

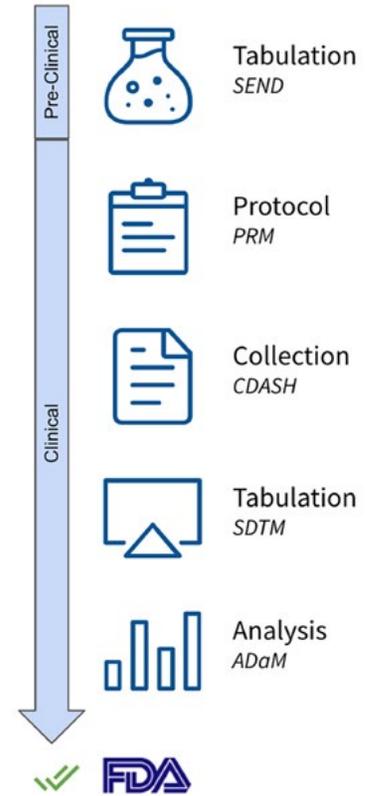
Tamr CDISC Conversion Solution

Unify messy clinical data into standard CDISC models

From designing better trials to spotting successful (or failed) outcomes earlier in the drug development process, having unified views of your clinical research data comprises a critical part of the modern R&D landscape for life sciences companies. The Clinical Data Interchange Standards Consortium (CDISC) is a widely used authority in managing these views and include standards such as the Standard for Exchange of Nonclinical Data (SEND), the Protocol Representation Model (PRM), the Study Data Tabulation Model (SDTM), and the Analysis Data Model (ADaM).

However, companies often fail to maintain unified views because data is stored in legacy formats, was acquired via M&A activity, or originated from disparate external sources with their own specific data standards. These challenges can easily result in hundreds (or thousands) of studies being inaccessible to their scientists—a situation referred to as having large data debt.

Tamr's CDISC Conversion Solution tackles data challenges with a powerful data harmonization process, driven by human-guided machine learning that can replace traditional in-house tools prone to delays and errors.



Key Benefits:

With Tamr, life science companies can make transformational changes to their research and drug development processes.



Integrate historic and new studies into a unified repository of clinical data



Leverage built-in systems based on industry data best practices



Drastically increase data processing speeds up to 10+ billion records per day



Automate and streamline data conversion from multiple sources



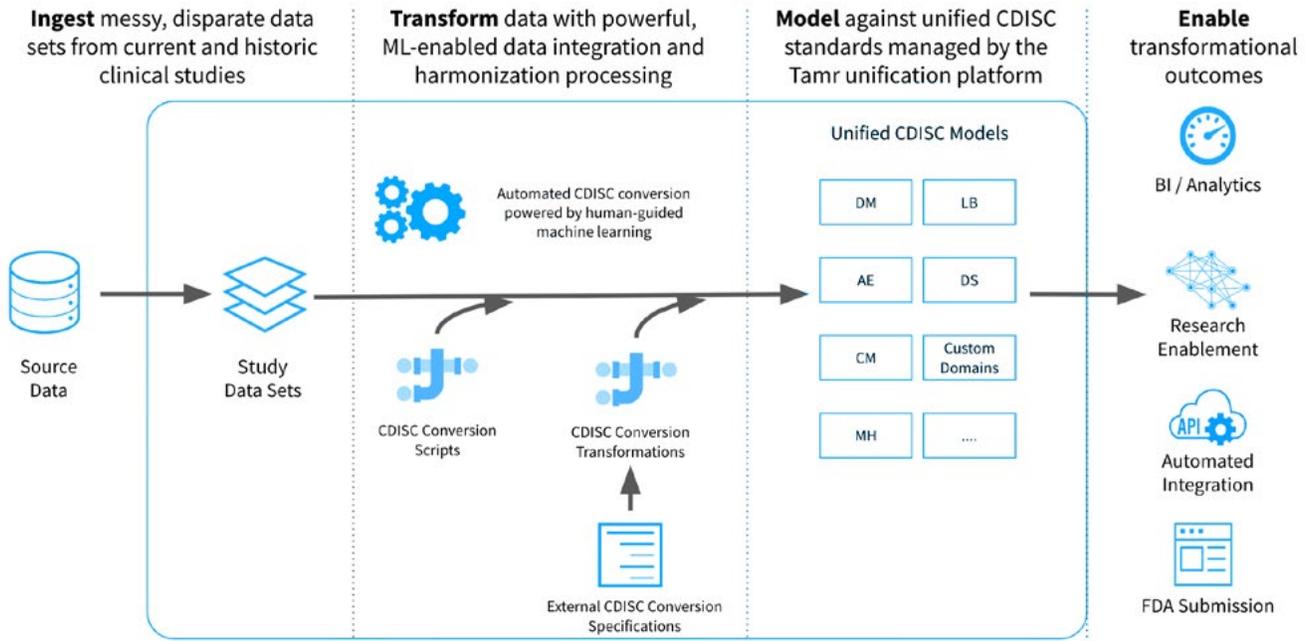
Consolidate existing conversion specs to accelerate data conversion timeline



