

## UPLOADING CURATED PPMI DATA TO A LOCAL TRANSMART INSTANCE

### BACKGROUND

Thank you for your interest in curated Parkinson's Progression Markers Initiative (PPMI) data! The Michael J. Fox Foundation for Parkinson's Research (MJFF) is extremely excited to make curated PPMI data available for download and in tranSMART, a knowledge management platform, through the MJFF data access center. PPMI is an ongoing, longitudinal observational study of idiopathic Parkinson's disease (PD) patients, age- and gender-matched controls, and affected and unaffected individuals with clinical risk factors (REM Sleep Behavior Disorder or hyposmia) or genetic risk factors (mutations in LRRK2, GBA, or SNCA) for PD.

The [PPMI database](#) is open access. Raw data from recently occurring clinical visits are added to the database weekly; results from analyses of PPMI biospecimen are added to the PPMI database as analyses are completed. Curated data is not as current as the raw data (i.e., not available in real-time) because data curation is a time-intensive process. There may be discrepancies between the curated data in the MJFF data access center and the raw data in the PPMI database. The curated data currently available in the MJFF data access center were pulled from the PPMI database on *August 19, 2016*.

MJFF is interested in researcher feedback on the utility of curated PPMI data. Please email [resources@michaelfox.org](mailto:resources@michaelfox.org) for additional information or opportunities to contribute funding to accelerate the curation and harmonization of updated data cuts.

### UPLOADING TRANSMART-READY CURATED PPMI DATA INTO LOCAL INSTANCE

For users who are familiar with tranSMART in general and ETL procedures for this platform, please find instructions for loading tranSMART-ready curated files into a local instance below. tranSMART-ready files released through the MJFF data access center are intended for use with the [tMDataLoader](#) ETL script (see [tMDataLoader github](#) page for further details). If you are planning to use another ETL loader, please be advised that some changes to a Mapping file may be required to load the files. These files were validated and should load without any problems. Additional training and manual materials can be found [here](#). If you encounter any issues, please contact technical support at [techsupport@ranchobiosciences.com](mailto:techsupport@ranchobiosciences.com).

#### Prerequisites

1. Access to a local tranSMART server
2. Access to tranSMART database
3. WinSCP installed and configured to view tranSMART server
4. Logins/passwords to both database and tranSMART server

### *Loading tranSMART-Ready Files*

Note: The same commands described below can be executed using PuTTY via Unix.

1. Open SFTP-client: **WinSCP**
2. Transfer tranSMART-ready files into your local hard drive, noting the location.
3. Log into the server with your credentials.
4. Locate the directory on the server associated with tranSMART ETL procedures. This directory should have a .jar file for ETL, usually called "tm\_etl.jar". Create directory "data" in this directory, and then as many data folders as you wish. The folder structure in the data directory will correspond to the root tree directory in tranSMART. Here we use a sample directory called "Studies".
5. Open the studies folder **/etl/data/Studies**.
6. Upload your dataset in the studies folder.
7. Open SSH-client: **PuTTY**
8. Log into the tranSMART server with your credentials.
9. Execute the following instructions:  
\$ **cd /etl/tm\_etl** (Note: this is where you have your ETL script)  
\$ **java -jar tm\_etl.jar -i -n -c Config.groovy**
10. Loader will find all studies that do not have "\_DONE\_" and will upload them to tranSMART. Wait for the end of the process and check results in the local tranSMART instance.

## RESOURCES AND HELPFUL HINTS

- For additional questions on uploading curated data to a local tranSMART instance, please consult the [FAQ page](#) or email [techsupport@ranchobiosciences.com](mailto:techsupport@ranchobiosciences.com).
- The following PPMI study materials can be found in the [Research Documents & SOPs](#) section of the [PPMI website](#):
  - Study protocol
  - Schedule of activities
  - Case report forms (CRFs)
- For questions on the PPMI study and/or data collection methods (non-tranSMART-related questions), please email [ppmi@indd.org](mailto:ppmi@indd.org).
- For questions about accessing PPMI biospecimen, please email [resources@michaeljfox.org](mailto:resources@michaeljfox.org).