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

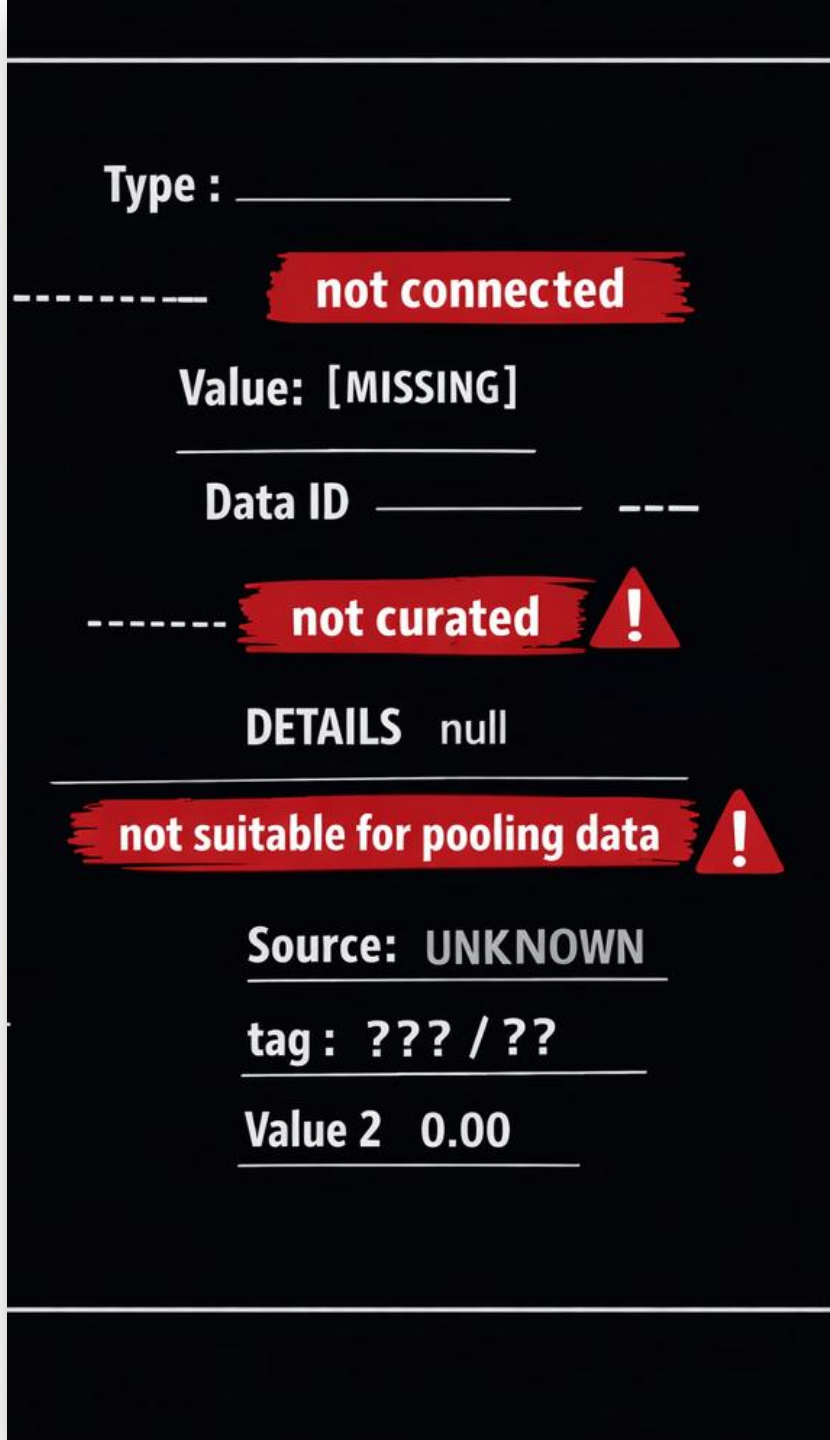
Martin Gram
Associate Standards Developer Director
Novo Nordisk

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!



