OMOP Abstractor

Michael Gurley
m-gurley@northwestern.edu
Northwestern Applied Research Informatics Group (NUARIG)

NUAIRG:
Firas Wehbe, MD, PhD
Yulia Bushmanova

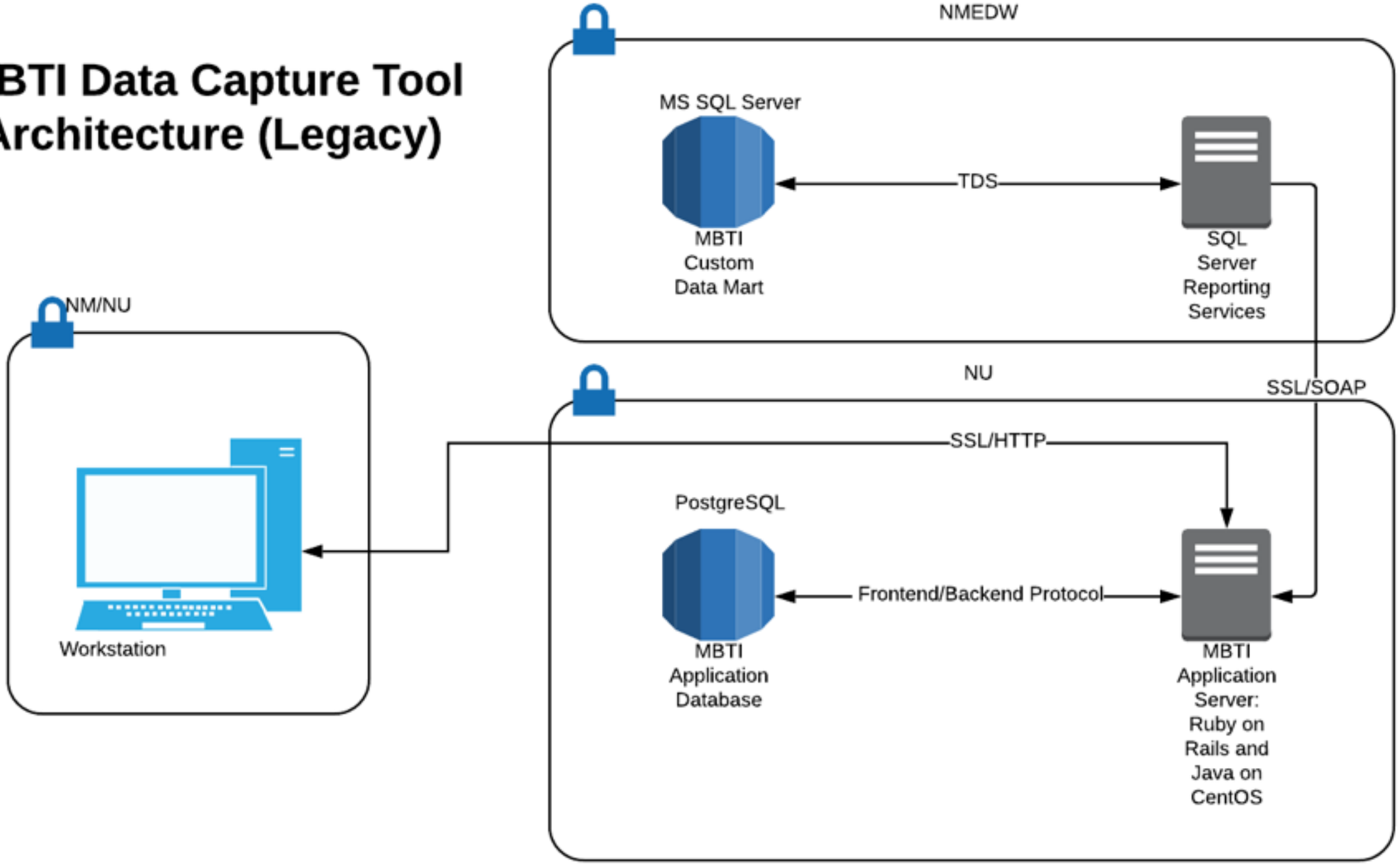
NMEDW:
Daniel Schneider
Prasanth Nannapaneni
Martin Borsje