Enable data scientists and researchers with previously inaccessible data

Solution Overview

The Tamr CDISC Conversion Solution provides an end-to-end approach to transform how life science companies manage their research and clinical data. Through the combination of human expertise and machine learning, Tamr provides a robust platform to sustain ongoing CDISC conversion of study data through data ingestion, transformation, modeling, and enablement.



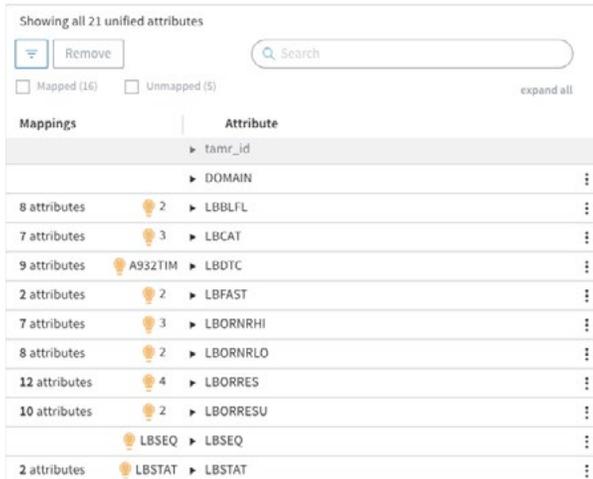
The Tamr CDISC Conversion Solution is made up of the following key features:

- **Unified CDISC model management**
All varieties of data models are managed within the Tamr platform throughout the data conversion process. With the help of built-in starting templates for CDISC data models, companies can accelerate their data conversion projects.
- **Data model discovery and maintenance**
Leveraging human-guided machine learning, Tamr's solution provides automated mapping suggestions for on-boarding new datasets into the platform.
- **Powerful data transformation engine for CDISC**
Transformation capabilities for CDISC conversion is accessible through intuitive SQL-like syntax, and processed using a high performance Spark engine that easily scales to fit the project.
- **Conversion of custom mapping specifications**
Projects with existing data harmonization specifications can upload existing mapping specs into the CDISC solution to accelerate the integration of all study data into unified data models.
- **Data lineage and audit trail**
All data conversion transformations and user interactions are tracked and logged by Tamr to provide clear data lineage and audit trails.

Key Features

Unified CDISC Model Management

Create, update, extend, import, and export domain data models



Working with the world’s largest life science companies, Tamr understands industry challenges of integrating clinical, biomarker, and lab data from disparate sources into unified data models through applying CDISC standards to harmonize study data.

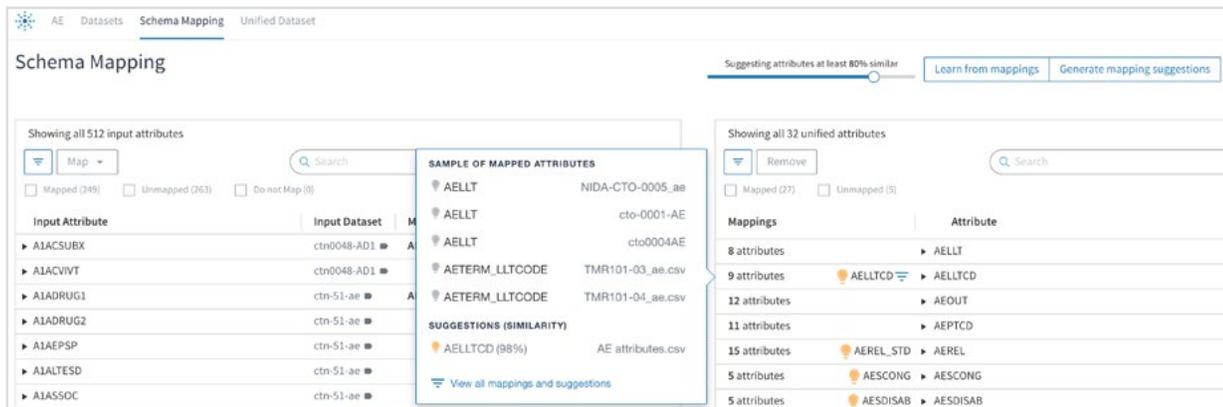
The Tamr solution provides a robust platform to manage predefined templates of target unified CDISC data models derived from the latest CDISC standards. Therefore, companies are provided with a framework to create and modify unified data models to quickly build out and maintain data modeling processes without impacting upstream clinical data sources.