OpenStudyBuilder Activity Concept - Documents

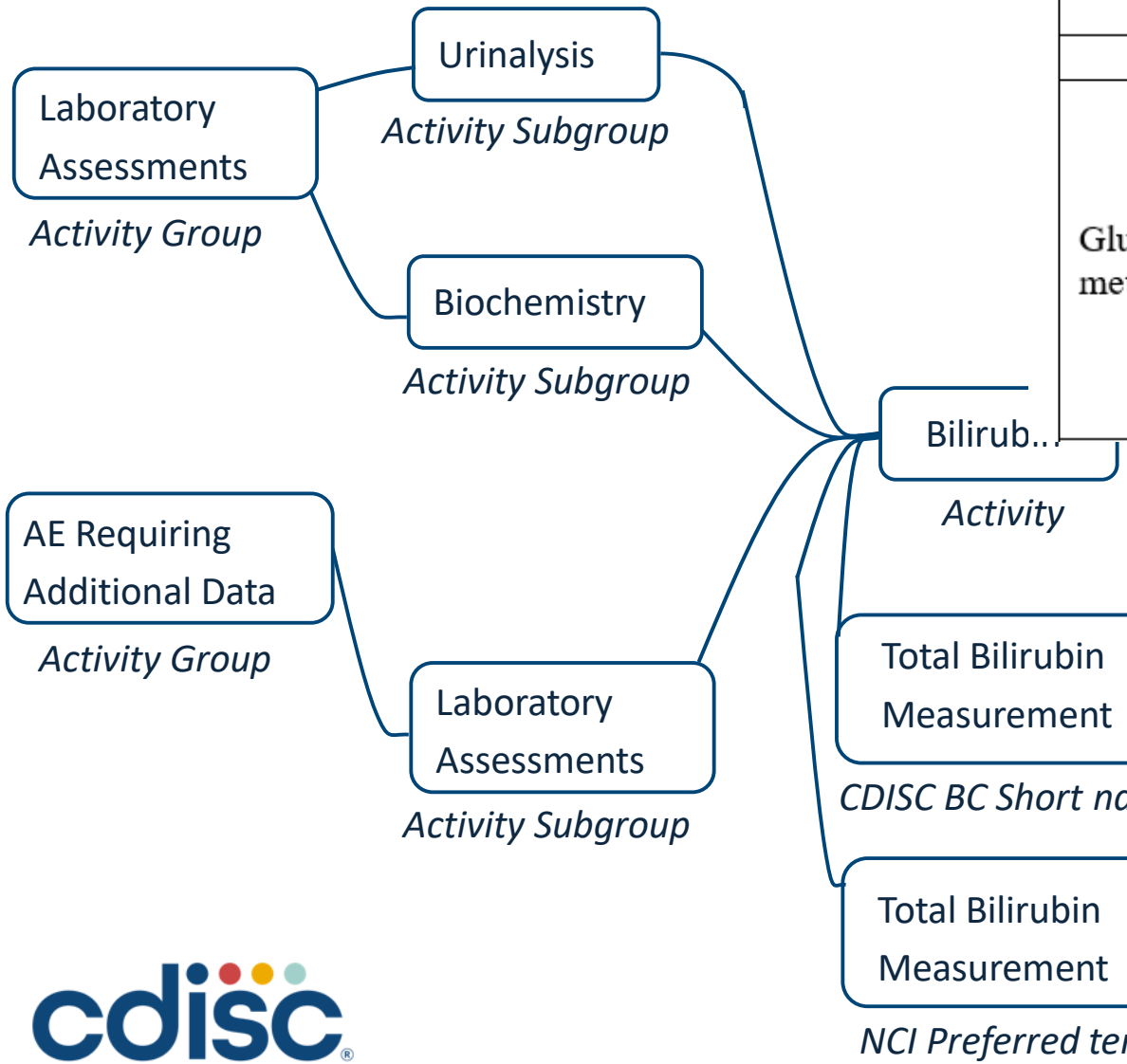
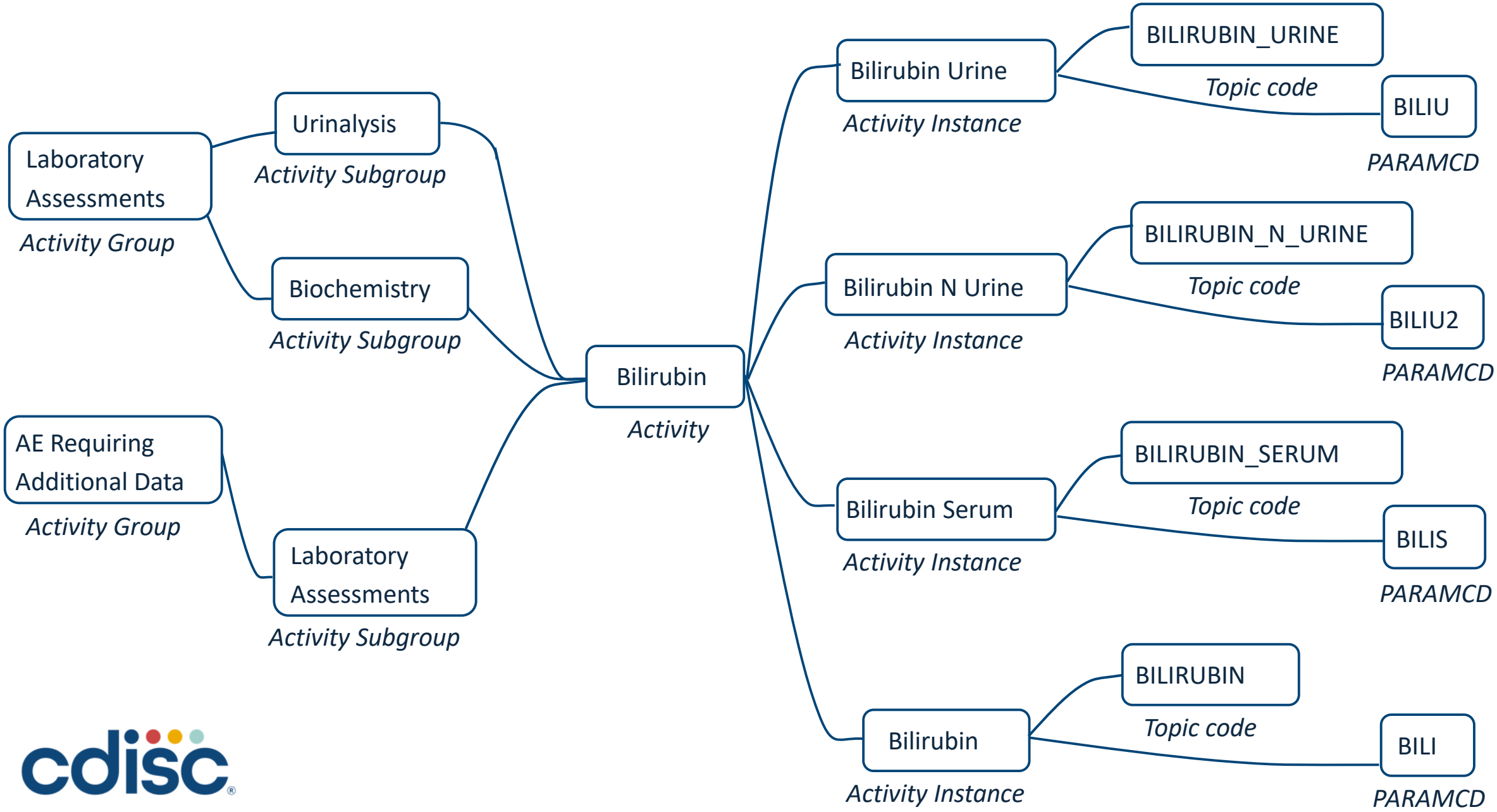