Northwestern

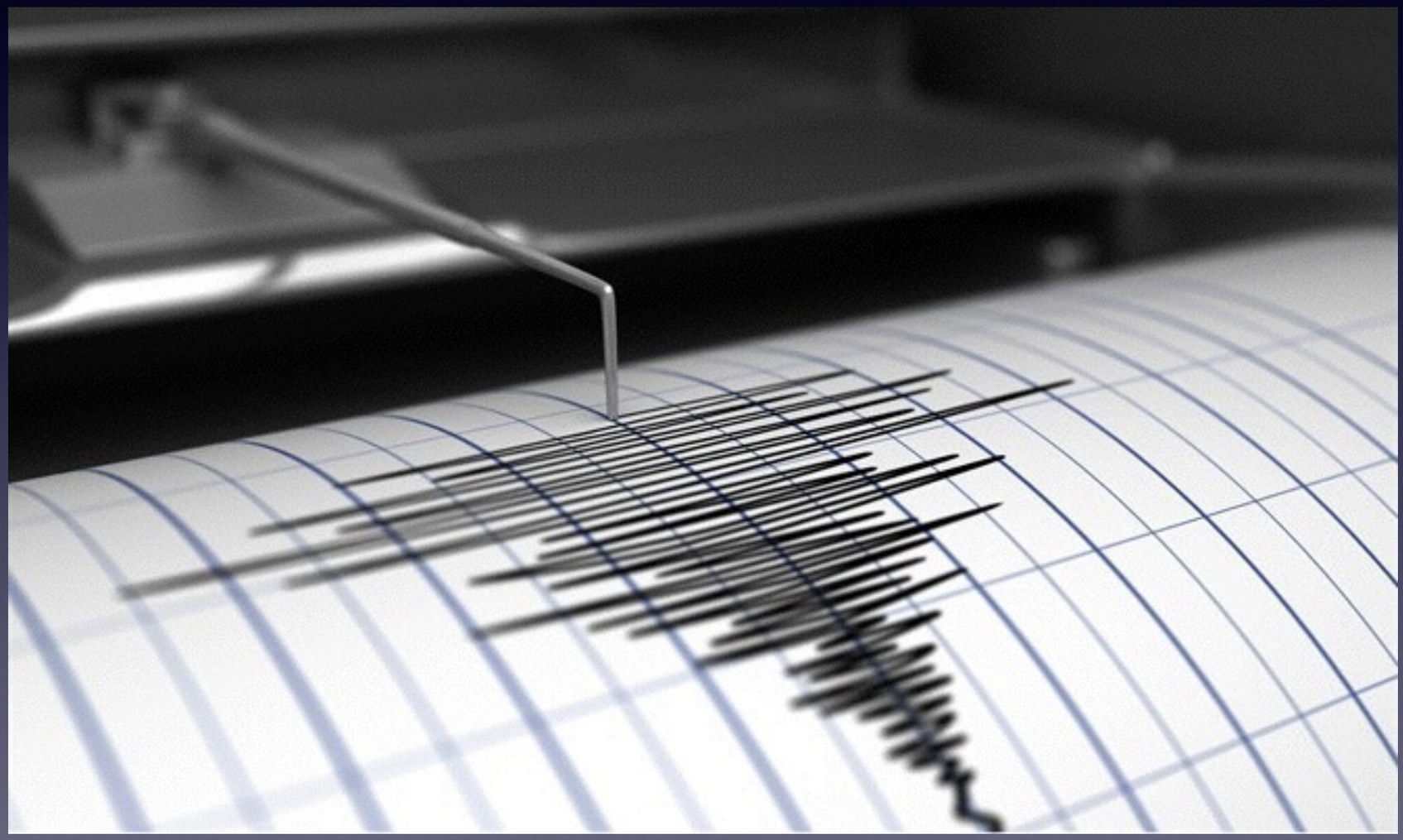
Background:
MBTI Data Capture Tool

- The Northwestern Malnati Brain Tumor Institute outcomes/research database: The MBTI Data Capture Tool (MBTI-DCT).
- Backed by a custom ETL/data mart within the Northwestern Enterprise Data Warehouse (NMEDW) (Microsoft SQL Server and SQL Server Integration Services).
- Aggregates data from Northwestern Medicine's (NM) clinical systems: Epic (Outpatient), Cerner (Inpatient, Surgery, Pathology, Radiology) and MOSAIQ (Radiation Oncology).
- Custom data model.
- Incremental loads of the custom ETL/data mart into a PostgreSQL database via SQL Server Reporting Services (SSRS) exposed as SOAP Web Services.
- The MBTI-DCT includes an NLP-aided chart abstraction user interface for the curation of non-discrete data points from clinical narratives:
    - Anatomic site/histology, WHO grade, recurrence status, pathology findings from pathology reports.
    - Extent of Resection from surgical procedure reports.
    - Anatomical target of radiotherapy from radiation oncology summaries.
    - Performance status declarations, outside treatments from clinic progress notes.
    - Recurrence/progression declarations from imaging exam reports.
- Ruby on Rails user interface. Custom NLP pipeline using the Stanford NLP Java Library and Lingscope.

MBTI Data Capture Tool Architecture (Legacy)



Data Earthquake



Change Happens

- NM merged with other hospital(s).
- Large project to migrate three Epic instances into one instance.
- NM migrated from using Cerner Surginet for tracking surgeries to Epic.
- NMEDW deprecated legacy traditionally modeled “Integrated Data Structures” to new fact/dimension modeled “Integrated Data Structures”.
- Programatic access to SSRS reports via SOAP services was disabled.

Remediation Choice

- All this change broke the MBTI-DCT.
- The custom ETL/data mart's flow of new data stopped because of reliance on deprecated structures and data access strategies.
- Decision:
    - Remediate the ETL's population of the custom data model.

      OR

    - Replace the custom data model with a common data model (CDM) and remediate the MBTI-DCT UI and application logic to work with a CDM.
- All of Us and eMerge CDM activities were ramping up around this time. Laying the ground work for NM investing in transforming its NMEDW into the OMOP CDM.

Choose OMOP: Challenges

