

Data Curators and Loaders

Introduction

These pages provide instructions on how to load data into a tranSMART installation. There are several open-source ETL tools developed and maintained by members of the tranSMART community.

Video tutorials on how to load data into tranSMART using two of the most frequently used tools are available from the links. For [transmart-batch](#) tutorial click [here](#). Follow this link for [tMDataLoader Training](#) video.

Additional tutorials and manuals for these and other ETL tools are available from dedicated Tool page (below).

Additional pages in this wiki section provide guidance on how to curate and organize your data prior to loading.

| Content | Description |
|--|---|
| Loading data with transmart-batch (Cell Line Use Case dataset) | This page describes how to load a very complete set of example cell line data from multiple platforms, using the transmart-batch ETL tool. Data are provided by the CTMM TraIT project. |
| Loading data using tMDataLoader | tMDataLoader is a command-line data loading interface developed by Clarivate Analytics (Formerly IP&Sciences, Thomson Reuters) (v2.1, May 2015). tMDataLoader Wiki |
| Loading data with Kettle (Step by step tranSMART ETL Guide) | A guide for clinical and expression data loading from JnJ (May 2015) |
| Loading data with transmart-data (Cell Line Use Case dataset) | This page describes how to load a very complete set of example cell line data from multiple platforms, using the transmart-data ETL tool. Data are provided by the CTMM TraIT project. |
| Loading data using Integrated Curation Environment (ICE) | The ICE is a GUI-based data loading framework developed by Sanofi (v1.4, Jan 2015). |
| Loading VCF files with transmart-data | How to load Variant Call Format (VCF) files into tranSMART using transmart-data. If you have your vcf file/files ready, you can also load them with tMDataLoader . |
| TranSMART Guide for Manual Data Deleting | Sometimes you may want to delete data and you can do it manually. Or you can use <code>--delete-study-by-id <delete_id></code> or <code>--delete-study-by-path <delete_path></code> tMDataLoader options. |
| Data Curation Standards | This page is intended to be a repository of information about strategies used and documents/forms created by tranSMART users to standardize tranSMART data loading. |
| CTMM TraIT OpenClinica-to-tranSMART Converter | For those who use OpenClinica to acquire eCRF data, the CTMM TraIT project has developed tools to help transfer data from OpenClinica to tranSMART. This GitHub repository contains the tools and complete documentation. |
| eTRIKS Standards Starter Pack | the eTRIKS project has developed a comprehensive approach to data standards for use with tranSMART. <i>Data harmonisation begins with standards.</i> |
| Supported Data Types | An overview of the types of data that tranSMART supports, and details of how to load each. |
| Curated Data Repository | This page contains curated data sets that are ready to load into tranSMART. Check back often to see what's been added. |