Tamr provides domain-specific CDISC models that can be modified to fit specific studies

Data Model Discovery and Maintenance

Data Modeling Engine Powered by Machine Learning

The Tamr CDISC Conversion Solution leverages human-guided machine learning to capture and apply domain knowledge from repeated CDISC conversion projects. By identifying similar clinical data sets, both historically neglected and new clinical data sets can adopt previous data modeling transformations to automatically map into CDISC data models with negligible additional effort. Moreover, the machine learning engine will automatically capture small schema discrepancies that may occur along the CDISC data conversion pipeline for ongoing data source inputs.



Automatic schema recommendations through Tamr’s human-guided machine learning accelerates the on-boarding of new domain data sets

Powerful Data Transformation Engine

Rapid Data Transformation Processing

Many life science companies are limited by their ability to efficiently move data through their DataOps infrastructure. Most ETL solutions are lacking in the capability to regularly transform and process disparate data sources into unified views. Further hampering the effort is that traditional ETL solutions may take several days to weeks when processing a large scale of disparate clinical data records.

The Tamr solution, however, provides a wide collection of CDISC transformation templates and capabilities that can be reused across various data sources to accelerate transformation processing. In addition, transformations are processed through Tamr's innovative implementation of Spark that can be scaled to efficiently process clinical data records at over 10+ billion records (or multiple terabytes) a day.



The screenshot displays two side-by-side panels. The left panel, titled 'Formula', shows a transformation rule for the column 'AESTDTC'. The formula is: `datetime_to_iso(get(AESTDTC, 0), array('dd MMM yyyy', 'dd.MM.yyyy', 'dd.M.yyyy', 'd.MM.yyyy', 'd.M.yyyy'))`. Below the formula, it states 'output column: AESTDTC for records from unified dataset'. The right panel, titled 'EDIT CODE', shows the equivalent SQL query: `SELECT *, VISIT, VSDTC; WINDOW row_number() AS VSSEQ BY USUBJID ORDER BY VISITNUM, VSDTC, VSTPTREF, VSTPTNUM, VSTESTCD;`. A 'Collapse' button is visible in the top right corner of the SQL editor.

Simple and complex transformations can be executed with easy, SQL-like syntax

Conversion of Custom Mapping Specifications

Repeatable, Automated Schema Mapping at Scale

One of the largest inefficiencies in clinical data management is the low repeatability of data wrangling processes applied to different datasets. While an organization may have existing processes to wrangle and transform data sources into CDISC standards, these tools leverage handmade data dictionaries and transformation logic that are tailored each time to study-specific data sources and rules.

To counteract this fragmenting of work, Tamr's data transformation engine provides the ability to automatically translate current mapping specifications to be used as part of Tamr's CDISC Conversion Solution. Since organizations do not have to replace existing processes, adoption of the Tamr solution can happen quickly - with some users continuing to leverage tools that they are most familiar with.

Leveraging machine learning and natural language processing, all CDISC conversion specifications can be processed through the solution to build robust transformations in the platform that can be reused on all similar data sources.

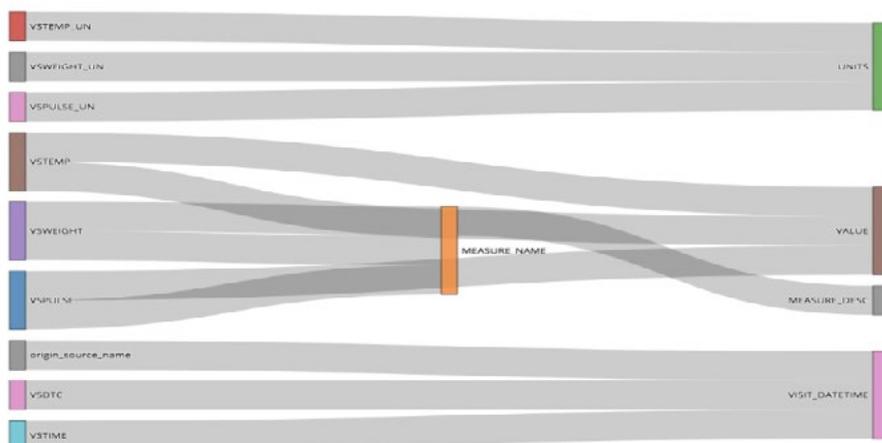
Below is an example of legacy conversion specifications consumed by Tamr to automatically generate SQL-like transformations and map clinical study source data into CDISC standard data models.