- We decided to remediate the MBTI-DCT UI and application logic to work with the OMOP CDM.
- Challenges:
    - Work with a truncate/reload data refresh model. Our abstraction/NLP output/curation tables need stable structures to hang off of across data reloads.
    - Remediating MBTI-DCT UI to work with the OMOP CDM.
    - Representing PHI.
    - Include all surgeries in the OMOP instance.
    - Include all pathology procedures in the OMOP instance.
    - Include all pathology reports sections in the OMOP instance.
    - Preserve and represent references between surgeries, pathology procedures and pathology report sections.
    - De-duping surgeries and pathology procedures.

Abstractor Library

- The legacy MBTI-DCT UI was backed by an NLP-aided assisted chart abstraction library, ‘abstractor’, created by NUAIRG.
- Abstractor’s existing data model had wide support for specifying NLP expectations for chart abstraction. Beyond what is supported via the OMOP ‘note_nlp’ table.
- Abstractor allows for the setup of ‘expectations’ necessary for NLP-aided assisted chart abstraction.
    - Event Cohort Expectations
        - Flows of clinical events. An abstractor ‘namespace’ is a set of saved search criteria to produce an ‘event cohort’. For example: a ‘Surgical Pathology’ namespace that binds to all ‘Final Diagnosis Section’ notes related to ‘Surgical pathology procedure’ procedure occurrences
    - Data Point Expectations
        - Sets of abstractable data points for a specific type of clinical note or clinical event (subject). For example, (‘Site, ‘Histology’, ‘WHO Grade’, ‘IDH1 Status’, etc.) for a clinical note (subject).
- Abstractor ties these two sets of expectations together to allow for an NLP library to generate suggestions for a defined set data points for a defined set of clinical events.
- Abstractor presents these flows of expectations (clinical events, data points and NLP suggestions) within a user interface for curation.

Abstractor: Namespaces

- An abstractor namespace is a named search criteria that is instantiated within the ‘abstractor_namespaces’ table. The search criteria are specified in the ‘joins_clause’ and ‘where_clause’ columns. The ‘subject_type’ column indicates the subject or clinical event that is the target of the namespace. The ‘id’ of the ‘abstractor_namespaces’ table can be referenced via in the ‘namespace_id’ column of the ‘abstractor_subjects’ table to bind data points to the ‘namespace’.

