



# cdisc® 2026 Europe Interchange

THE FUTURE IS CONNECTED: STANDARDS AND AI POWERING DIGITAL TRANSFORMATION



MILAN, ITALY | MAIN CONFERENCE: 20-21 MAY | TRAININGS & WORKSHOPS: 18, 19, & 22 MAY

**Buy One – Get Three! Rule-Based Naming and Standardization of SoA, SDTM, and ADaM Metadata**  
Martin Gram, Associate Standards Developer Director, Novo Nordisk



# Speakers

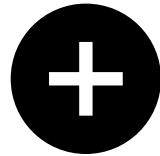


**Martin Gram**

Associate Standards Developer Director  
Novo Nordisk



Add new metadata





## We want metadata that is:

- Defined once, used many times, without manual editing
- Suitable across use cases (data and documents)
- Aligned to external standards, don't reinvent the wheel
- Support automation
- Fast to create
- High quality
- Easy to maintain

Concept Based!

Type : \_\_\_\_\_

----- **not connected**

Value: [MISSING]

\_\_\_\_\_

Data ID \_\_\_\_\_

----- **not curated** !

DETAILS null

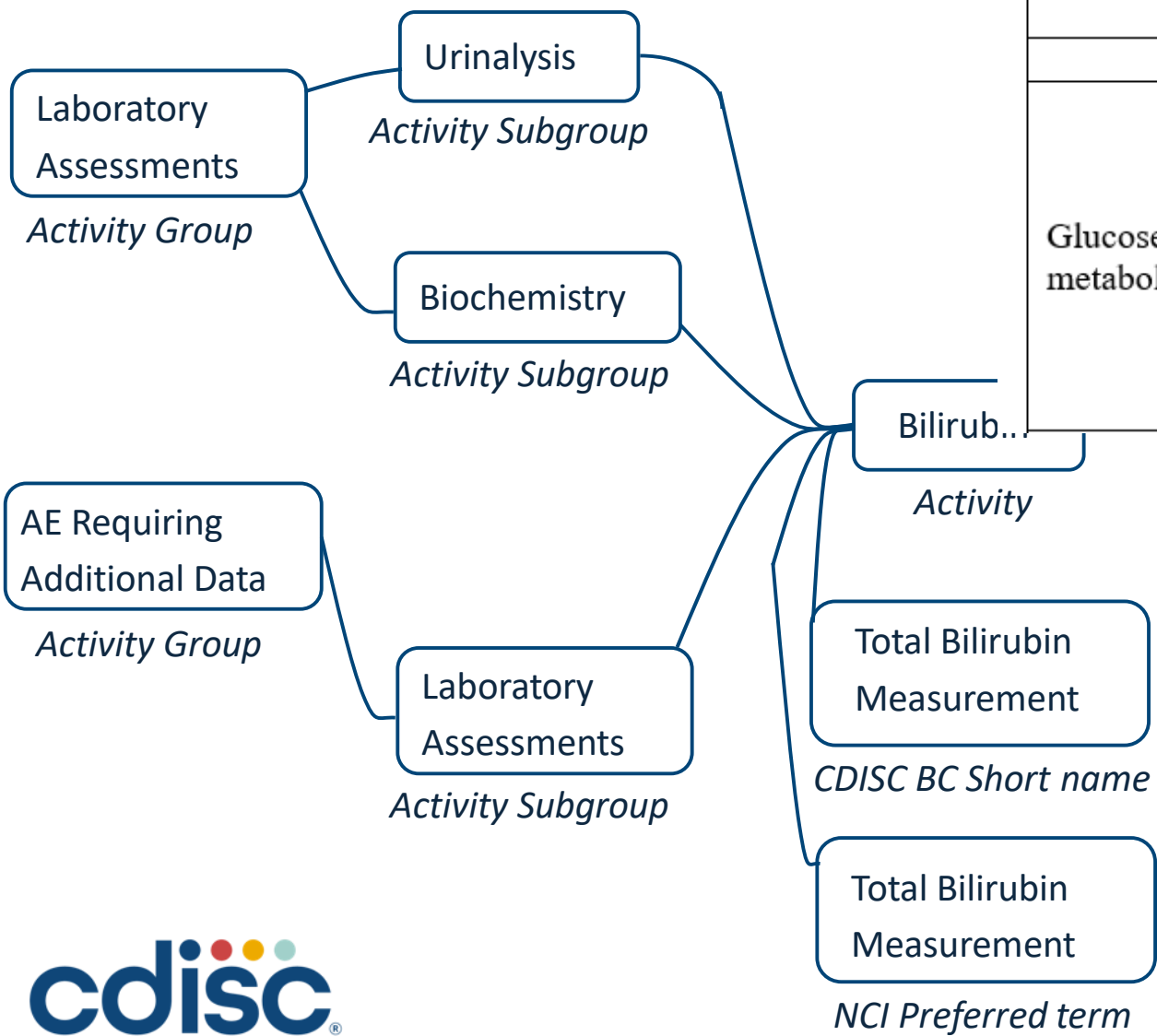
----- **not suitable for pooling data** !

