**Bundle Splitting Data:**

**Goal:** To make the most utilization of synthetic data.

**Background:**

Synthetic data is a free resource that tests the capabilities of our code, where much of the data (patient demographics, conditions, observations, and encounters) is already provided in a realistic manner, such as spacing between appointments. Additionally, security is not a concern, as HIPAA does not apply to synthetic data. The prewritten code will save code writing time and cost in the long run. Synthetics allows testing of some the interactions and corner cases with CQL, such as age and hypertensive medication. Synthetics provides templates for areas we currently have no code, such as medications.

Additionally, we need to merge code from CDS Author, which has limitations in flexibility though easy to use, with hand-written CQL code within Atom. Therefore, the MITRE data must be imported into CQL-Atom.

BundleSplit.py is a Python program that splits all the synthetic patient bundled data from MITRE into the files and file structure (condition, patient, encounter, and observation) expected by ATOM/CQL. It is anticipated that the Python code can be modified, so that medications recommendations, and future grants requirements can easily be added with relatively facile modification.

**MITRE SYNTHETIC DATA:**

The Synthea data can be found here:

<https://synthea.mitre.org/downloads>

For this work the 1k patient data set was used.

<https://storage.googleapis.com/synthea-public/synthea_sample_data_fhir_stu3_sep2019.zip>

**INSTALLATION AND EXECUTION:**

**To use Bundle.py to split bundled patient data:**

1. Python3 must be installed, which comes standard with LINUX bundles.
2. The MITRE Synthetic data must be downloaded, *vide supra*.
3. A folder named JSON placed in the same directory as Bundle.py.
4. Place all the bundled .json files inside JSON directory.
5. Execute Bundle.py , which will make the correctly named patient folders, subdirectories, and files.
6. Place these folders into the FHIR/input/tests/individual\_test\_folder
7. Execute the CQL code

**RULES:**

Rules for importing any data have been generated from empirical development:

**General:**

* White space does not matter.
* Bracket position does not matter, but brackets must be paired.
* Directories must be closed before deleting in ATOM, strange but true.
* Commas must be added between items within brackets.
* No comma can be present at the end of the last group within a bracket.

**Patient:**

* “ID”: needs to be the same name as the directory

**Condition:**

* The condition needs to be named “condition-patient\_name-X”, where X represents the enumerated conditions, with a number.
* Subject Reference must be the “Patient/patient\_name”
* Note that the encounter name does not matter. All encounters can even be deleted and work.

**Observation:**

* The observation name needs to be “condition-patient\_name-X”, where X represents the enumerated observations, with a number.
* Strangely the line beginning with “issued”: will produce an error. The line has milliseconds in the time stamp, though if milliseconds or even the range are removed, the error still manifests.

**Encounter:**

* Currently, we do not use encounters in this grant, though we may in the future.

**EXAMPLE EXECUTION OF BundleSplit.py FOR 7 .json FILES**

root@DAP-PC:~/HBP/HBP/Scripts# BundleSplit.py

Splitting all the .json files in the JSON directory

Currently in directory: /mnt/c/Informatic/HBP/Scripts

Moving to: /mnt/c/Informatic/HBP/Scripts/JSON

Clearing out old files/directories the JSON directory

Splitting apart bundle file: Adolph80\_Turcotte120\_52b1b75f-2b8a-4319-9542-7abc39502cab.json

Searching through 11866 lines.

2 Condition file(s) from bundle Adolph80\_Turcotte120\_52b1b75f-2b8a-4319-9542-7abc39502cab.json

11 Encounter file(s) from bundle Adolph80\_Turcotte120\_52b1b75f-2b8a-4319-9542-7abc39502cab.json

86 Observation file(s) from bundle Adolph80\_Turcotte120\_52b1b75f-2b8a-4319-9542-7abc39502cab.json

1 Patient file(s) from bundle Adolph80\_Turcotte120\_52b1b75f-2b8a-4319-9542-7abc39502cab.json

Splitting apart bundle file: Agnes294\_Jenkins714\_185d26ad-fb9f-40ae-afb0-94d72827d887.json

Searching through 15293 lines.

8 Condition file(s) from bundle Agnes294\_Jenkins714\_185d26ad-fb9f-40ae-afb0-94d72827d887.json

21 Encounter file(s) from bundle Agnes294\_Jenkins714\_185d26ad-fb9f-40ae-afb0-94d72827d887.json

77 Observation file(s) from bundle Agnes294\_Jenkins714\_185d26ad-fb9f-40ae-afb0-94d72827d887.json

1 Patient file(s) from bundle Agnes294\_Jenkins714\_185d26ad-fb9f-40ae-afb0-94d72827d887.json

Splitting apart bundle file: Ana\_Mari╠üa762\_Balderas66\_7b0c45bd-3266-4e6e-b85f-436e909688b2.json

Searching through 14858 lines.

5 Condition file(s) from bundle Ana\_Mari╠üa762\_Balderas66\_7b0c45bd-3266-4e6e-b85f-436e909688b2.json

19 Encounter file(s) from bundle Ana\_Mari╠üa762\_Balderas66\_7b0c45bd-3266-4e6e-b85f-436e909688b2.json

82 Observation file(s) from bundle Ana\_Mari╠üa762\_Balderas66\_7b0c45bd-3266-4e6e-b85f-436e909688b2.json

1 Patient file(s) from bundle Ana\_Mari╠üa762\_Balderas66\_7b0c45bd-3266-4e6e-b85f-436e909688b2.json

Splitting apart bundle file: Ana\_Mari╠üa762\_Paredes726\_c85e59cb-50b9-4ae5-9961-12a5008e640a.json

Searching through 121866 lines.