- Abstractor has a schedulable ‘suggestor’ job that can cycle through all setup namespaces, execute each namespace’s saved search criteria to return a list of subjects, discover all the data points points bound to the subject via the namespace and ask each data point’s NLP configuration to generate NLP suggestions for the subject in hand.

- Bookkeeping that a subject for a given namespace has been processed is tracked in the ‘abstractor_namespace_events’ table.

Abstractor:
Subject, Predicate, Object

- Abstractor uses the familiar paradigm of subject, predicate, object to model data point expectations.
    - **Subject**: the ‘abstractor_subjects’ table points to type of entity via the ‘subject_type’ column and an abstractor namespace via the ‘namespace_id’ column. For OMOP, normally the note table (or more precisely, as we will later see, the note_stable_identifier table) is the value of the ‘subject_type’ column. The ‘abstractor_subjects’ table also points to a ‘predicate’ via the ‘abstractor_abstraction_schema_id’ column.
    - **Predicate**: the abstractor_abstraction_schemas table describes a property of the subject. For example, the ‘has_icdo3_histology’ property of declaring the histology of a cancer diagnosis. A predicate can have different types of possible answers. This is captured in the abstractor_object_type_id column: list, number, boolean, string, date and text Some predicates can have multiple syntactical expressions or synonyms. This is specified in the ‘abstractor_abstraction_schema_predicate_variants’ table.
    - **Object**: the ‘abstractor_object_values’ table describes, for ‘list’ predicates, the range of possible values that can satisfy the property of the subject. Some objects can have multiple syntactical expressions or synonyms. This is specified in the ‘abstractor_object_value_variants’ table.
- For those data points that are only meaningfully collected in groups (like a cancer diagnosis ‘site’ and ‘histology) and with possible cardinality > 1, abstractor allows for repeating grouping of data points via the ‘abstractor_subject_groups’ table.

Abstractor: Sources

- Abstractor has the concept of ‘sources’ for each data point for each subject. The ‘abstractor_abstraction_sources’ table specifies in the ‘from_method’ column the textual content of the subject. Normally, for OMOP this will be the ‘note_text’ column of the ‘note’ table but could also be the ‘note_title’ column.

- Each source entry must specify the expected format of how a predicate appears in content for the subject in the abstractor_rule_type_id’ column. Here are the possible rule types:

    - ‘name/value’: both the predicate and the value are expected within a suggestible mention. For example, ‘KPS: 20’ or ‘The patient has a KPS of 20.’

    - ‘value’: only the value is expected within a suggestible mention. For example: ‘Pituitary adenoma (see Microscopic Description).’.

- Each source entry must specify a strategy for making suggestions in the ‘abstractor_abstraction_source_type_id’ column. Here are the possible strategies:

    - ‘nlp suggestion’: delegate the generation of suggestions to abstractor’s built-in NLP library. A simple sentence-splitting, rule-based named entity recognizer with negation detection.

    - ‘custom nlp suggestion’: delegate the generation of suggestions to an external NLP library via communication over a RESTful interface. Not confined to rule-based NLP. Abstractor sends a document to the configured endpoint with an embedded endpoint URL containing the dictionary for each data point and a call back endpoint to send back suggestions.

Abstractor: Suggestions and Abstractions

- Abstractor has the concept of ‘suggestions’ for each data point for each subject. The ‘abstractor_suggestions' table contains the NLP suggested value or suggestion of ‘unknown’ or ‘not applicable’. The underlying textual evidence supporting an NLP library’s suggestion is contained within the ‘match_value’ and ‘sentence_match_value’ columns of the ‘abstractor_abstraction_sources’ table. Abstractor automatically makes a suggestion of ‘unknown’ and ‘not applicable’ for all data points.

- A curator’s accepted ‘suggestion’ for a data point is recorded within the ‘abstractor_abstractions’ table. This table is the center of the abstractor universe.

- Grouped, relating accepted data points are collected within the ‘abstractor_abstraction_groups’ table.

