PharmaSUG 2025 - Paper MM-23

# Enhancing Dictionary Management and Automation with NCI EVS: A Deeper Dive for the CDISC Community

Anthony Chow, CDISC

## ABSTRACT

For many in the CDISC community, the US National Cancer Institute's Enterprise Vocabulary Services (NCI EVS) may primarily serve as a source for downloading CDISC Controlled Terminology (CT) files or setting data submission values. Yet, this view underestimates the critical, long-standing partnership between CDISC and NCI EVS in developing semantic frameworks that underpin CDISC's standards. Through advanced semantic management tools, NCI EVS maintains CDISC's codelists and controlled terms in a structured and automated environment—far beyond the spreadsheet-based systems. Thanks to a comprehensive API, users can readily access these resources without navigating complex backend processes.

This paper will explore two core NCI EVS tools that can support and enhance CDISC workflows: the EVS Explore, an interactive tool for concept browsing, and the API driving this robust application. We will highlight practical use cases that illustrate their value to standards professionals. For instance, in curating CDISC Biomedical Concepts (BCs), NCI EVS allows users to group concepts into meaningful taxonomies, facilitating systematic concept curation. Another powerful application is for the CT Relationships deliverables. We use these NCI EVS tools to validate content, such as codelists and terms, ensuring consistency and accuracy across CDISC domains in scope. By utilizing these freely available resources, standards implementers can significantly improve the precision, efficiency, and automation of terminology management, streamlining processes for enhanced compliance and data integrity.

## INTRODUCTION

### OVERVIEW OF CDISC AND NCI EVS COLLABORATION

CDISC and NCI EVS have shared a long-standing partnership dating back to the early 2000s. Originally formed to support CDISC's emerging data standards, this collaboration has evolved into a cornerstone of CDISC's approach to CT management. NCI EVS terminology experts are responsible for the development, harmonization, publication, and maintenance of all CDISC CT across foundational and therapeutic area standards.

Controlled terminology plays a critical role in the adoption and implementation of CDISC standards across clinical research workflows. Every term within CDISC CT is assigned a unique code and a clear definition, promoting consistent meaning and usage across studies. This level of semantic precision is especially important in regulatory contexts. Since 2017, agencies such as the FDA and Japan's PMDA have mandated the use of CDISC formats with controlled terms, while China's NMPA and Europe's EMA have encouraged their use. The collaboration with NCI EVS ensures CDISC can meet these regulatory expectations and foster semantic interoperability across the research community.

### COMMON MISCONCEPTIONS ABOUT NCI EVS

A common misconception is that the NCI EVS is merely a static repository for downloading terminology files. In reality, NCI EVS is a robust, enterprise-grade system that leverages advanced semantic technologies to manage and deliver terminology in a dynamic, structured, and automated environment.

Contrary to the belief that users are simply retrieving spreadsheets or flat files, these formats are actually convenient renderings generated from a semantic data store. Tools like Protégé, online browsers, and APIs enable users to interact with the data in far more sophisticated ways than file downloads. For example, CDISC codelists and terms are not managed manually in spreadsheets, but through a comprehensive infrastructure that ensures consistency, automation, and traceability.

Over the past few years, EVS has significantly modernized its technology stack to improve usability and performance. A key advancement was the 2021 launch of EVS Explore — a next-generation, cloud-based terminology browser built on a native triple-store database with Elasticsearch indexing. This new architecture powers fast, intelligent search capabilities, allowing users to retrieve detailed concept information (definitions, synonyms, code list memberships, etc.) using queries that leverage the architecture. The user interface further enhances the experience with features like hierarchical term views and a query speed indicator, which early users have praised as a major improvement over the legacy browser.

### PURPOSE OF THIS PAPER

The primary aim of this paper is to explore the capabilities of the EVS Explore interface and its accompanying API, with a focus on demonstrating practical use cases for process automation and terminology management.