Source: UNKNOWN

tag : ??? / ??

Value 2 0.00

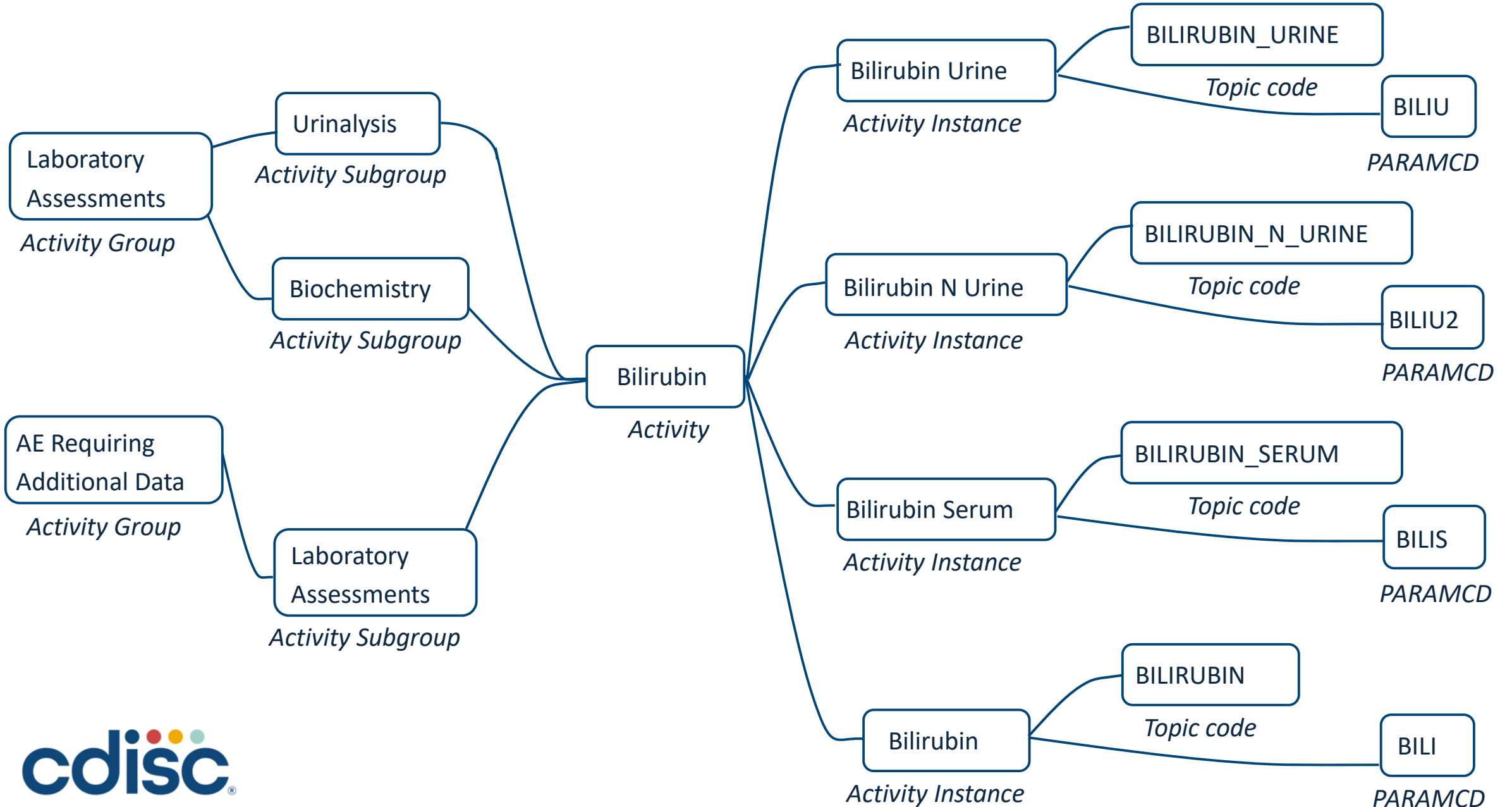
# OpenStudyBuilder Activity Concept - Documents