OMOP Abstractor

Green = OMOP CDM Table

Yellow = OMOP CDM Extension Table

Blue = OMOP Abstractor Expectation Table

Pink = OMOP Abstractor Abstraction Table

ABSTRACTOR_OBJECT_VALUE_VARIANTS

PK	id
FK	abstractor_object_value_id
	vocabulary_code
	value
	case_sensitive

ABSTRACTOR_OBJECT_TYPES

PK	id
	value

ABSTRACTOR_ABSTRACTION_SCHEMA_PREDICATE_VARIANTS

PK	id
FK	abstractor_abstraction_schema_id
	value

ABSTRACTOR_OBJECT_VALUES

PK	id
	value
	vocabulary_code
	vocabulary
	vocabulary_version
	case_sensitive

ABSTRACTOR_ABSTRACTION_SCHEMAS

PK	id
	display_name
	predicate
FK	abstractor_object_type_id
	preferred_name

ABSTRACTOR_ABSTRACTION_SOURCE_TYPES

PK	id
	name

ABSTRACTOR_RULE_TYPES

PK	id
	name
	description

ABSTRACTOR_ABSTRACTION_SCHEMA_OBJECT_VALUES

PK	id
FK	abstractor_abstraction_schema_id
FK	abstractor_object_value_id
	display_order

ABSTRACTOR_SUBJECTS

PK	id
FK	abstractor_abstraction_schema_id
	subject_type
	dynamic_list_method
	namespace_type
	namespace_id

ABSTRACTOR_ABSTRACTION_SOURCES

PK	id
FK	abstractor_subject_id
	from_method
	custom_method
FK	abstractor_abstraction_source_type_id
FK	abstractor_rule_type_id
	section_name
	custom_nlp_provider

ABSTRACTOR_NAMESPACES

PK	id
	name
	subject_type
	joins_clause
	where_caluse

ABSTRACTOR_SUBJECT_GROUPS

PK	id
	name
	cardinality
	subtype
	enable_workflow_status
	workflow_status_submit
	workflow_status_pend

ABSTRACTOR_ABSTRACTION_OBJECT_VALUES

PK	id
FK	abstractor_abstraction_id
FK	abstractor_object_value_id

ABSTRACTOR_SUGGESTION_SOURCES

PK	id
FK	abstractor_abstraction_source_id
FK	abstractor_suggestion_id
	match_value
	sentence_match_value
	source_id
	source_method
	source_type
	custom_method
	custom_explanation
	section_name

ABSTRACTOR_ABSTRACTIONS

PK	id
FK	abstractor_subject_id
	value
	about_type
	about_id
	unknown
	not_applicable
	workflow_status
	workflow_status_whodunnit

ABSTRACTOR_NAMESPACE_EVENTS

PK	id
FK	abstractor_namespace_id
	eventable_type
	eventable_id

NOTE_STABLE_IDENTIFIER

PK	id
FK	note_id
	stable_identifier_path
	stable_identifier_value

NOTE

PK	note_id
	person_id
	note_date
	note_datetime
	note_type_concept_id
	note_class_concept_id
	note_title
	note_text
	encoding_concept_id
	language_concept_id
	provider_id
	visit_occurrence_id
	visit_detail_id
	note_source_value

ABSTRACTOR_SUBJECT_GROUP_MEMBERS

PK	id
FK	abstractor_subject_id
FK	abstractor_subject_group_id
	display_order

ABSTRACTOR_ABSTRACTION_GROUPS

PK	id
FK	abstractor_subject_group_id
	about_type
	about_id
	system_generated
	subtype

ABSTRACTOR_ABSTRACTION_GROUP_MEMBERS

PK	id
FK	abstractor_abstraction_group_id
FK	abstractor_abstraction_id
	about_id
	system_generated
	subtype

ABSTRACTOR_SUGGESTIONS

PK	id
FK	abstractor_abstraction_id
	suggested_value
	unknown
	not_applicable
	accepted

