

# Loading data with Kettle (Step by step tranSMART ETL Guide)

## Contributors

- Denny Verbeeck, J&J. [dverbeec@its.jnj.com](mailto:dverbeec@its.jnj.com)
- Francisco Bonachela Capdevilla, J&J. [fcapdevi@its.jnj.com](mailto:fcapdevi@its.jnj.com)

This document is intended to be a guide for performing Extraction, Transformation and Loading (ETL) processes in tranSMART. It covers the ETL pipeline on Windows and Linux systems, for tranSMART instances running on a Postgres database. Chapter one in the Dataset Explorer ETL Guide [1] gives an excellent discussion on how to plan and build your ontologies, which will determine how data appears in tranSMART. This guide aims to combine the dispersed information on ETL from various sources, such as the tranSMART Foundation Wiki [2] and the ETL Guide mentioned above.

This guide assumes that you have a working instance of tranSMART installed.

[1 Introduction](#)

[2 Dependencies](#)

[3 Set-Up](#)

[4 Kitchen/Spoon Basics](#)

[5 Clinical Data](#)

[6 Gene Expression Data](#)

[7 Gene Expression Platform Definitions](#)

[8 Uploading Data with ICE](#)

[9 Removing Data](#)

[10 Troubleshooting](#)

[Appendix A Sample kettle.properties file](#)

[Appendix B Removing a study](#)

[Original Document \(PDF\)](#)

[Example files \(zip\)](#)

[References](#)

[1] Recombinant Data Corp., Dataset Explorer ETL Guide, 2012, Available at [https://wiki.transmartfoundation.org/download/attachments/131201/tranSMART\\_DSE\\_ETL\\_Guide.pdf](https://wiki.transmartfoundation.org/download/attachments/131201/tranSMART_DSE_ETL_Guide.pdf)

[2] tranSMART Foundation Wiki, ETL Section, Available at [Data Curators and Loaders](#)

[3] NCBI-GEO, Available at <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>.

## Version History

Version 1.0 May 13, 2015 Initial Version

Copyright ©2015 eTRIKS. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is available at <http://www.gnu.org/licenses/fdl.html>