11 Condition file(s) from bundle Ana\_Mari╠üa762\_Paredes726\_c85e59cb-50b9-4ae5-9961-12a5008e640a.json

140 Encounter file(s) from bundle Ana\_Mari╠üa762\_Paredes726\_c85e59cb-50b9-4ae5-9961-12a5008e640a.json

1390 Observation file(s) from bundle Ana\_Mari╠üa762\_Paredes726\_c85e59cb-50b9-4ae5-9961-12a5008e640a.json

1 Patient file(s) from bundle Ana\_Mari╠üa762\_Paredes726\_c85e59cb-50b9-4ae5-9961-12a5008e640a.json

Splitting apart bundle file: Andres25\_Konopelski743\_611e75ee-bca5-4cb4-9ccd-10f4b98e7732.json

Searching through 10980 lines.

9 Condition file(s) from bundle Andres25\_Konopelski743\_611e75ee-bca5-4cb4-9ccd-10f4b98e7732.json

10 Encounter file(s) from bundle Andres25\_Konopelski743\_611e75ee-bca5-4cb4-9ccd-10f4b98e7732.json

77 Observation file(s) from bundle Andres25\_Konopelski743\_611e75ee-bca5-4cb4-9ccd-10f4b98e7732.json

1 Patient file(s) from bundle Andres25\_Konopelski743\_611e75ee-bca5-4cb4-9ccd-10f4b98e7732.json

Splitting apart bundle file: Arturo47\_Cormier289\_b6db8b3b-db2d-499c-be9d-be5ea89bf0a3.json

Searching through 8100 lines.

7 Condition file(s) from bundle Arturo47\_Cormier289\_b6db8b3b-db2d-499c-be9d-be5ea89bf0a3.json

9 Encounter file(s) from bundle Arturo47\_Cormier289\_b6db8b3b-db2d-499c-be9d-be5ea89bf0a3.json

40 Observation file(s) from bundle Arturo47\_Cormier289\_b6db8b3b-db2d-499c-be9d-be5ea89bf0a3.json

1 Patient file(s) from bundle Arturo47\_Cormier289\_b6db8b3b-db2d-499c-be9d-be5ea89bf0a3.json

Splitting apart bundle file: Aurore252\_Shanahan202\_24b1778f-54bf-4d39-b6b7-6cc780bd3f24.json

Searching through 125495 lines.

22 Condition file(s) from bundle Aurore252\_Shanahan202\_24b1778f-54bf-4d39-b6b7-6cc780bd3f24.json

140 Encounter file(s) from bundle Aurore252\_Shanahan202\_24b1778f-54bf-4d39-b6b7-6cc780bd3f24.json

1391 Observation file(s) from bundle Aurore252\_Shanahan202\_24b1778f-54bf-4d39-b6b7-6cc780bd3f24.json

1 Patient file(s) from bundle Aurore252\_Shanahan202\_24b1778f-54bf-4d39-b6b7-6cc780bd3f24.json

Finished. A total of 7 bundled .json file(s) processed

root@DAP-PC:~/HBP/HBP/Scripts

**EXAMPLE CQL-ATOM EXECUTION AFTER BundleSplit.py (only 1 of 7 shown)**

Test Andres25\_Konopelski743\_611e75ee-bca5-4cb4-9ccd-10f4b98e7732

Fri Nov 27 2020 15:34:32 GMT+0300 (Moscow Standard Time)

Patient=org.hl7.fhir.dstu3.model.Patient@32be0d82

Adult at start of Measurement Period=true

Blood Pressure Observations=[org.hl7.fhir.dstu3.model.Observation@fb350b5b, org.hl7.fhir.dstu3.model.Observation@87c0b548, org.hl7.fhir.dstu3.model.Observation@56c6b08a]

Avg Systolic BP=116.71118704 'mm[Hg]'

Avg Diastolic BP=79.31592410 'mm[Hg]'

Avg BP=Tuple {

"systolic": 116.71118704 'mm[Hg]'

"diastolic": 79.31592410 'mm[Hg]'

}

Normal BP=null

Elevated BP=null

HTN Stage 1 BP=null

HTN Stage 2 BP=null

HTN Crisis BP=null

Systolic BP Variability=10.10130053 'mm[Hg]'

Systolic BP Variability String=10.1 mm[Hg]

Last 4 Blood Pressure Observations=[org.hl7.fhir.dstu3.model.Observation@56c6b08a, org.hl7.fhir.dstu3.model.Observation@87c0b548, org.hl7.fhir.dstu3.model.Observation@fb350b5b]

Avg Last 4 Systolic BP=116.71118704 'mm[Hg]'

Avg Last 4 Diastolic BP=79.31592410 'mm[Hg]'

Avg Last 4 BP=Tuple {

"systolic": 116.71118704 'mm[Hg]'

"diastolic": 79.31592410 'mm[Hg]'

}

Avg Last 4 BP String=117/79 mm[Hg]

Avg BP String=117/79 mm[Hg]

elapsed: 176.74 seconds