A suite of specialized tools and platforms support the curation and dissemination of CDISC terminology within the NCI EVS ecosystem, which includes these core components:

- **NCI Thesaurus (NCIt):** A comprehensive biomedical ontology that serves as the authoritative repository for CDISC controlled terminology, among other biomedical dictionaries. All CDISC terms and codelists are modeled within NCIt using rich semantic relationships and uniquely assigned concept code (C-code).
- **Ontology Curation with Protégé:** NCI EVS maintains NCIt content using a dedicated instance of Protégé, a powerful ontology editing platform. NCI EVS has customized this environment with specialized plugins and business rules to streamline terminology curation, quality control, and version management within a formal ontology framework.
- **EVS REST API:** Recently redesigned to leverage a triple-store backend, this API allows for advanced, semantically rich queries across the ontology. Applications and developers can retrieve CDISC terminologies, subsets, and metadata on demand — a critical capability for enabling automated workflows.

By showcasing these components, this paper illustrates how organizations can automate key processes such as validating dataset terms or populating electronic case report forms with standardized terminology, driving greater efficiency and consistency in clinical research data management.

## UNDERSTANDING NCI EVS TOOLS

### EVS EXPLORE BROWSER

EVS Explore is a web-based, interactive browser designed to help users navigate, search, and understand biomedical terminologies, including the NCIt and other controlled vocabularies. Freely accessible via any modern web browser at https://evsexplore.semantics.cancer.gov, this tool offers a user-friendly interface for exploring complex semantic structures.

EVS Explore provides stable, unique codes for biomedical concepts, along with rich metadata such as preferred terms, synonyms, research codes, and mappings to external source vocabularies. Users can tailor their experience by limiting searches to specific terminology sources, for example, filtering results to return only CDISC-related concepts.

One of the most powerful and visually intuitive features of EVS Explore is its hierarchical interface, which presents terminology in a tree-like structure. This view highlights the full path from the root of the hierarchy to the selected concept, while also showing sibling and ancestor terms for greater context. As users browse, the interface dynamically scrolls to keep the active concept centered, making it easy to stay oriented within the broader taxonomy.

By simplifying the navigation of complex code systems, EVS Explore serves as a practical tool for researchers, data managers, and developers working with biomedical terminologies.

### EVS REST API

This NCI EVS API is the backbone of EVS Explore, offering a powerful set of tools that enable seamless integration of EVS terminologies into external systems and custom applications. Designed primarily for software developers, the API allows programmatic access to the rich semantic content maintained by the NCI EVS.

Currently, the EVS API provides read-only endpoints, which serve as specific URLs through which applications can query and retrieve data. These endpoints support a variety of key functions, including concept searching, metadata access, and content retrieval across terminologies like the NCIt and CDISC terminologies.

By using the EVS API, developers can automate tasks that rely on standardized terminology, like checking dataset terms and filling in digital forms with the right codes. This helps cut down on manual work and keeps the data more consistent.

For technical specifications and interactive documentation, the API can be explored at: https://api-evsrest.nci.nih.gov/swagger-ui/index.html. And for advanced usage such as integrations with HL7 FHIR and Regenstrief Institute's LOINC, the EVS GitHub repository offers additional resources: https://github.com/NCIEVS/evsrestapi-client-SDK.

Whether building lightweight tools or enterprise-level applications, the NCI EVS API provides a robust foundation for integrating standardized biomedical terminology into workflows.

## USE CASES

### CDISC BIOMEDICAL CONCEPTS (BC) CURATION

BCs are structured representations of clinical ideas such as lab measurements or procedures, combining a concept with related data elements such as result, unit, and method. In CDISC workflows for standards authoring, BCs help standardize how data and metadata are defined and reused across studies, ensuring clarity, consistency, and regulatory readiness. Because BCs are inherently interconnected, they require a taxonomy-driven approach. Organizing them hierarchically with clear definitions and codes is essential to managing complexity and supporting consistent use across domains.

To support this, CDISC uses the NCIt as the authoritative source for BC definitions. NCIt's rich content includes definitions, hierarchical relationships, semantic types, and standardized codes, enabling curators to distinguish between similar terms and select the correct one based on context.

The concept of glucose provides a helpful example of how this process works in practice. When curators search for a term in NCIt, they may find several similarly named entries. For instance, "glucose" is classified as an organic chemical, whereas "glucose measurement" is categorized as a laboratory procedure. This distinction is critical: one term refers to the substance, and the other to the act of measuring it. In the context of findings such as tests and procedures, curators typically select the measurement term. Metadata such as semantic type and hierarchical placement in EVS Explorer play a vital role in guiding curators to the correct selection. Without these tools, the appropriate choice may not be immediately obvious. This example illustrates why NCIt's taxonomical structure is essential for accurate and meaningful curation.