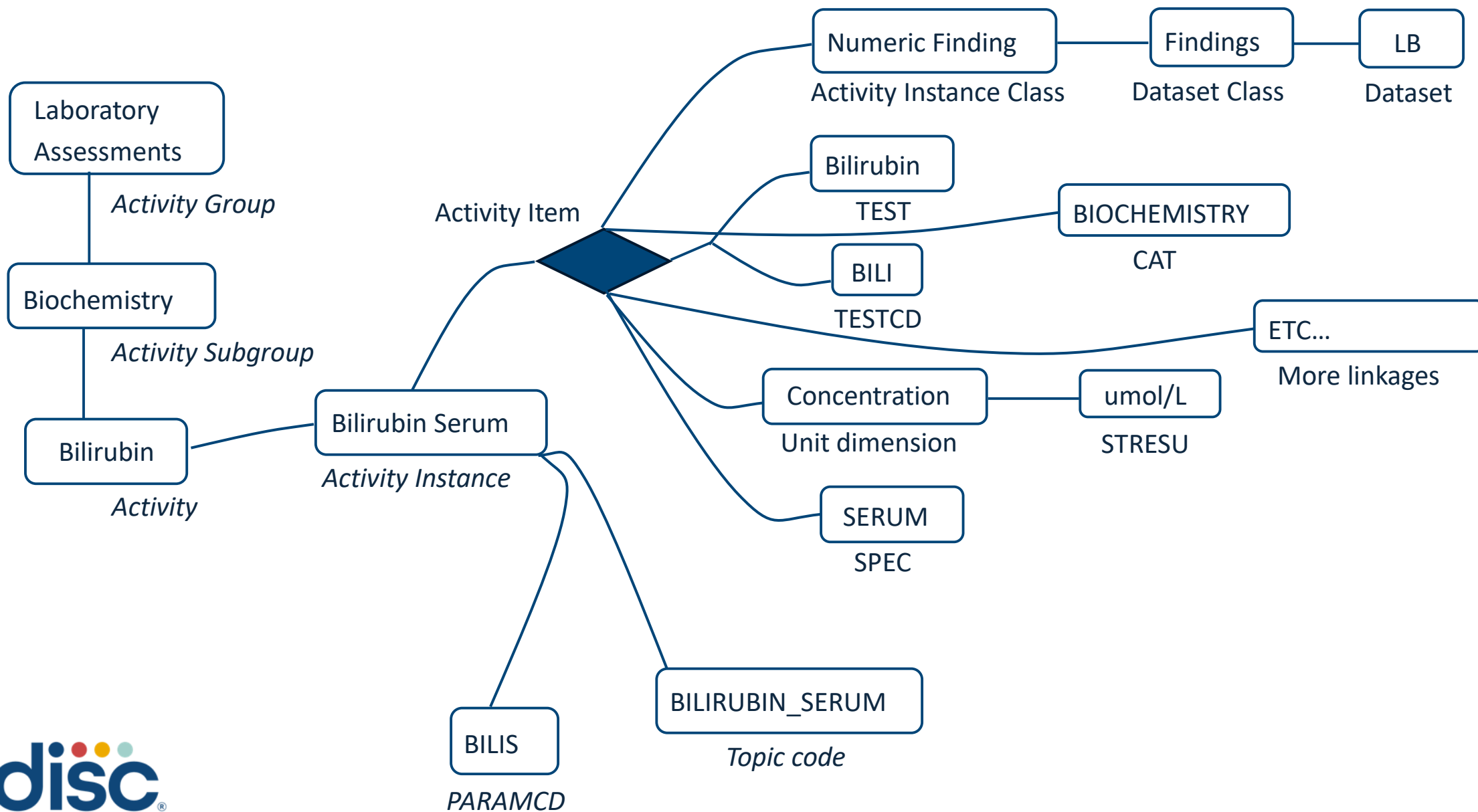
**Table 10-1 Protocol-required efficacy laboratory assessments**

	Parameters			
	Term ↑↓	Source ↑↓	Term Type ↑↓	6
Glucose metabolism	BILI	CDISC	PT	
	Bilirubin	CDISC	PT	
	Bilirubin	CDISC	SY	
	Total Bilirubin	CDISC	SY	
	Total Bilirubin	GDC	PT	
	Total Bilirubin	PCDC	PT	
	Total Bilirubin Level	CPTAC	PT	
	Total Bilirubin Measurement	caDSR	SY	
	Total Bilirubin Measurement	NCI	PT	
	Total Bilirubin Measurement		(Preferred_Name)	