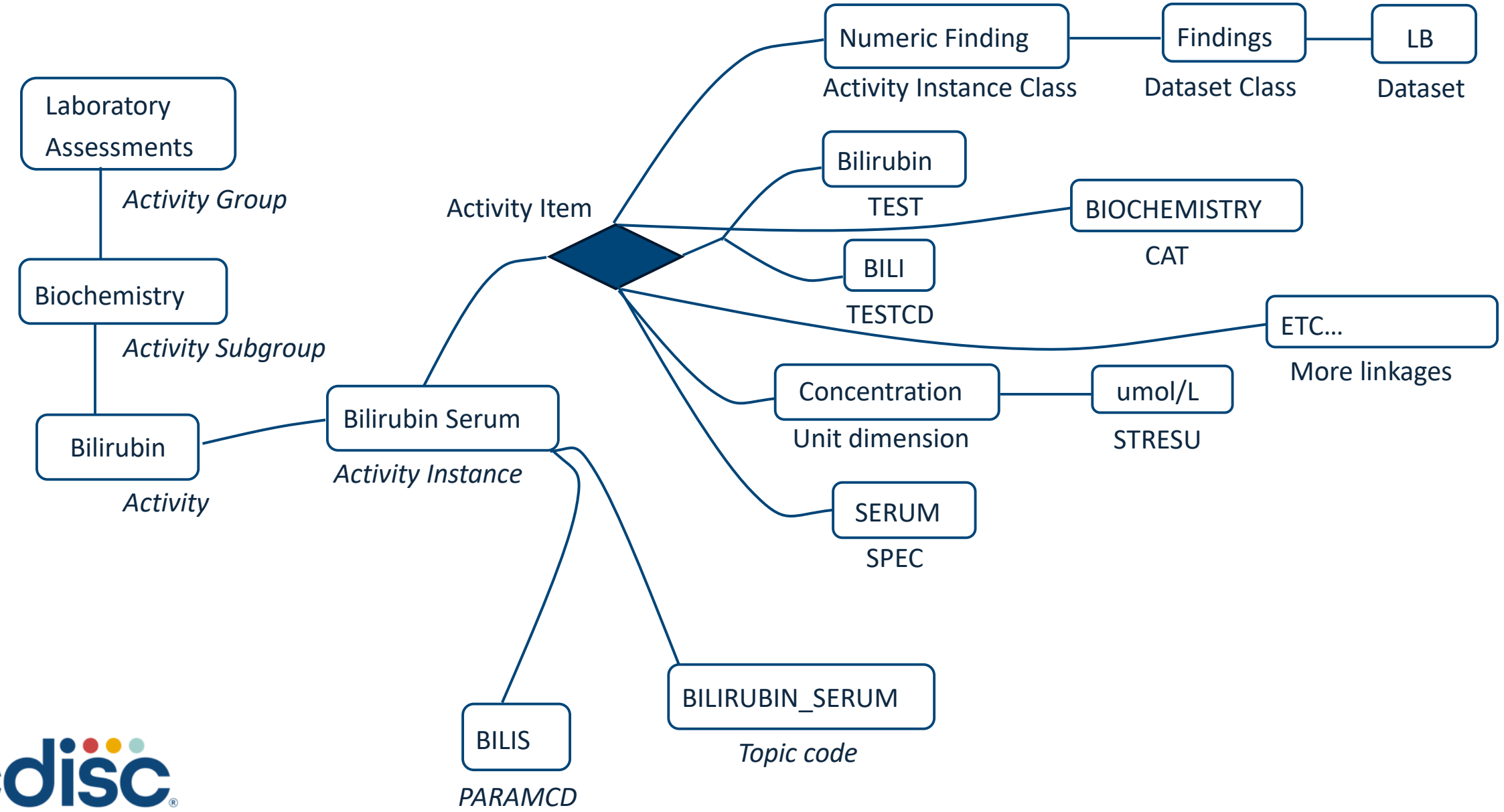
Table 10-1 Protocol-required efficacy laboratory assessments

	Parameters		
	Term ↑↓	**Source** ↑↓	**Term Type** ↑↓
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