ABSTRACTOR_SUGGESTION_OBJECT_VALUES

PK	id
FK	abstractor_suggestion_id
FK	abstractor_object_value_id

ABSTRACTOR_SUGGESTION_OBJECT_VALUE_VARIANTS

PK	id
FK	abstractor_suggestion_id
FK	abstractor_object_value_variant_id

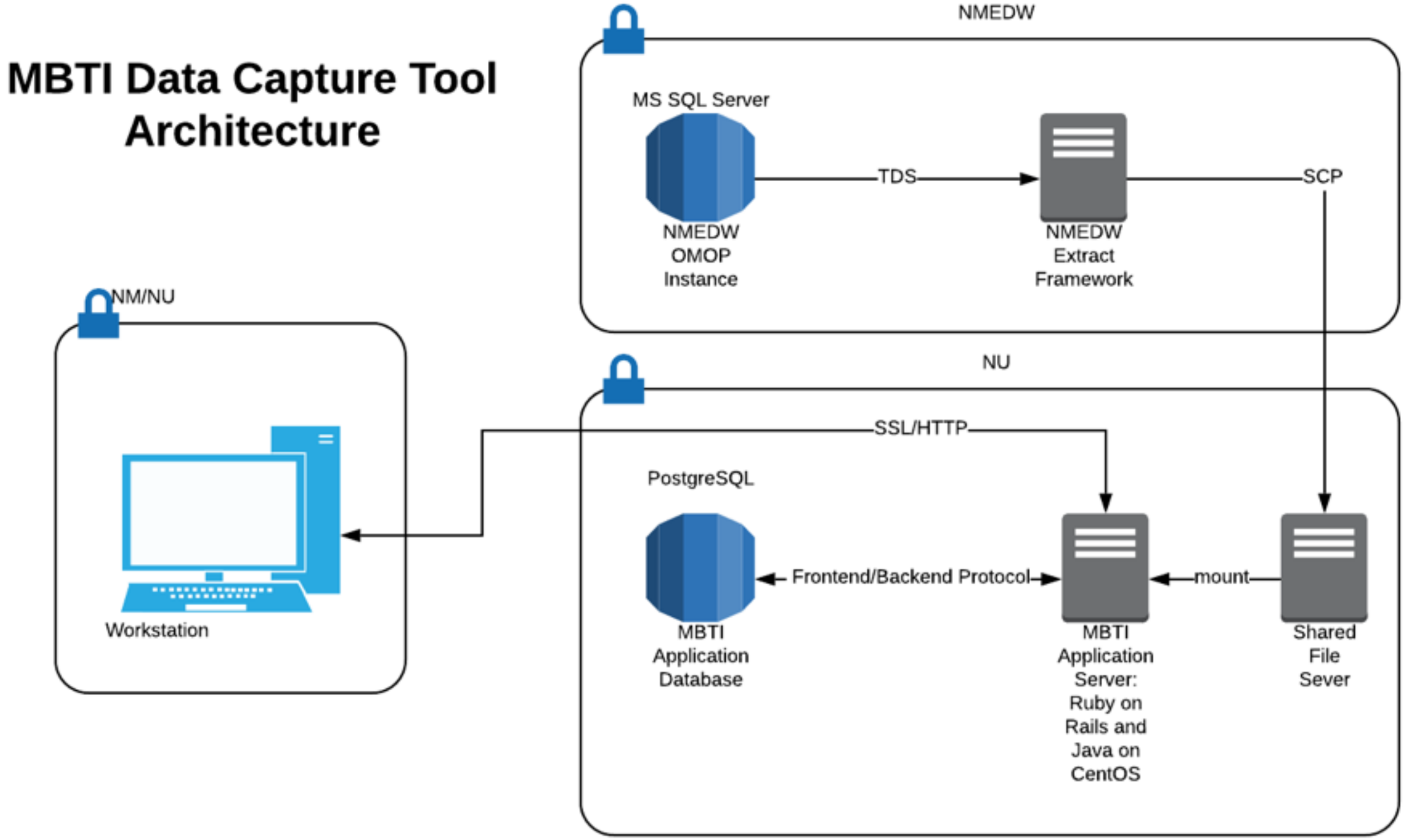
Challenge:
Work with a truncate/reload data refresh model.

