Synthetic Health Data Challenge Technical Guidance and Tips

What is Synthea? What are clinical disease modules?

Synthea is an open-source, fully synthetic set of electronic health record data developed by the MITRE Corporation that can be used to model a vast array of disease states and populations. More information regarding Synthea can be found in this article: https://academic.oup.com/jamia/article/25/3/230/4098271. Detailed information for using Synthea can be found here: https://github.com/synthetichealth/synthea/wiki.

Synthea's clinical disease modules are created by the community using a combination of clinical care protocols, publicly available disease incidence and prevalence statistics, and clinical expert feedback. Synthea uses these modules to generate individual synthetic patient health records, simulating the progression and treatment of disease from birth to death. Currently there are more than 120 modules and submodules available in Synthea. You can view many of the modules in the Synthea Module Gallery: https://github.com/synthetichealth/synthea/wiki/Module-Gallery

How do I install and run Synthea?

To install Synthea, see the basic and developer instructions available on the Synthea wiki.

- To install a local version of Synthea and generate synthetic patient records, use the basic setup instructions: https://github.com/synthetichealth/synthea/wiki/Basic-Setup-and-Running.
- To examine source code, create and test new modules, or edit modules, use the developer installation instructions: https://github.com/synthetichealth/synthea/wiki/Developer-Setup-and-Running.

QUICK TIP: For Windows users, be sure to create an environment variable called JAVA_HOME (if not automatically generated) that stores the file path to your JDK (Java Developer Kit) download. In addition, add "%JAVA_HOME%\bin" (no quotations) to the PATH environment variable.

How do I set Synthea parameters?

For information on Synthea basic setup, including setting parameters for altering the generated population, see examples on the Synthea wiki page: Synthea Basic Setup and Running Wiki page.

Figure 1: Common Configuration Options

```
This command takes additional parameters to specify different regions or common run options. Any options not specified are
left at the default value.

java -jar synthea-with-dependencies.jar [-h]
                                        [-s seed]
                                        [-cs clinician seed]
                                        [-p populationSize]
                                        [-g gender]
                                        [-a minAge-maxAge]
                                        [-m moduleFilter]
                                        [-c localConfigFilePath]
                                        [-d localModulesDirPath]
                                        [state [city]]
```

The Synthea Developer Setup and Running wiki provides syntax and examples for using parameters when running Synthea as a developer.

How do I specify a type of output for my Synthea files?

By default, Synthea stores the generated patient record outputs in FHIR R4 format; however, there are many formats available for output. The Synthea properties file can be modified directly after download. It is located at synthea/src/main/resources/synthea.properties in your local Synthea repository. The user may override settings in the Synthea properties file to modify how data are exported. For example, users may choose to enable exports to CSV format and disable exports to FHIR R4 format. Alternatively, users may choose to export to both formats.

Another option is to configure the output type when using the run command. For example, in order to export patient records to CSV format ONLY, include the following commands directly following the run command:

```
--exporter.fhir.export false --exporter.csv.export true
```

The "false" keyword disables the default export of patient data to FHIR R4 format while the "true" keyword enables exports to CSV. To record only the three most recent years of patient data in the output files, use the following command:

```
java -jar synthea-with-dependencies.jar --exporter.years_of_history 3
```

A list of all output configuration settings can be viewed at the following link:
https://github.com/synthetichealth/synthea/wiki/Common-Configuration

All generated output will be stored in a folder labeled "output" which the software automatically creates for the user. Within this folder, output in various formats will be stored in their own respective folders.
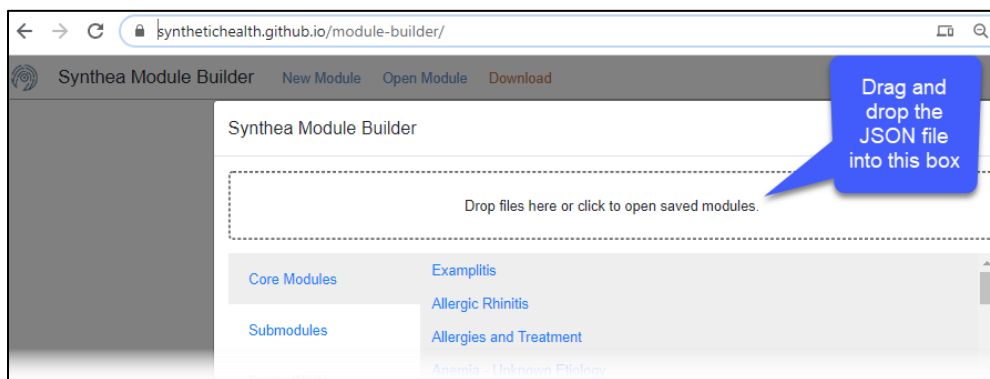
How do I view a module in the Synthea Module Builder?