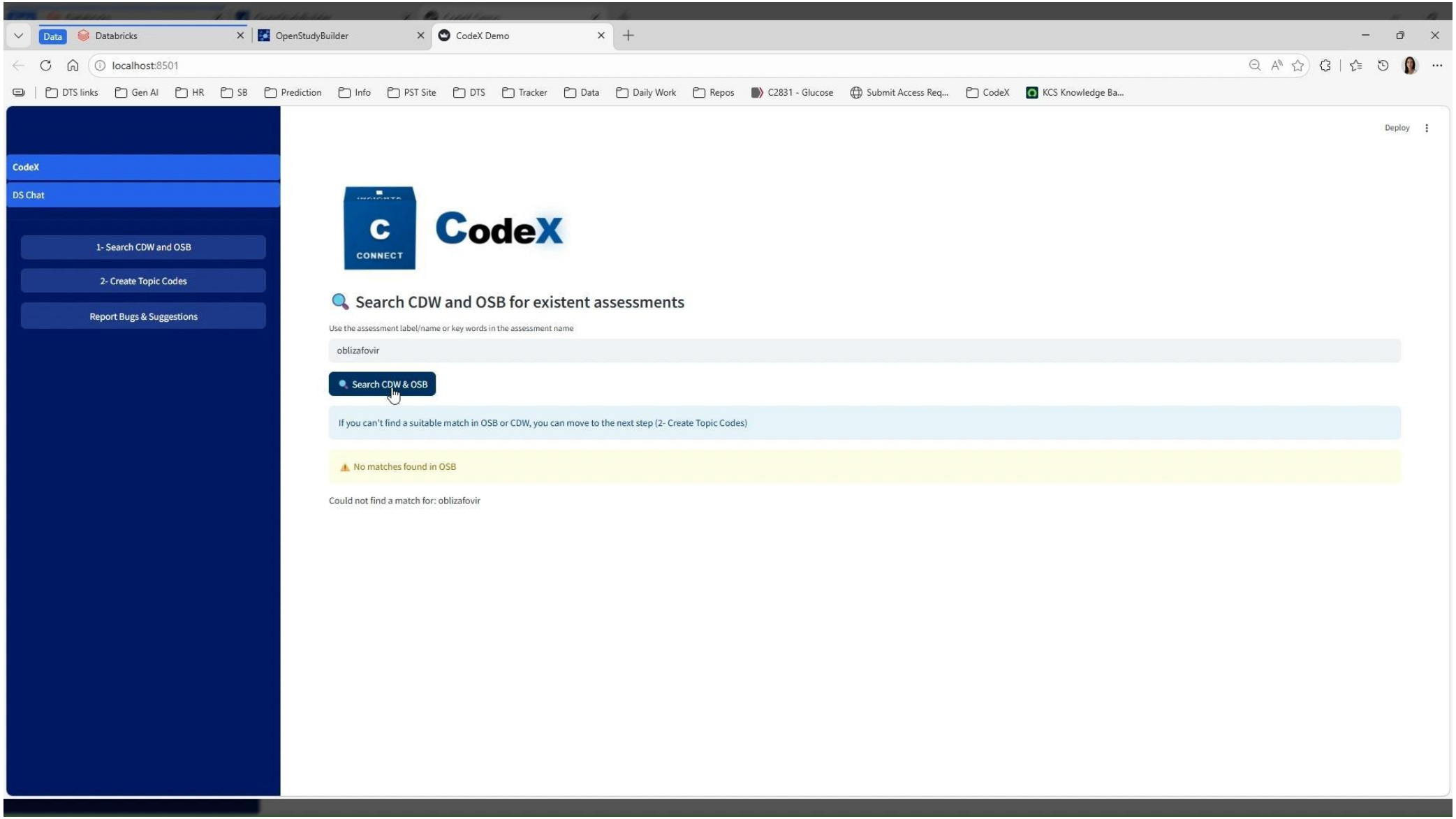


OpenStudyBuilder Activity Concept - Data

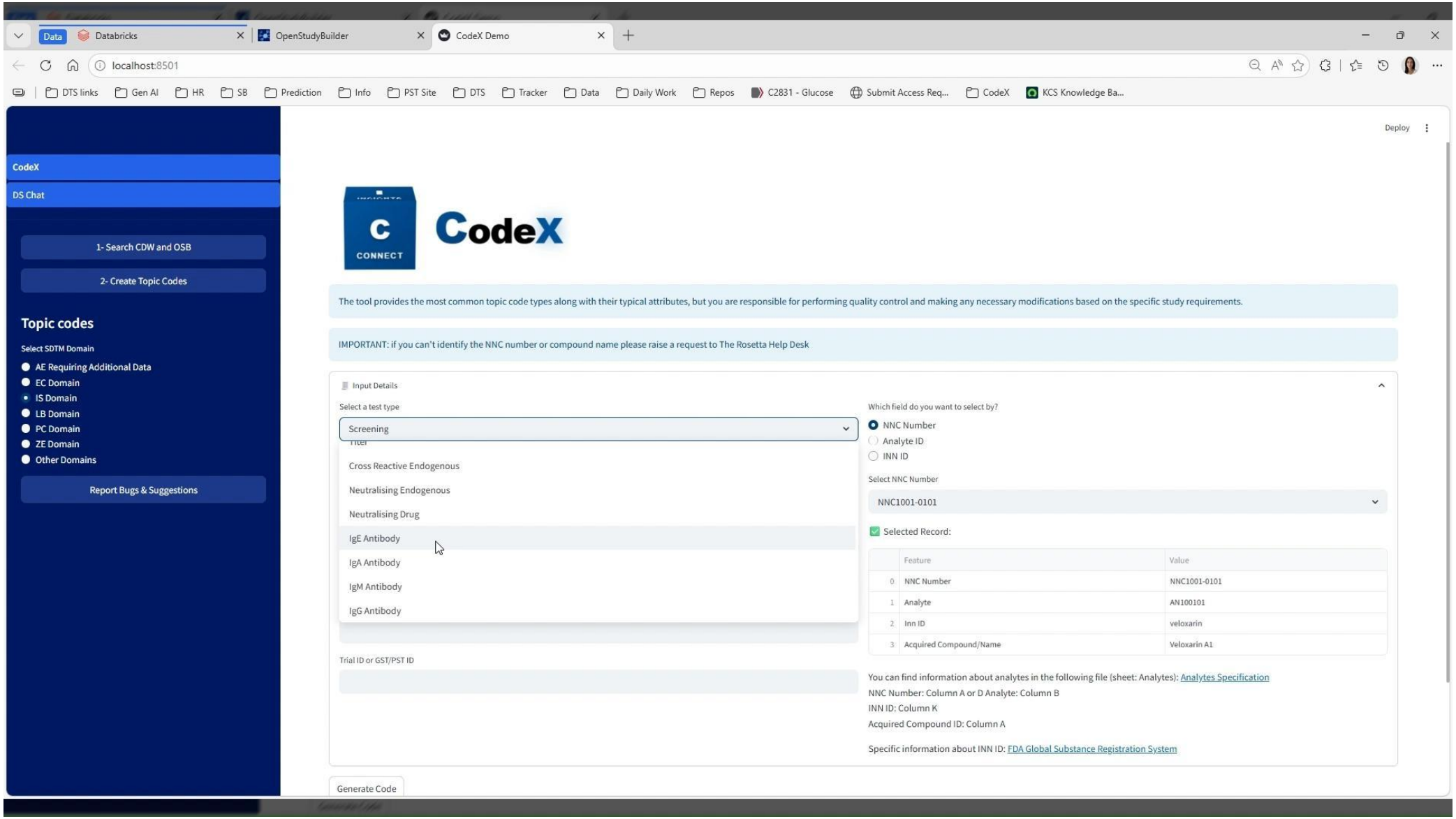


Demo CodeX

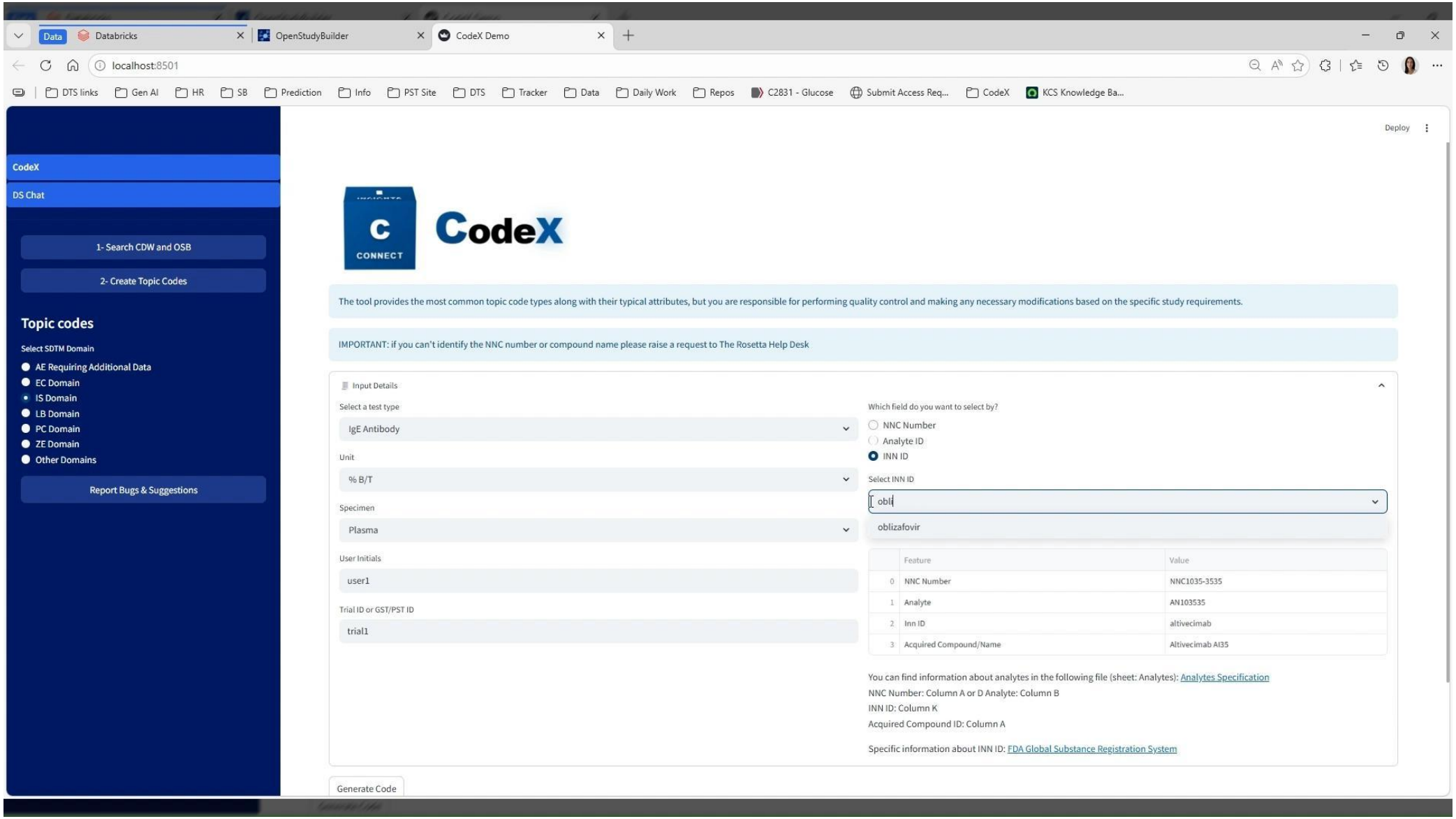
Demonstration Screenshot