- New load strategy:
    - Replacement of programmatic access to data via SSRS.
    - NMEDW has a single OMOP instance that is partitioned during extraction by a cohort definition for the MBTI.
    - The NMEDW extract framework deposits files on a shared folder mounted to the MBTI-DCT application server.
        - Incremental loads not feasible.
- Stable Identifiers
    - OMOP internal IDs change across truncate/reload data refreshes.
    - Our abstraction/NLP output/curation tables need stable structures to hang off of across data reloads.
    - Asked the NEMEDW OMOP data architect team to populate stable identifier tables for NOTE and PROCEDURE_OCCURRENCE.
    - The stable identifier tables contains an invariant ‘id’ column that stays stable across loads, an OMOP internal id column (‘note_id’, ‘procedure_occurrence_id’) that changes across loads and invariant pointers to source-system row-level provenance via the ‘stable_identifier_path’ and ‘stable_identifier_value’ columns.

NOTE_STABLE_IDENTIFIER

```
CREATE TABLE public.note_stable_identifier
(
  id bigint NOT NULL DEFAULT nextval('note_stable_identifier_id_seq'::regclass),
  note_id bigint NOT NULL,
  stable_identifier_path character varying NOT NULL,
  stable_identifier_value character varying NOT NULL,
  CONSTRAINT note_stable_identifier_pkey PRIMARY KEY (id)
)
```

MBTI Data Capture Tool Architecture



Challenge:
Remediating MBTI-DCT UI
to work with the OMOP CDM
Index Page

- Legacy MBTI-DCT UI was based on a custom data model. Each type of note targeted for NLP-aided chart abstraction had its own dedicated table: pathology reports, imaging exam reports, clinic progress notes and radiation oncology summaries. Each note type had its own dedicated index and edit page.
- Index Page
    - Remediate one index screen to display notes needing curation by abstraction ‘namespaces’.
    - Remediate the index screen to allow for searching:
        - By keyword search across OMOP PHI tables and the note text.
        - By note date.
        - By providers associated with a first level procedures (for example, pathology procedures) and second level procedures (for example, surgeries).

Challenge:
Remediating MBTI-DCT UI
to work with the OMOP CDM
Edit Page

- Edit Page
  - Remediate the edit screen to display data from the OMOP PHI tables.
  - Remediate the edit screen to display the list of abstractable data points for a note based on the set of abstractable data points bound to a ‘namespace’.
  - Remediate the edit screen to display associated first level and second level procedures.
  - Remediate the edit screen to display other note entries associated to first level procedures (for example, all other “sections” of the current pathology report section)

MBTI | Notes

0.0.0.0:3000/notes?utf8=✓&abstraction_status=needs+review&namespace_id=19&search=&de

Search

Northwestern MBTI Data Capture Tool

Notes

Reviewed?

needs review

Namespace

- Molecular Pathology
- Surgical Pathology
- Outside Surgical Pathology

Search

From

To

Provider

× HORBINSKI, CRAIG

SEARCH Clear

1 2 3 4 5 6 7 8 9 10 11 12

Note Date	Note Type	Note Title	First Name	Last Name	MRN(s)	
01/01/1900	Note	Final Diagnosis	Bob	Jones	**Northwestern** 000000000	Review
01/01/1900	Note	Final Diagnosis	Bob	Jones	**Northwestern** 000000000	Review
01/01/1900	Note	Final Diagnosis	Bob	Jones	**Northwestern** 000000000	Review
01/01/1900	Note	Final Diagnosis	Bob	Jones	**Northwestern** 000000000	Review

Note

Back | Notes | Previous | Next

Patient	MRN(s):	Note Date	Note Type	Note Class	Title	Provider
Bob N Jones	**Northwestern** 000000000	01/01/1900	Note	No matching concept	Final Diagnosis	

Procedures

Procedure	Date	Provider	Specimens	Notes	
Surgical pathology procedure	01/01/1900	HORBINSKI, CRAIG M.		Specimen/Gross Description	VIEW
				Intraprocedural Consultation Findings	VIEW
				Clinical Information	VIEW
				Surg Path Non-Chartable Comment	VIEW
				Addendum	VIEW
Craniectomy, trephination, bone flap craniotomy; for excision of meningioma, supratentorial	01/01/1900	CHANDLER, JAMES P.			