Source	Source Table	Source Column	Target Table	Target Column	Mapping Logic (psuedo code)	Transpose Rule
Source 1	DOUBLE_SUBJECT	SUBJECT_NUMBER	RN	INVID	Lookup to DM on SUBJECT_NUMBER = DM.SUBJID, return DM.INVID	RN1
Source 1	DOUBLE_SUBJECT	SUBJECT_NUMBER	RN	SPID	'RV'-<SUBJECT_NUMBER>	RN1
Source 1	DOUBLE_SUBJECT	SUBJECT_NUMBER	RN	VISIT	'UNSCHEDULED'	RN1
Source 1	DOUBLE_SUBJECT	SUBJECT_NUMBER	RN	VISITNUM	IF UPPER(FOLDERNAME) LIKE '%UNSCHE%' THEN 900 ELSE LOOKUP{VISIT= FOLDERNAME; VISITNUM}	RN1
Source 1	DOUBLE_SUBJECT	TREATMENT_LONG_NAME	RN	RANDCD	Study specific	RN1
Source 2	S_AE_LT_RAW	AE_DESCRIPTION_M2	AE	AEDICTV	Lookup to TMS Coding-MEDDRA, Return <Dictionary Version>	AE2
Source 2	S_AE_LT_RAW	AE_DESCRIPTION_M2	AE	AEDICTV_QEVAL	CLINICAL STUDY SPONSOR'	AE2
Source 2	S_AE_LT_RAW	AE_DESCRIPTION_M2	AE	AEDICTV_QORIG	'ASSIGNED'	AE2
Source 2	S_AE_LT_RAW	AE_DESCRIPTION_M2	AE	AEHLGT	Lookup to TMS Coding- MEDDRA, Pass AE_DESCRIPTION_Mx, Return AEHLGT	AE2

Life science companies can access Tamr’s powerful CDISC conversion engine while still using legacy conversion specifications for CDISC (e.g. ‘AE’) as well as custom(e.g. ‘RN’) domains of data

Data Lineage and Audit Trail

Documented logging of every data transformation and user activity



All transformations and user activities are recorded throughout the Tamr platform to document how source data attributes are transformed into the CDISC data model, as well as the users making the transformations. The documentation of data lineage provides transparency into the process and dependencies around the CDISC conversion projects for better user collaboration, as well as for auditing purposes.

Data lineage is kept intact with detailed tracking of all record transformations in the platform

Transformational Impact

Organizations achieved results that were previously impossible

The Tamr CDISC Conversion Solution provides a process to replace inefficiencies involved in manually unifying disparate data sources with a platform powered by automation and machine learning. As a result, life science companies go through a transformational shift when they deliver results that were previously impossible.



Amgen leveraged Tamr as a key component of their data integration pipeline: "Tamr made a huge impact in our ability to process over 200 Amgen legacy clinical study data in less than a year with only two resources. When you consider that such schema mapping was conducted on over 4000 source tables, the benefit of machine learning really added up."



GlaxoSmithKline (GSK) implemented Tamr and saw incredible initial results in their goal of unifying data from three different domains (assays, clinical trial data, and genetic data). With Tamr, GSK mapped over 40,000 datasets across 1,000 studies to 36 different SDTM domains. "We are doing a step change on machine learning [...]. We simply have to have more machine learning skills to deal with all the available data now" - Mark Ramsey, R&D Chief Data & Analytics Officer

In addition, the Tamr platform scales big data infrastructure to meet any volume of data processing demand. Overtime, Tamr customers have been able to efficiently tackle additional domains while setting up the capability to **process over 10 billion records a day**.

Tamr's transformational results have reshaped the landscape of what is possible with a company's research data no matter where they are in their data transformation maturity. With Tamr, pharma customers have been able to invigorate a vibrant, data-driven culture among their scientists who can now efficiently harmonize historic and new study data in the pursuit of finding better treatments and outcomes for patients.

About Tamr

Tamr is the enterprise-scale data unification company trusted by industry leaders like GE, Toyota, Thomson Reuters, and GSK. The company's patented software platform uses machine learning supplemented with customers' knowledge to unify and prepare data across myriad silos to deliver previously unavailable business-changing insights. With a co-founding team led by Andy Palmer (founding CEO of Vertica) and Mike Stonebraker (Turing Award winner) and backed by founding investors NEA and GV, Tamr is transforming how companies get value from their data.

To find out more or register for a demo visit tamr.com