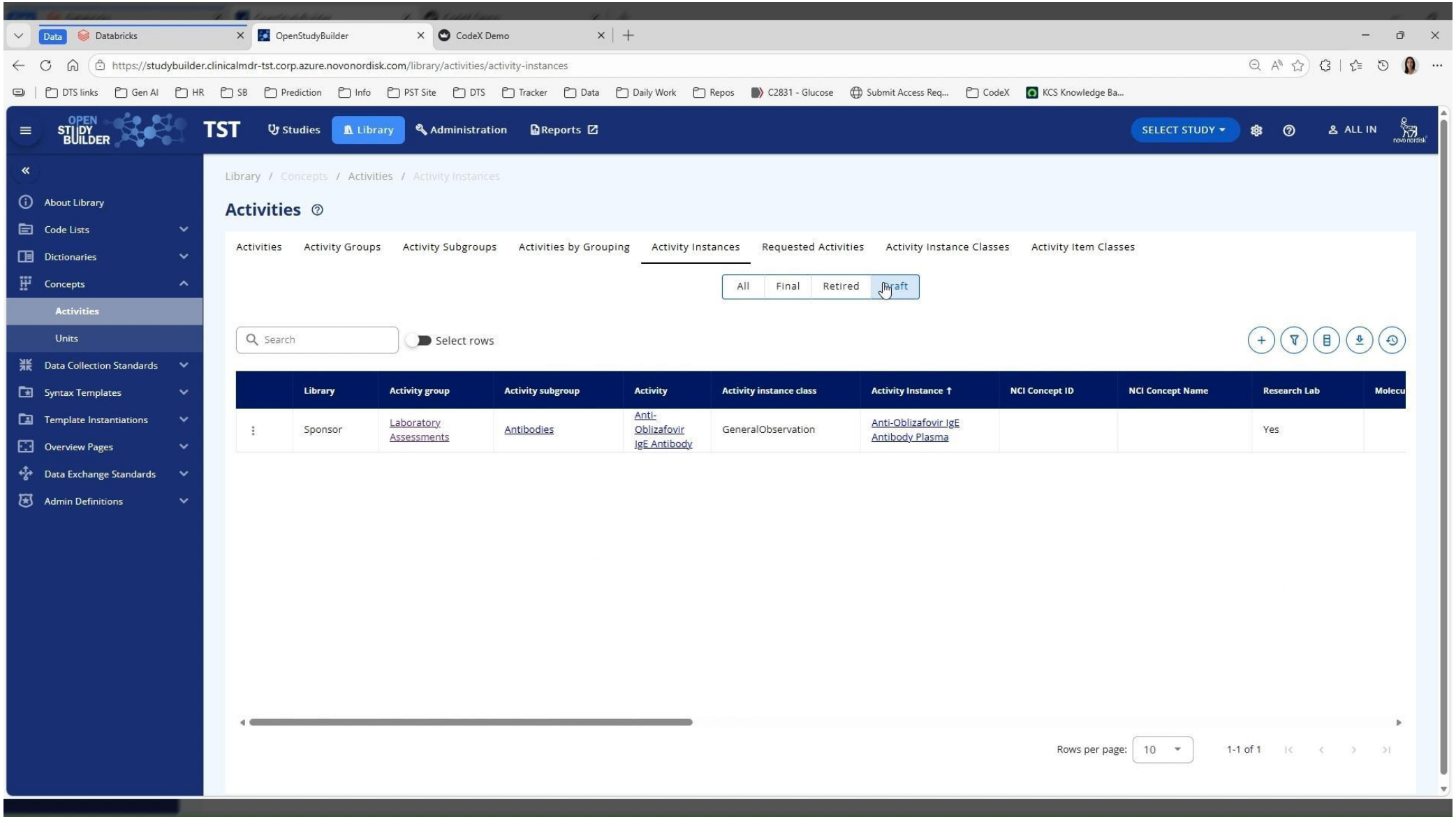
Demonstration Screenshot



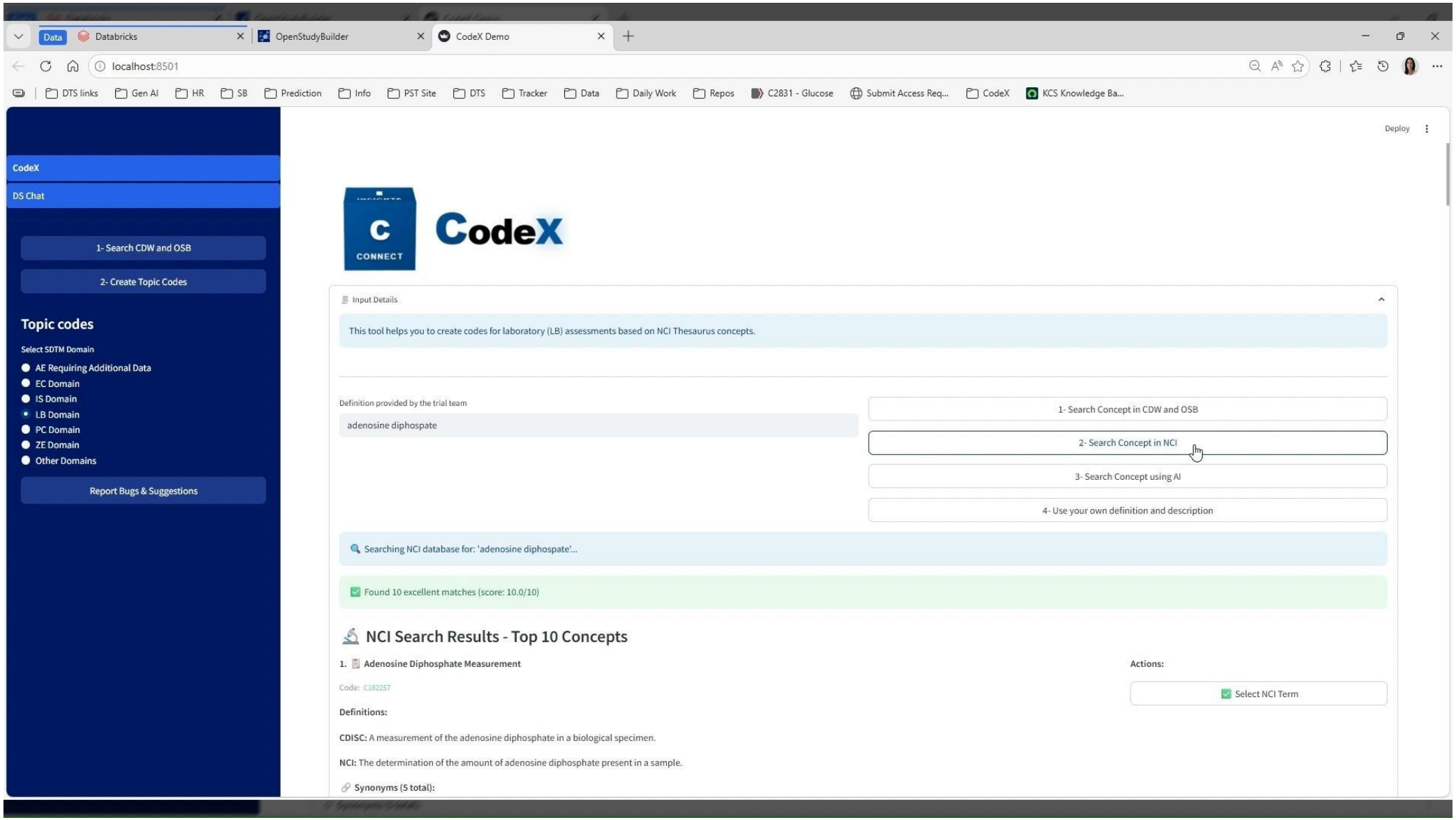
Demonstration Screenshot



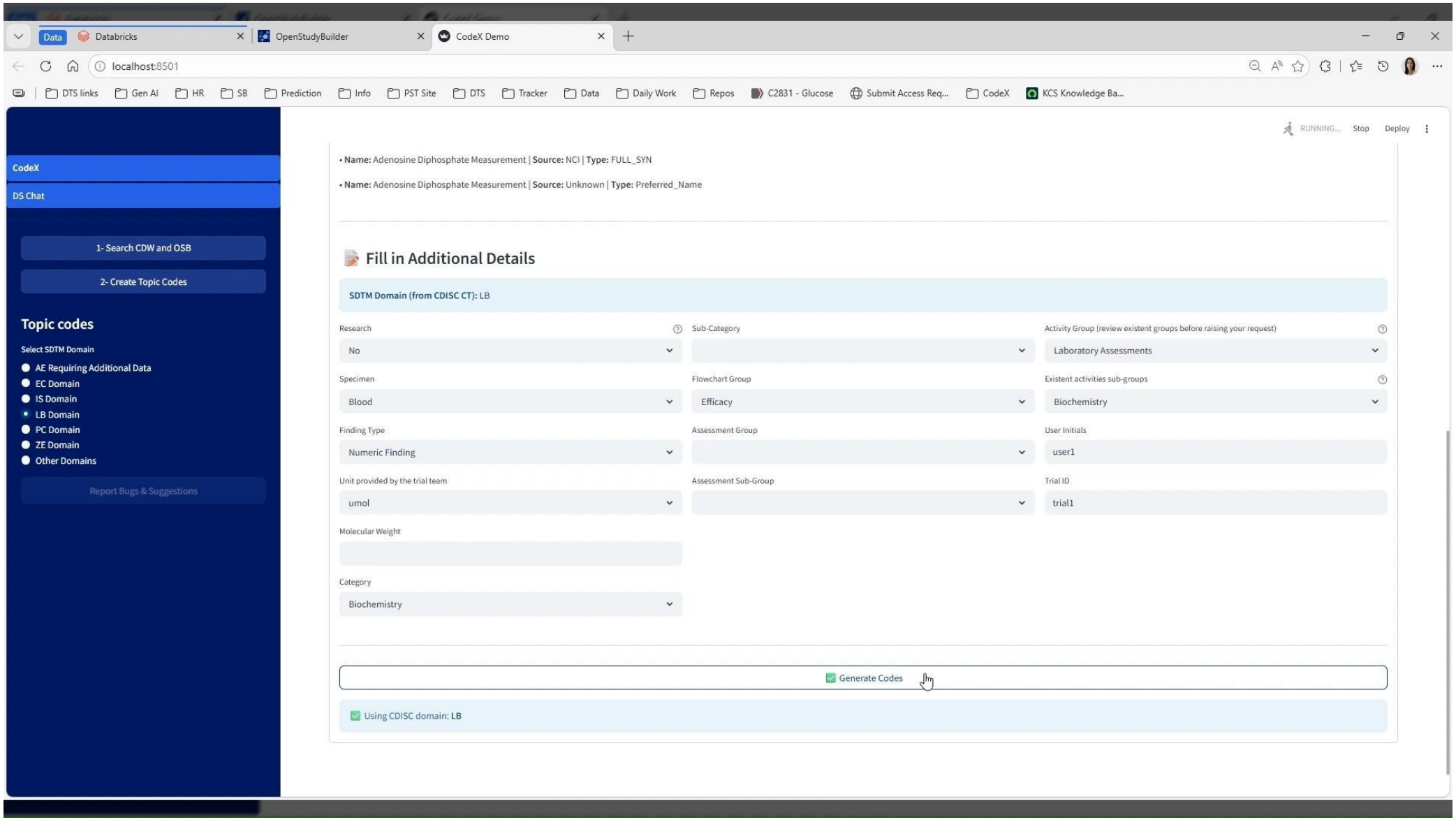
Demonstration Screenshot