# OpenStudyBuilder Activity Concept - Data



# OpenStudyBuilder Activity Concept - Data





Demo CodeX



# Demonstration Screenshot

The screenshot shows a web browser window with three tabs: 'Data', 'OpenStudyBuilder', and 'CodeX Demo'. The address bar shows 'localhost:8501'. The browser's bookmark bar contains several items, including 'DTS links', 'Gen AI', 'HR', 'SB', 'Prediction', 'Info', 'PST Site', 'DTS', 'Tracker', 'Data', 'Daily Work', 'Repos', 'C2831 - Glucose', 'Submit Access Req...', 'CodeX', and 'KCS Knowledge Ba...'. The application interface is divided into a dark blue sidebar on the left and a main white content area on the right. The sidebar contains the following elements: a 'CodeX' header, a 'DS Chat' section, and three buttons: '1- Search CDW and OSB', '2- Create Topic Codes', and 'Report Bugs & Suggestions'. The main content area features the CodeX logo (a blue square with a white 'C' and the word 'CONNECT' below it) and the text 'CodeX'. Below the logo is a search section titled 'Search CDW and OSB for existent assessments'. It includes a sub-instruction: 'Use the assessment label/name or key words in the assessment name'. A search input field contains the text 'oblizafovir'. Below the input field is a dark blue button with a magnifying glass icon and the text 'Search CDW & OSB'. A light blue message box below the button reads: 'If you can't find a suitable match in OSB or CDW, you can move to the next step (2- Create Topic Codes)'. A yellow warning box with a triangle icon contains the text: 'No matches found in OSB'. At the bottom of the search section, a message reads: 'Could not find a match for: oblizafovir'. In the top right corner of the application area, there is a 'Deploy' button with a dropdown arrow.

# Demonstration Screenshot

The screenshot displays the CodeX web application interface. The browser tabs include 'Data', 'Databricks', 'OpenStudyBuilder', and 'CodeX Demo'. The address bar shows 'localhost:8501'. The sidebar on the left contains the following elements:

- CodeX
- DS Chat
- 1- Search CDW and OSB
- 2- Create Topic Codes
- Topic codes
- Select SDTM Domain
  - AE Requiring Additional Data
  - EC Domain
  - IS Domain
  - LB Domain
  - PC Domain
  - ZE Domain
  - Other Domains
- Report Bugs & Suggestions

The main content area features the CodeX logo and a description: "The tool provides the most common topic code types along with their typical attributes, but you are responsible for performing quality control and making any necessary modifications based on the specific study requirements." An important note states: "IMPORTANT: if you can't identify the NNC number or compound name please raise a request to The Rosetta Help Desk".

The 'Input Details' section includes a dropdown menu for 'Screening' and a list of test types: Cross Reactive Endogenous, Neutralising Endogenous, Neutralising Drug, IgE Antibody, IgA Antibody, IgM Antibody, and IgG Antibody. A 'Trial ID or GST/PST ID' field is also present.

The 'Which field do you want to select by?' section has radio buttons for NNC Number (selected), Analyte ID, and INN ID. Below this is a 'Select NNC Number' dropdown menu with 'NNC1001-0101' selected.

The 'Selected Record:' section contains a table:

	Feature	Value
0	NNC Number	NNC1001-0101
1	Analyte	AN100101
2	Inn ID	veloxarin
3	Acquired Compound/Name	Veloxarin A1

Below the table, there is a note: "You can find information about analytes in the following file (sheet: Analytes): [Analytes Specification](#)". Further instructions are provided: "NNC Number: Column A or D Analyte: Column B", "INN ID: Column K", and "Acquired Compound ID: Column A". A link for "Specific information about INN ID: [FDA Global Substance Registration System](#)" is also included.

A 'Generate Code' button is located at the bottom of the form.

# Demonstration Screenshot

The screenshot displays the CodeX web application interface. The browser tabs include 'Data', 'Databricks', 'OpenStudyBuilder', and 'CodeX Demo'. The address bar shows 'localhost:8501'. The sidebar on the left contains the following elements:

- CodeX
- DS Chat
- 1- Search CDW and OSB
- 2- Create Topic Codes
- Topic codes
- Select SDTM Domain
  - AE Requiring Additional Data
  - EC Domain
  - IS Domain
  - LB Domain
  - PC Domain
  - ZE Domain
  - Other Domains
- Report Bugs & Suggestions

The main content area features the CodeX logo and a description: "The tool provides the most common topic code types along with their typical attributes, but you are responsible for performing quality control and making any necessary modifications based on the specific study requirements." An important note states: "IMPORTANT: if you can't identify the NNC number or compound name please raise a request to The Rosetta Help Desk".