The integration of NCIt and EVS tools into CDISC's BC curation process brings major benefits: enhanced clarity, reduced errors, and improved collaboration among curators and implementers. Anchoring concepts to a centralized ontology ensures long-term maintainability and interoperability of standards. This approach provides both structure and flexibility, forming a strong foundation for evolving CDISC standards and improving the quality of clinical data across the research ecosystem.

### CDISC CT RELATIONSHIPS

The CDISC CT Relationships publication explains which controlled terms or subsets are needed or expected for certain SDTM variables. For example, the AEDECOD variable should use the Preferred Term from the MedDRA dictionary that's maintained by MSSO. Similarly, the AESER variable in the Adverse Events (AE) domain should reference the CDISC CT code list "NY", which indicates whether an

adverse event is serious. In some cases, only specific values are permitted. For instance, the AESER variable only accepts the terms "N" (No) or "Y" (Yes) out of the four possible entries in the "NY" code list.

The latest deliverable, CT Relationships Version 1.0, supports SDTM Version 1.7, SDTMIG Version 3.3, and SDTMIG-MD Version 1.1. It includes over 2,100 data element entries and their controlled terminology stipulations. A significant portion of the content covers CDISC CT metadata such as the code list name, the associated NCIt C-code, and the corresponding submission values. These elements must be aligned for accuracy. Given the volume, manual validation is not feasible. To ensure correctness and efficiency, the team implemented an automated quality control process using the NCI EVS API.

CT Relationships provides significant value to users by offering a high-quality product that clearly identifies and documents all relationships among CDISC variables, non-standard variables, TEST/PARMs, and their associated terminology codelists or subsets. All content is carefully reviewed and verified for consistency and accuracy. This was made possible using the NCI EVS API.

## PRACTICAL IMPLEMENTATION

To implement controlled terminology effectively within clinical data standards, organizations should adopt a strategic and practical approach that leverages resources from NCI EVS and CDISC. The first step is to ensure that key personnel become familiar with the EVS Explore interface. This tool provides hierarchical views and definitions of CDISC CT, helping users identify discrepancies between internal data dictionaries and standard terminology. This process promotes alignment with authoritative sources.

Organizations should then map their workflows to identify where terminology is used and evaluate how EVS tools can automate or enhance these points. Key stages to consider include study setup, case report form design, and data validation. By leveraging the EVS REST API, teams can access CT in a scalable and on-demand manner. Integrating this functionality into enterprise systems can eliminate reliance on static files. When designing these systems, teams should incorporate version control to enhance change management and data caching for efficient access.

Integration efforts also extend to governance frameworks. Organizations should update policies to require the use of NCI EVS and the CDISC Library as primary terminology sources. Appointing a terminology curator and monitoring CT updates helps maintain compliance and readiness for change. The CDISC Library also supports governance by validating the correct use of codelists in context. Encouraging participation in public reviews further supports proactive terminology management.

A strong terminology strategy requires a cultural shift. Organizations should promote regular reviews, foster exploration, and train staff to use tools such as EVS Explore and the CDISC Library effectively. Transitioning from spreadsheets to API-driven workflows may require dedicated training and, in some cases, the development of user-friendly internal tools powered by these APIs.

Alignment with regulatory and data standards remains essential. Regulatory agencies, including the FDA and PMDA, mandate the use of CDISC CT. Internal semantic models must reflect these standards to support submission readiness. The CDISC Library ensures that variables are linked to appropriate codelists, while NCI EVS offers mappings to external vocabularies such as MedDRA, SNOMED, and LOINC. These mappings enable broader data integration. Staying informed about regulatory guidelines and maintaining a reference matrix for CT versions used across domains helps ensure compliance and supports data interoperability.

## FUTURE OPPORTUNITIES

Looking ahead, the development of a *CT Browser* presents exciting new opportunities that expand what was previously possible in managing and using controlled terminology. This web-based application is designed to significantly improve both accessibility and usability of CDISC CT. With interactive browsing and advanced search capabilities, including semantic and natural language search, users can quickly locate the terminology they need. Dictionary maintainers will gain new insights through graphical visualizations that make it easy to explore the lineage of packages, codelists, and terms.