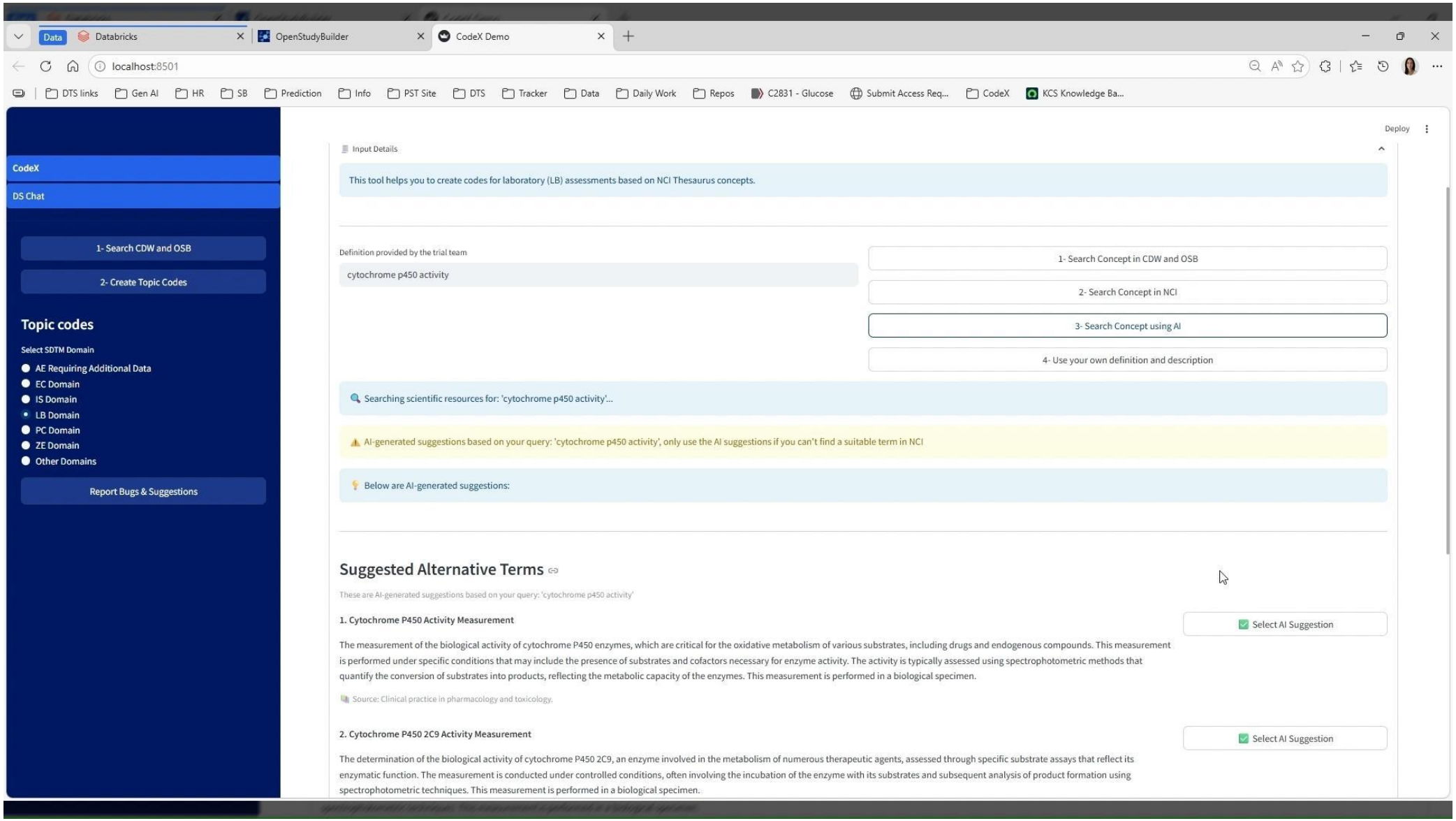
Demonstration Screenshot



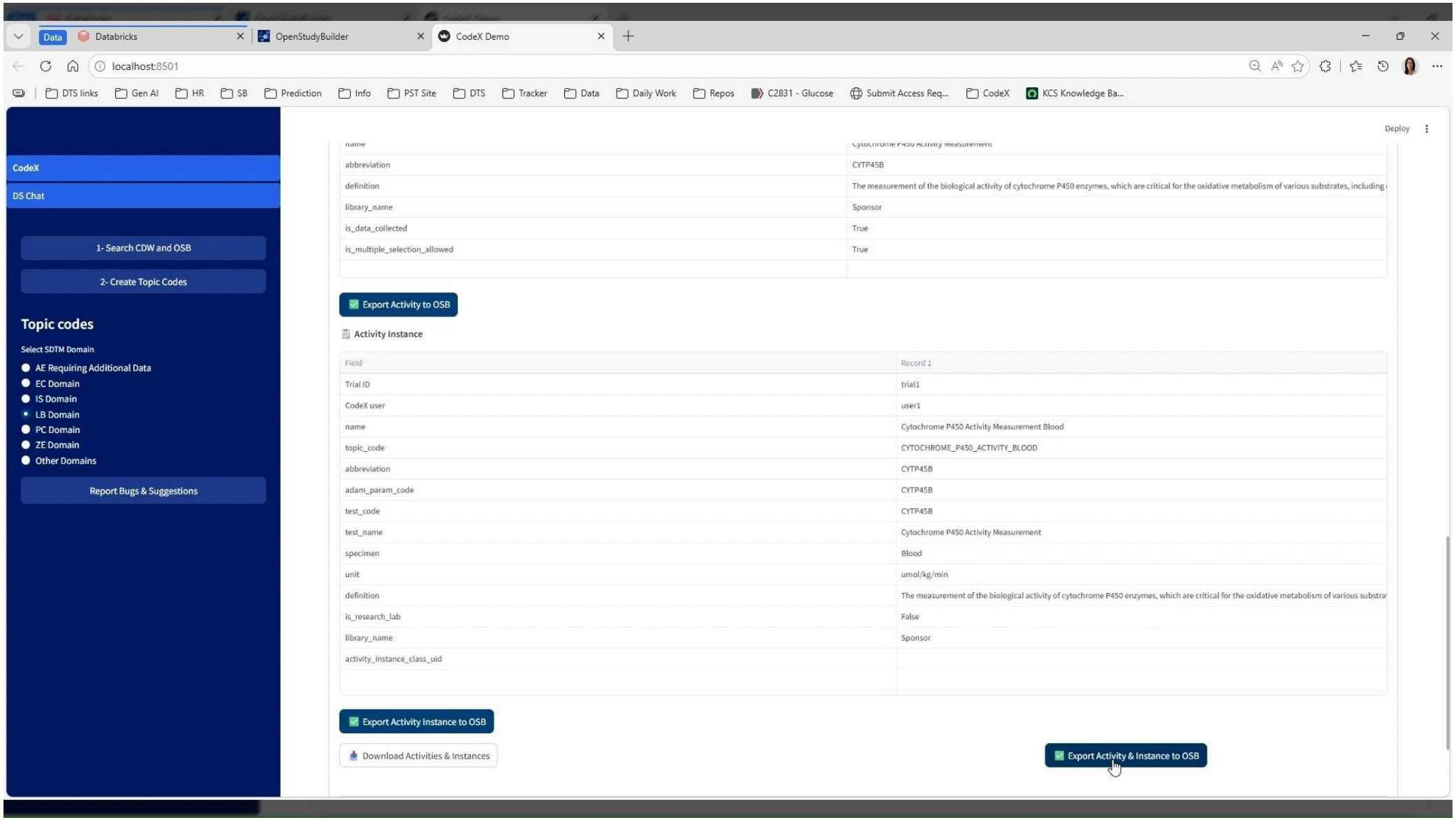
Demonstration Screenshot



Demonstration Screenshot

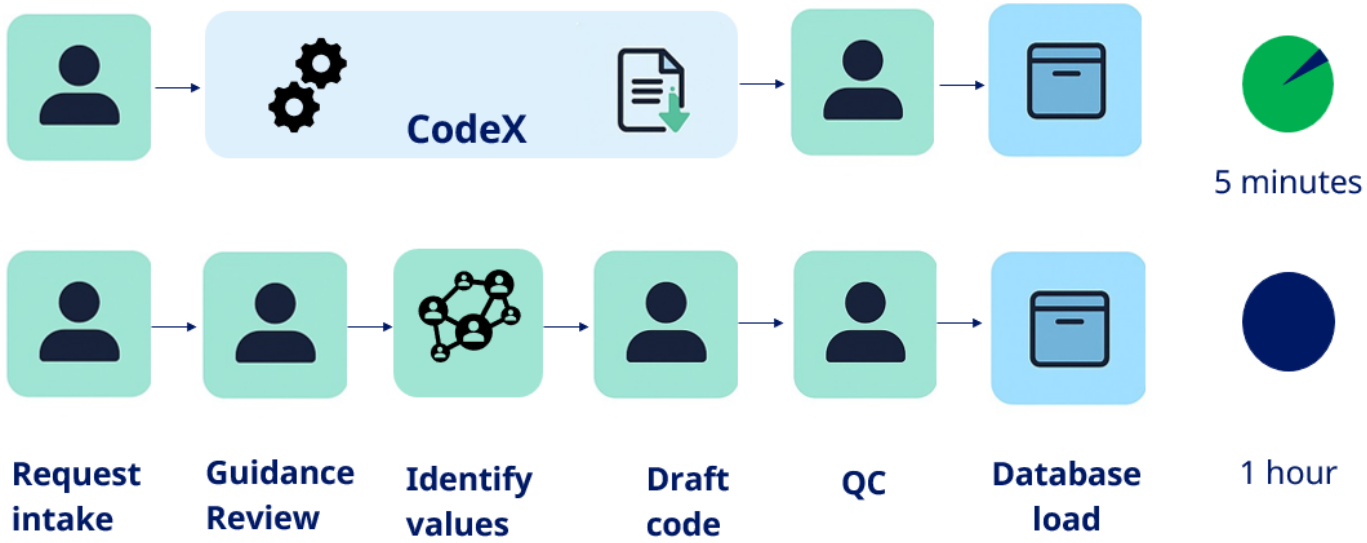


Demonstration Screenshot



