pre-instance User Defined			<b>+ 7 8 ≥ ⊙</b>	are using 'parameters'
Concepts	Search Q			٥	(in Green)
Endpoint Templates Time Frame Templates Criteria Templates Activity Templates	Sequence number Indication or disorder	Objective category Confirmatory testing Not Applicable Yes	Parent template Modified 4 To compare the effect of Comparator on Activity/instance when something	Status Version	that will be replaced
Footnote Templates           Template Instantiations         V           Template Collections         V           Template Collections         V           Data Exchange Standards         V	E O Not Applicable	Not Applicable Yes	Time to first occurrence of jul 4, 2023, 3:08 MACE, a composite PM endpoint consisting of: CV death, nonfatal KM, nonfatal stroke, or hospitalization for unstable angina	(Final) 1.0	Study level
E Lue V	Add objective template pre-instantiation 0      or template parameter values      to campare the effect of methamic rotates is up insulat     methods         X + @ right insulat         Unformatted plain text representation         Xincome and informatter parameters         Xincome and informatter plain text representation	in diabetic retricepathy - catavant when something $X + \bigotimes^{(n)} \frac{diabetic c_m}{diabetic c_m} > X$ in diabetic retrinopathy - catavant when something	· 0	Index samplate  Converse  Converse  3	
	The compare one sheet of metorman readore to riph insular o	numbered reamonable in - calls acc much something			

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).

	PCIDE GUIDE SUT	IIIG AP SOTHIG MD									
3.4 Status Final	Effective 2021-11-2	date 0		tengile SDTM	ments v2.0						/
1.3										/	
L2 ( Cor	ments Demogra	phics Events Find	fings Findings	storie	ntiona Non-No	at Organism Identifier	s Belalad Records Rol	alad Specimens	elated Subjects	Subject Disease Milestenes Subject	tiements >
1.3 c 1	CP CP	CV DA	00 86	FT.	64 18	15 1.0	MB MI	MK MS	NV OF	PC PE PP	Q1 >
Laborator	r Test Results	Ordinal Descript 24 A finding microbio	don 5 domain thát cont Rogy or pharmacol	ains Taboratory b inetic data, which	st data such as he are stored in sepi	matology, clinical chem prate domains.	intry and urinalysis. This dom	ain does not include			
Orthnai											
	Name	Label	Data Type	Pale	Core	Codelist.	Described Value Domain	Implaiments	Value List.	Description	
1	Name	Label Study identifier	Data Type Char	Role Mercifier	Core	Codalist	Described Value Domain	implaiments STUDHD	Value List.	Description Unique Identifier for a study.	
1 2	STUDVID DOMAIN	Label Study Identifier Domain Abbreviation	Data Type Char Char	Role Identifier Identifier	Core Reg Reg	Codelist	Described Value Domain	impiaments STUDIO DOMAD	Value List	Description Unique identifier for a study. Two character abbreviation for the domain	
1 2.) 8.)	Name STUDVID DOSMIN USUBJD	Label Study Identifier Domain Abtreviation Unique Subject Identifier	Data Type Chie Chie Chie	Ruie Identifier Identifier Identifier	Core Reg Reg	Codelint	Described Value Domain	implaiments STUDING ROMAN USUBIQ	Value List	Description Unique identifier for a study. Two-character obtreation for the domain studies for all applications or submissions is product.	arross all solving the
1 2 3	Kame STLOND DOMUN USUBJD LBSED	Label Study Identifier Donaio Abbreviation Unique Subject Identifier Sequence Number	Data Type Char Char Char Char	Role Mercifier Mercifier Mercifier	Core Reg Reg Reg	Codellat	Described Value Domain	Implements STADMS ROMAN LOUGHD .SED	Value List	Description Unique Identifier for a study. Two-character addression for the domain therefore and to uniquely identify a unique studies for all approximes or submissions a graduit. Sequence number given to ensure singuen ensemble addressionals, Kay for any unit	across all solving the ess of subject unities.
1 2. 3. 4. 5.	Kame STLDND DOSKUN USUBJD LBSED LBSED	Label Study identifier Domain Abbrevation Unique Subject Identifier Sequence Number Group ID	Data Type Our Our Our Num Our	Ruie Mercifier Mercifier Mercifier Mercifier	Core Reg Reg Reg Reg Perm	Codellas	Described Value Domain	Emplements STUDING COMMIN LOUGHER LOUGHER LOUGHER	Value List	Description Unique Identifier for a shudy. The characer observation for the domain identifier and to imparty identifier of a subjec- graduate. In all optimises on administration of the impact of all optimises on administration forgament number of the allocat in totaled into damain for a subject.	across all solving the ess of subject unitier, ds in a single
1 2) 3 4 5 6	Name STUDID DOMIN USUBD LBR/D LBR/D	Label Study identifier Domain Albrevallen Umger Subject identifier Seguence Number Orsug ID Specimen ID	Data Type Our Our Our State Our Our	Ruie Klentifer Klentifer Klentifer Klentifer	Eore Req Req Req Req Perm Perm	Codellast	Described Value Demain	Implements IT-Ump COMM COMM COMM COMM COMM COMM COMM COM	Value List	Description unages detected for a shulp; Tare-character adheridation for the downars shuder for all approximate sensitivity a subje- minute. Sequence number game to enumeration under the all approximates or undersensing instant. Sequence number game to enumeration data and an approximates and and sensitivity and approximates and approximates data and approximates and approximates and approximates and approximates and approximates and approximates and approximates and approximates and approximates and approximates and approximates and approximates and approximates and approximates and approximates and approximates and approximates and approximates and approximat	across all incluing the ess of subject unities; ds is a single esple; specimes
1 2 3 4 5 4 7	FIRIND TRIND TRIND COMMA TRIND COMMA TRIND COMMA	Label Label Study Identifier Doman Alterwalten Umgas Salgert Identifier Seguence Number Groug ID Specimen ID Specimen ID Specimen ID	Data Type Our Our Our Our Our Our Our	Ruie Mercher Kercher Kercher Kercher Kercher Kercher	Eore Reg Reg Reg Perm Perm Perm	Codeline	Described Value Demain	Implements STURING SOMAN USURIO -SEO -SEO -SEO -SEO	Value Lint	Description Unsame the description Unsame the description The description of the density Interference on the sense Interference on the sense of the density Interference on the sense of the density Interference on the density of the density Interference on the density of the density Interference on the density	across all solving the ess of subject sumber, ds is a single mple: specim preprinted o ed in the

