

Synthea Data in i2b2 Home



Synthetic patient data generated by Synthea can now be loaded into i2b2. The Synthea SyntheticMass sample files have been converted to i2b2-ACT format, and scripts to load Synthea data from scratch are available here: <https://github.com/i2b2/i2b2-synthea>

Synthea Load Process:

1. Set up an [i2b2 project with the ACT ontology](#).
2. Either download the SyntheticMass 63k sample in i2b2 format from https://github.com/i2b2/i2b2-synthea/blob/main/syntheamass_63K_sample.zip, or follow the instructions below to load any Synthea dataset from scratch.

Loading Synthea data from scratch

1. Download [SyntheticMass Data, Version 2 \(24 May, 2017\)](#)
 - All data sets (1k, COVID 10k, COVID 100k) have been verified to work EXCEPT the 100k patients in the large SyntheticMass Version 2 download.
 - The 100k patients in the large SyntheticMass Version 2 download needs an extra step to delete invalid records before import. In this case, download [synthea_cleanup.pl](#) to your disk, and then run "synthea_cleanup <directory-for-synthea-csv-files>" The fixed csv files will be in <directory-for-synthea-csv-files>/fixcsv.2. Set up an [i2b2 project with the ACT ontology](#).
 - Download the scripts from <https://github.com/i2b2/i2b2-synthea>
2. Run `create_synthea_table_<your_dbServerType>.sql` in your project to create the Synthea tables.
3. Import the Synthea data you downloaded in step one into the Synthea tables in your project.
4. Load the i2b2-to-SNOMED table in this repository into your project. https://www.nlm.nih.gov/healthit/snomedct/us_edition.html
 - Click on the "Download SNOMED-CT to ICD-10-CM Mapping Resources" link to download. (You will need a UMLS account.)
 - Unzip the file
 - Import the TSV file into a table called SNOMED_to_ICD10 in your database.
5. In Postgres and Oracle, follow the additional instructions in the comments at the top of `synthea_to_i2b2_<your_dbServerType>.sql` to clean up the date formatting.
6. Run `synthea_to_i2b2_<your_dbServerType>.sql` to convert synthea data into i2b2 tables (this will truncate your existing fact and dimension tables!)
 - Replace references to `i2b2metadata.dbo` in the script. Use the database and schema where your ACT ontology tables are.

Recent space activity



[Jeffrey Klann](#)

[Synthea Data in i2b2 Home](#) updated Jun 01, 2022 • [view change](#)

Space contributors

- [Jeffrey Klann](#) (700 days ago)