The 'Input Details' section contains the following fields:

- Select a test type: IgE Antibody
- Unit: % B/T
- Specimen: Plasma
- User Initials: user1
- Trial ID or GST/PST ID: trial1

The 'Which field do you want to select by?' section has radio buttons for NNC Number, Analyte ID, and INN ID (selected). Below this is a 'Select INN ID' dropdown menu with 'oblizafovir' selected.

	Feature	Value
0	NNC Number	NNC1035-3535
1	Analyte	AN103535
2	Inn ID	altivecimab
3	Acquired Compound/Name	Altivecimab A135

Additional information provided at the bottom of the form:

- You can find information about analytes in the following file (sheet: Analytes): [Analytes Specification](#)
- NNC Number: Column A or D Analyte: Column B
- INN ID: Column K
- Acquired Compound ID: Column A
- Specific information about INN ID: [FDA Global Substance Registration System](#)

A 'Generate Code' button is located at the bottom of the form.

# Demonstration Screenshot

The screenshot displays the OpenStudyBuilder web application interface. The browser address bar shows the URL: <https://studybuilder.clinicalmdr-tst.corp.azure.novonordisk.com/library/activities/activity-instances>. The application header includes the 'OPEN STUDY BUILDER' logo, 'TST' branding, and navigation tabs for 'Studies', 'Library', 'Administration', and 'Reports'. A 'SELECT STUDY' dropdown menu is visible on the right.

The left sidebar contains a navigation menu with the following items: About Library, Code Lists, Dictionaries, Concepts, Activities (selected), Units, Data Collection Standards, Syntax Templates, Template Instantiations, Overview Pages, Data Exchange Standards, and Admin Definitions.

The main content area is titled 'Activities' and shows a breadcrumb trail: Library / Concepts / Activities / Activity Instances. Below the title, there are tabs for 'Activities', 'Activity Groups', 'Activity Subgroups', 'Activities by Grouping', 'Activity Instances' (selected), 'Requested Activities', 'Activity Instance Classes', and 'Activity Item Classes'. A filter bar includes buttons for 'All', 'Final', 'Retired', and 'Draft'.

A search bar and a 'Select rows' toggle are located above the table. The table below displays a list of activity instances with the following columns: Library, Activity group, Activity subgroup, Activity, Activity instance class, Activity Instance ↑, NCI Concept ID, NCI Concept Name, Research Lab, and Molecu. The first row of data is as follows:

Library	Activity group	Activity subgroup	Activity	Activity instance class	Activity Instance ↑	NCI Concept ID	NCI Concept Name	Research Lab	Molecu
Sponsor	Laboratory Assessments	Antibodies	Anti-Oblizafovir IgE Antibody	GeneralObservation	Anti-Oblizafovir IgE Antibody Plasma			Yes	

At the bottom of the page, there is a pagination control showing 'Rows per page: 10' and '1-1 of 1'.

# Demonstration Screenshot

The screenshot displays the CodeX web application interface. On the left is a dark blue sidebar with navigation options: 'CodeX', 'DS Chat', and two buttons: '1- Search CDW and OSB' and '2- Create Topic Codes'. Below these are 'Topic codes' with radio buttons for 'Select SDTM Domain' (AE, EC, IS, LB, PC, ZE, Other) and a 'Report Bugs & Suggestions' button.

The main content area features the CodeX logo and a 'Deploy' button. Below the logo is an 'Input Details' section with a light blue background containing the text: 'This tool helps you to create codes for laboratory (LB) assessments based on NCI Thesaurus concepts.' Below this is a text input field with 'adenosine diphosphate' and a 'Definition provided by the trial team' label.

To the right of the input field are four buttons: '1- Search Concept in CDW and OSB', '2- Search Concept in NCI' (highlighted with a mouse cursor), '3- Search Concept using AI', and '4- Use your own definition and description'.

Below the buttons is a search progress bar: 'Searching NCI database for: 'adenosine diphosphate...''. A green bar indicates 'Found 10 excellent matches (score: 10.0/10)'. The 'NCI Search Results - Top 10 Concepts' section shows the first result: '1. Adenosine Diphosphate Measurement' with code 'C182257'. It includes definitions for CDISC and NCI, and a note about 5 synonyms.

An 'Actions:' section at the bottom right contains a button 'Select NCI Term' with a green checkmark icon.