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

### THE OPENSTUDYBUILDER COME WITH VISUALISATION TOOL: NEODASH

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).



# DIGITAL DATA FLOW ADAPTOR

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.

The TransCelerate DDF envisions a future state of fully automated, dynamic study start-up readiness, which is also the goal of the OpenStudyBuilder. Within the context of DDF, the OpenStudyBuilder serves as a Study Definition Repository (SDR).

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.



Figure 17: From OpenStudyBuilder to Digital Data Flow example

# 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, SDTM define.xml specifications and more. This metadata can be cleverly connected, creating a concrete platform that heavily utilizes biomedical concepts to enhance end-to-end processes for 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: <u>https://openstudybuilder.com/</u>

2. Project Resources, if you would like to see various information resources available, you can check out the resource-list: <u>https://openstudybuilder.com/info_resources/</u>

3. OpenStudyBuilder demonstration video, we have a 30' demonstration video available showing the functionality of the OpenStudyBuilder in version 0.4: <u>https://www.youtube.com/watch?v=dL5CY0BwfEs</u>

4. GitLab repository, the complete solution is available in GitLab. A docker setup is available as well: <a href="https://gitlab.com/Novo-Nordisk/nn-public/openstudybuilder/OpenStudyBuilder-Solution/">https://gitlab.com/Novo-Nordisk/nn-public/openstudybuilder/OpenStudyBuilder-Solution/</a>

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-Jvrhtmy7hGstgyilvB1Wsw

# **CONTACT INFORMATION**

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: <u>OpenStudyBuilder@gmail.com</u> OpenStudyBuilder Slack: <u>https://join.slack.com/t/openstudybuilder/shared_invite/zt-19mtauzic-Jvrhtmy7hGstgyilvB1Wsw</u>

Brand and product names are trademarks of their respective companies.