CodeX

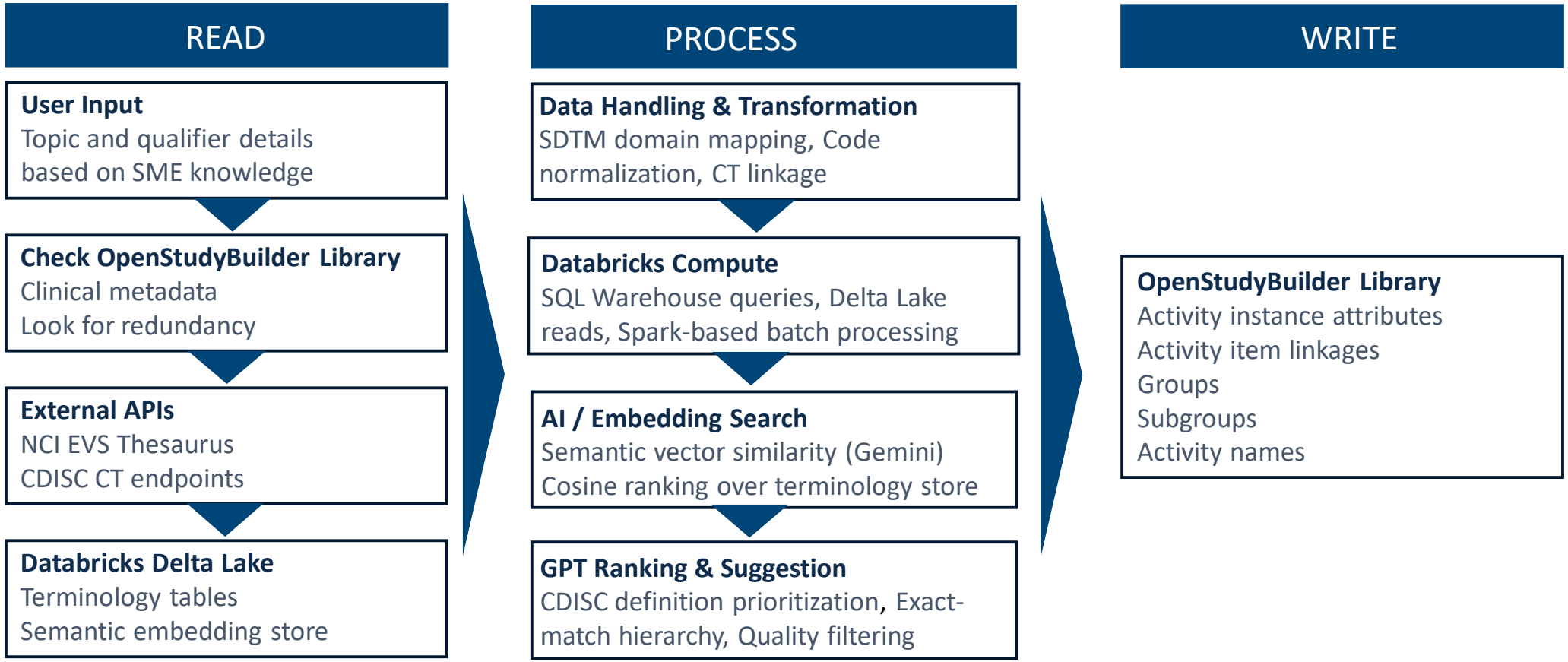
- Focuses attention on the relevant details
- Reduces reliance on manual input
- Leverages predefined rules and AI
- Aligned with industry standards
- Automates the mechanical, enforces consistency and connects workflows



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

cdisc®

Acknowledgements

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

Questions?