By integrating essential data directly into *CT Browser*, including codetable mapping files and coded terms with their corresponding decodes (e.g., TESTCD and TEST or PARMCD and PARM pairings), users will

significantly reduce the need to reference external resources. The browser will also offer clear insight into how codelists support BCs, helping users understand where specific codelists function as value domains.

Customizable subsetting features will allow users to filter content based on specific needs, such as QRS codelists, domain-specific content, or therapeutic areas. An export function will enable users to save selected content to a personalized shopping cart for easy download. During public review periods, draft codelists and terms can be explored visually, making the feedback process more informed and efficient. The browser will also support change management by allowing users to compare different versions of the controlled terminology, view the type and impact of changes, and generate reports that support process automation. Altogether, these capabilities offer a smarter, more agile way to work with controlled terminology that was not feasible before.

## CONCLUSION

The CDISC community continues to benefit from its long-standing partnership with NCI EVS. This collaboration has become a cornerstone of semantic precision and regulatory compliance in clinical research. Tools such as EVS Explore and the EVS REST API now offer users a powerful, structured, and automated environment for managing controlled terminology. These tools enable organizations to move beyond manual, spreadsheet-based methods and adopt scalable, integrated approaches to terminology curation, validation, and governance.

By incorporating these tools into enterprise systems and governance frameworks, CDISC implementers can position terminology as an active, governed element of the clinical data lifecycle. This integration supports automation, promotes consistency, and enhances regulatory readiness. Proactive terminology management—combined with organizational alignment and targeted training—ensures that controlled terminology is not only maintained but meaningfully embedded into day-to-day workflows.

Looking ahead, broader adoption of these tools and the introduction of innovations such as the upcoming CT Browser will further advance terminology management. Features like interactive navigation, visual lineage tracking, and context-aware subsetting will help users apply and evolve CDISC terminology with greater agility and insight.

By leveraging openly available resources, organizations can build on a solid foundation and tailor solutions to meet specific needs. In doing so, they ensure that terminology remains a strategic asset that supports both innovation and compliance.

## REFERENCE


Muhlbradt, E. E., Li, J. V., Whiteman, L. & Remennik, L., (2023) “NCI-EVS: Building the Semantic Infrastructure to Support CDISC Data Standards and Real-World Data”, Journal of the Society for Clinical Data Management 3(1). doi: https://doi.org/10.47912/jscdm.134

“CDISC Library Data Standards Browser.” Accessed March 28, 2025. Available at https://library.cdisc.org/browser.

“CDISC Library API Developer Portal.” Accessed March 28, 2025. Available at https://api.developer.library.cdisc.org.

CDISC BC Machine-Readable Resources in GitHub. “CDISC Conceptual and Operational Standards Metadata Services (COSMoS).” Accessed March 28, 2025. Available at https://github.com/cdisc-org/COSMoS.

CDISC CT Relationships. “Controlled Terminology Relationships v1.0 for SDTM v1.7, SDTMIG v3.3, SDTMIG-MD v1.1”. Accessed March 28, 2025. Available at https://www.cdisc.org/standards/terminology/controlled-terminology#standard__Resources.

“NCI EVS Explore.” Accessed March 28, 2025. Available at https://evsexplore.semantics.cancer.gov.

NCI EVS REST API Endpoints Documentation. “NCI EVS Rest API Version 2.1.0”. Accessed March 28, 2025. Available at https://api-evsrest.nci.nih.gov/swagger-ui/index.html.

NCI EVS Software Development Kit in GitHub. “EVSRESTAPI Client SDK.” Accessed March 28, 2025. Available at https://github.com/NCIEVS/evsrestapi-client-SDK.

## ACKNOWLEDGEMENTS

I gratefully acknowledge the generous review and insightful feedback provided by Linda Lander (CDISC), Jordan Li, PhD (NCI EVS), and Erin Muhlbradt, PhD (NCI EVS). This paper would not have been possible without their valuable contributions.

## CONTACT INFORMATION

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

Anthony Chow
CDISC
achow@cdisc.org