# Demonstration Screenshot

The screenshot displays the OpenStudyBuilder web application interface. The browser tabs include 'Data', 'Databricks', 'OpenStudyBuilder', and 'CodeX Demo'. The address bar shows 'localhost:8501'. The application has a dark blue sidebar on the left with the following elements:

- CodeX**
- DS Chat**
- Buttons: '1- Search CDW and OSB', '2- Create Topic Codes'
- Topic codes**
- Select SDTM Domain
  - AE Requiring Additional Data
  - EC Domain
  - IS Domain
  - **LB Domain**
  - PC Domain
  - ZE Domain
  - Other Domains
- Report Bugs & Suggestions

The main content area shows a form titled 'Fill in Additional Details' for a finding. The finding name is 'Adenosine Diphosphate Measurement' with source 'NCI' and type 'FULL\_SYN'. The SDTM Domain is set to 'LB'. The form includes the following fields:

- Research: No
- Sub-Category: (empty)
- Activity Group: Laboratory Assessments
- Specimen: Blood
- Flowchart Group: Efficacy
- Existent activities sub-groups: Biochemistry
- Finding Type: Numeric Finding
- Assessment Group: (empty)
- User Initials: user1
- Unit provided by the trial team: umol
- Assessment Sub-Group: (empty)
- Trial ID: trial1
- Molecular Weight: (empty)
- Category: Biochemistry

At the bottom of the form, there is a 'Generate Codes' button with a green checkmark and a mouse cursor hovering over it. Below the form, a status bar indicates 'Using CDISC domain: LB' with a green checkmark.

# Demonstration Screenshot

The screenshot displays the CodeX application interface. On the left is a dark blue sidebar with the following elements:

- CodeX** (header)
- DS Chat** (header)
- Buttons: "1- Search CDW and OSB", "2- Create Topic Codes"
- Topic codes** (header)
- Section: "Select SDTM Domain" with radio buttons for:
  - AE Requiring Additional Data
  - EC Domain
  - IS Domain
  - LB Domain
  - PC Domain
  - ZE Domain
  - Other Domains
- Button: "Report Bugs & Suggestions"

The main content area is titled "Input Details" and contains:

- Introductory text: "This tool helps you to create codes for laboratory (LB) assessments based on NCI Thesaurus concepts."
- Input field: "Definition provided by the trial team" with the value "cytochrome p450 activity".
- Four search options in a list:
  - 1- Search Concept in CDW and OSB
  - 2- Search Concept in NCI
  - 3- Search Concept using AI
  - 4- Use your own definition and description
- Search progress indicator: "Searching scientific resources for: 'cytochrome p450 activity'..."
- Warning message: "AI-generated suggestions based on your query: 'cytochrome p450 activity', only use the AI suggestions if you can't find a suitable term in NCI"
- Section: "Below are AI-generated suggestions:"
- Suggested Alternative Terms** section:
  - Text: "These are AI-generated suggestions based on your query: 'cytochrome p450 activity'"
  - 1. Cytochrome P450 Activity Measurement**
    - Description: "The measurement of the biological activity of cytochrome P450 enzymes, which are critical for the oxidative metabolism of various substrates, including drugs and endogenous compounds. This measurement is performed under specific conditions that may include the presence of substrates and cofactors necessary for enzyme activity. The activity is typically assessed using spectrophotometric methods that quantify the conversion of substrates into products, reflecting the metabolic capacity of the enzymes. This measurement is performed in a biological specimen."
    - Source: "Clinical practice in pharmacology and toxicology."
    - Button: "Select AI Suggestion" (checked)
  - 2. Cytochrome P450 2C9 Activity Measurement**
    - Description: "The determination of the biological activity of cytochrome P450 2C9, an enzyme involved in the metabolism of numerous therapeutic agents, assessed through specific substrate assays that reflect its enzymatic function. The measurement is conducted under controlled conditions, often involving the incubation of the enzyme with its substrates and subsequent analysis of product formation using spectrophotometric techniques. This measurement is performed in a biological specimen."
    - Button: "Select AI Suggestion" (checked)