The following steps summarize how to view a Synthea module JSON file using the Synthea Module Builder. The Synthea Module Builder allows users to view, interact, and edit Synthea modules.
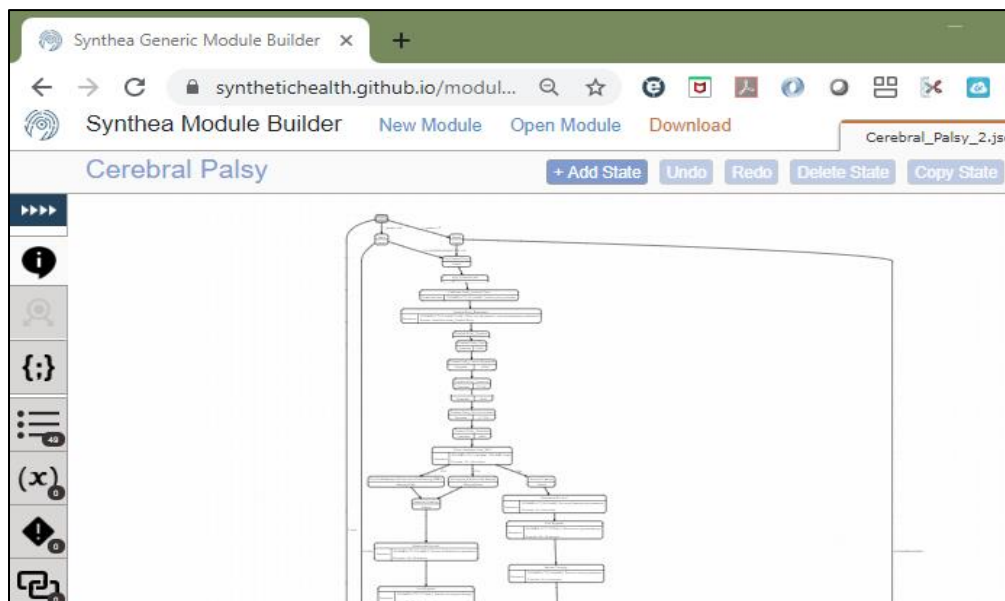
1. To view a JSON module, save the Synthea module JSON file to a location of your choice (e.g., desktop).
2. Visit the Synthea Module Builder using an internet browser: https://synthetichealth.github.io/module-builder/
3. To view an existing Synthea module or submodule, click on the module to view. If uploading a JSON module to view, drag and drop the saved JSON file to the Synthea Module Builder "drop files here" dialog box in your internet browser window (Figure 2):

Figure 2: Opening a Module in the Synthea Module Builder

4. After the file loads, the module will display in the Synthea Module Builder (Figure 3):

Figure 3: Viewing a Module in the Synthea Module Builder

3

How do I navigate a Synthea module in the Synthea Module Builder?

The left navigation bar in the Synthea Module Builder provides options for viewing and performing functions specific to a Synthea module.

Expand: Select this icon to expand and view additional module details. When the icon is not visible, content is expanded.

Hide: Select this icon to hide module details. When the icon is not visible, additional module details are hidden.

Module Remarks: Displays module remarks associated with the module that is under review.

State Editor: Provides the ability to add, edit, delete, copy, and paste states within the module that is under review. Note: This icon is only active for modules that are opened or created locally. This icon will be inactive when viewing core modules or submodules that reside within the Synthea Module Builder.

JSON Code Editor: Provides the ability to view and edit the module's JSON code directly.

State List: Displays a list of all states and types contained in the module. Double clicking on a state provides the ability to view all of the state details in the State Editor.

Attribute List: Displays a list of all attributes and states contained in the module. Double clicking on an attribute provides the ability to view all of the attribute details in the State Editor.

Warnings: Displays any issues/warning associated with the module. If the icon is displayed with a "0" in the bottom right corner, there are no warnings to review. If the icon contains a red circle with a number in the bottom right corner, warnings are present and should be reviewed.

Related Modules: Displays relationships to related submodules and their states. If the icon is displayed with a "0" in the bottom right corner, there are no related modules to review. If the icon contains a circle with a number in the bottom right corner, related modules exist and can be reviewed. Double clicking on a submodule state provides the ability to review the submodule state in the State Editor.

Where do I build a module for Synthea?

The Synthea Module Builder is available at: https://synthetichealth.github.io/module-builder/. Be sure to review the Generic Module Framework (GMF) Documentation prior to authoring modules. Familiarity with the GMF is required before authoring modules using the Synthea module builder.

How do I test and run a module built for Synthea?

To test or run a module built for Synthea, Synthea must be cloned locally to your desktop. See **How do I install and run Synthea** instructions above.

To copy the Synthea repository locally and install the necessary dependencies:

1. Download the module as a JSON file from the Synthea Module Builder.
2. Resolve any errors identified by the Synthea Module Builder.
3. Save the JSON file into the Synthea/src/main/resources/modules folder into your local Synthea repository.

To generate a patient population for a single module, run Synthea with the "-m" parameter in Terminal/Command Prompt, followed by the name of the module (e.g., m_COPD). To generate a patient population using all existing Synthea modules, use .\run_synthea command as instructed in the Synthea wiki.