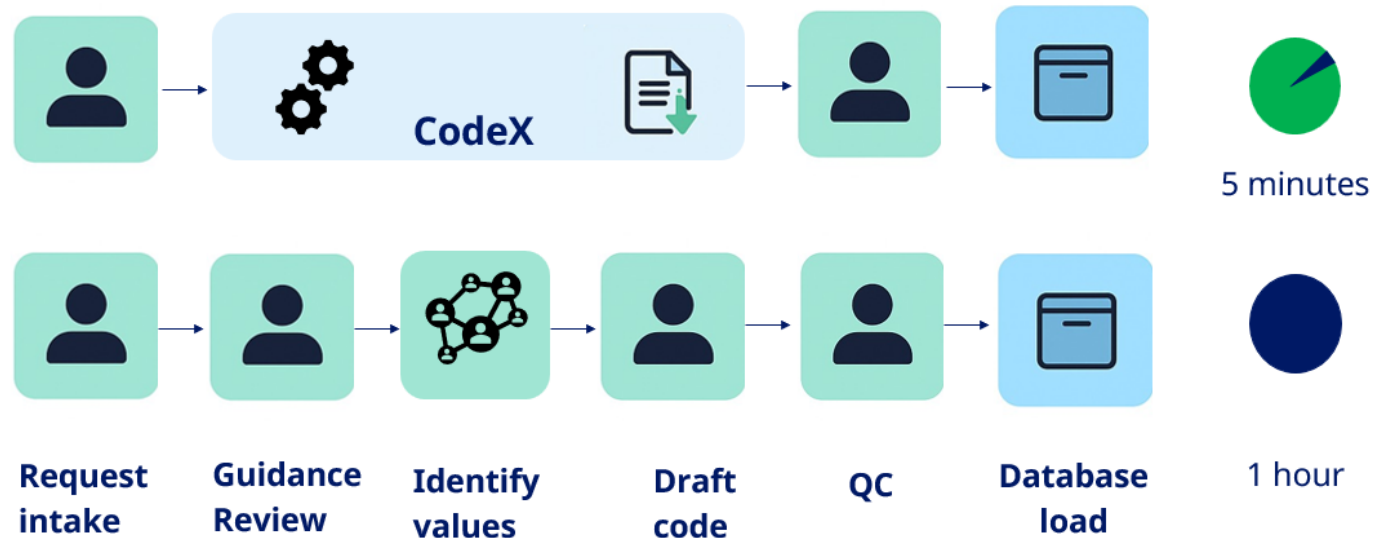
# Demonstration Screenshot

The screenshot displays the CodeX web application interface. On the left is a dark blue sidebar with navigation options: 'CodeX', 'DS Chat', '1- Search CDW and OSB', '2- Create Topic Codes', 'Topic codes' (with radio buttons for 'AE Requiring Additional Data', 'EC Domain', 'IS Domain', 'LB Domain', 'PC Domain', 'ZE Domain', and 'Other Domains'), and 'Report Bugs & Suggestions'. The main content area is light gray and contains two data tables. The top table lists properties for 'Cytochrome P450 Activity Measurement', including 'name', 'abbreviation', 'definition', 'library\_name', 'is\_data\_collected', and 'is\_multiple\_selection\_allowed'. Below this table is a green 'Export Activity to OSB' button. The second table, titled 'Activity Instance', shows a single record with fields like 'Field', 'Record 1', 'Trial ID', 'CodeX user', 'name', 'topic\_code', 'abbreviation', 'adam\_param\_code', 'test\_code', 'test\_name', 'specimen', 'unit', 'definition', 'is\_research\_lab', 'library\_name', and 'activity\_instance\_class\_uid'. Below this table are buttons for 'Export Activity Instance to OSB' and 'Download Activities & Instances'. At the bottom right of the main content area is a large green button labeled 'Export Activity & Instance to OSB' with a mouse cursor pointing to it.

Field	Record 1
Trial ID	trial1
CodeX user	user1
name	Cytochrome P450 Activity Measurement Blood
topic_code	CYTOCHROME_P450_ACTIVITY_BLOOD
abbreviation	CYTP45B
adam_param_code	CYTP45B
test_code	CYTP45B
test_name	Cytochrome P450 Activity Measurement
specimen	Blood
unit	umol/kg/min
definition	The measurement of the biological activity of cytochrome P450 enzymes, which are critical for the oxidative metabolism of various substrates, including
is_research_lab	False
library_name	Sponsor
activity_instance_class_uid	

# CodeX

- Focuses attention on the relevant details
- Reduces reliance on manual input
- Leverages predefined rules and AI
- Aligned with industry standards
- Automates the mechanical, enforces consistency and connects workflows

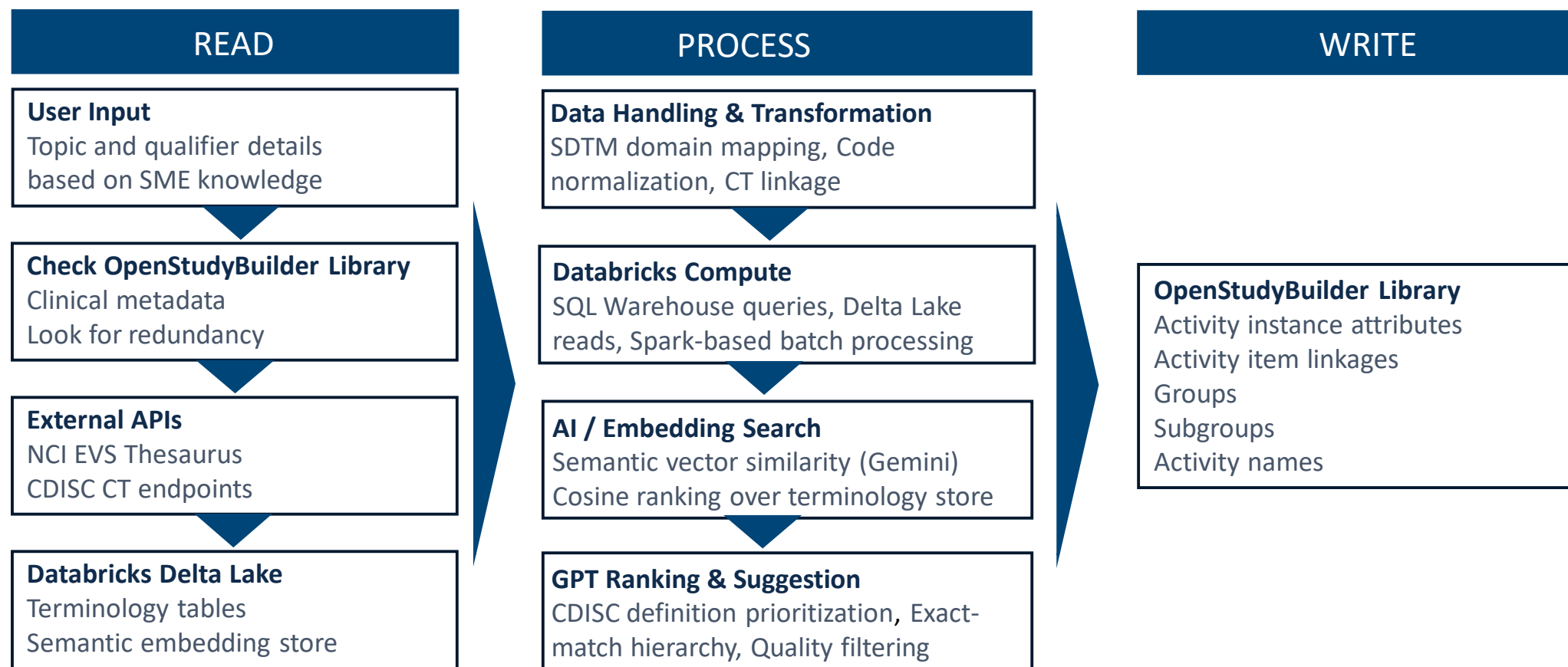


# CodeX outside OpenStudyBuilder?

- Companies have custom business rules and logic when creating metadata
- When standards evolve updates to CodeX are faster
- CodeX could become part of the OpenStudyBuilder extensions framework



# CodeX Application Architecture



# CodeX AI Components

## LLM

### GPT-4 Code Generation

- Generates submission values, topic codes and labels from clinical concept names
- Tailored prompts per SDTM domain
- Converts free-text descriptions to structured SDTM output

## EMBEDDINGS

### Semantic Embeddings

- Clinical terms stored as vectors in Databricks Delta Lake
- Finds semantically similar concepts - not just exact matches
- Powers search across thousands of existing codes

## SEMANTIC SEARCH

### NCI Thesaurus Search

- Queries NCI EVS API with abbreviation expansion
- Multi-word clinical phrases preserved as single query
- CDISC TESTCD extracted directly from NCI metadata

## RANKING

### Result Ranking

- Exact matches ranked first, semantic matches below
- CDISC definitions prioritised over generic NCI definitions
- GPT abbreviates only when length limits are exceeded

## AUTOMATION

### Code & Label Automation

- GPT selects medically relevant terms from user input
- Medical abbreviations generated and length-capped automatically
- Short labels use trusted dictionary first, GPT as fallback



# Acknowledgements

Angelica Prado for her major contribution to the technical implementation of CodeX and her continued input to improving the business logic





**Questions?**