NOT APPLICABLE ALL | UNKNOWN ALL

Note text

NOT APPLICABLE ALL | UNKNOWN ALL

Metastatic Cancer

Histology

- [x] adenocarcinoma, metastatic (8140/6)
- [ ] not applicable
- [ ] unknown

Edit | CLEAR

Site

- [ ] cerebellum, nos (c71.6)
- [ ] not applicable
- [ ] unknown

Edit | CLEAR

Primary Site

- [ ] not applicable
- [ ] unknown

Edit | CLEAR

Laterality

- [ ] not applicable
- [ ] unknown

Edit | CLEAR

Recurrent

- [ ] not applicable
- [ ] unknown

Edit | CLEAR

NOT APPLICABLE GROUP | UNKNOWN GROUP

ADD METASTATIC CANCER

Note text

A and B. Tumor, cerebellum, resection:
Metastatic adenocarcinoma (see Note).

Note: This tumor shows tall columnar cells with lumenal necrosis, the combination of which is a classic hallmark of colorectal adenocarcinoma.

IDH1 Status

- [ ] not applicable
- [ ] unknown

Edit | CLEAR

IDH2 Status

- [ ] not applicable
- [ ] unknown

Edit | CLEAR

1P Status

- ○ deleted ○ non-deleted

SAVE Cancel

19q Status

- [ ] not applicable
- [ ] unknown

Edit | CLEAR

10q/PTEN Status

- [ ] not applicable
- [ ] unknown

Edit | CLEAR

MGMT promoter methylation status Status

- [ ] not applicable
- [ ] unknown

Edit | CLEAR

ki67

- [ ] not applicable
- [ ] unknown

Edit | CLEAR

p53

- [ ] not applicable
- [ ] unknown

Note text

A and B. Tumor, cerebellum, resection:
Metastatic adenocarcinoma (see Note).

Note: This tumor shows tall columnar cells with lumenal necrosis, the combination of which is a classic hallmark of colorectal adenocarcinoma.

Challenge:
PHI

- The MBTI-DCT displays PHI. The display of PHI within the MBTI-DCT is necessary to meet curation and outcomes use cases.

- Solution: adopted and reused the PHI table specified and populated for the All of Us project.
  - pii_address
  - pii_email
  - pii_mrn
  - pii_name
  - pii_phone_number

Challenge:
Include all surgeries

- Validate that surgeries were ***not*** being included in the current OMOP build.
- Surgeries at NM are tracked in the new consolidated Epic instance.
- For Epic, the current OMOP build was populating the PROCEDURE_OCCURRENCE table exclusively from charge-oriented tables
    - HSP_ACCT_PX_LIST
    - HSP_ACCT_CPT_CODES
    - HSP_TRANSACTIONS
    - ARPB_TRANSACTIONS
- Asked the NEMEDW OMOP data architect team to pull from actual Epic surgery tables.
    - or_log
    - or_log_all_proc
    - or_proc
    - or_proc_cpt_id
- Automatic mapping to Procedure domain standardized vocabulary entries was achieved by using the or_proc_cpt_id.real_cpt_code field.

Opinion Sidebar:
Prefer Small (accurate) Data
over
Big (messy) Data

- Simple Determinism is better than Clever Probabilism.
  - If a source system represents a class of clinical events in a discrete manner to support a clinical workflow (like Epic does for surgeries), prefer this canonical representation to the clinical event welter caused by charge-oriented representations.
- Open question:
  - Exclude charge-related representations?


YEAH, WELL, YOU KNOW,
THAT'S JUST, LIKE,
YOUR OPINION, MAN.
imgflip.com

Challenge:
Include all pathology procedures

- Validate that pathology procedures were ***not*** being included in the current OMOP build.
- Pathology procedures at NM are tracked in Cerner Pathnet Anatomic Pathology.
- For Cerner, the current OMOP build was populating the PROCEDURE_OCCURRENCE table from charge-oriented tables
    - encounter (from Cerner)
    - EPSI charge tables
- Asked the NEMEDW OMOP data architect team to pull from actual Cerner Pathnet Anatomic Pathology tables.
    - pathology_case
    - prefix_group
    - case_specimen
    - case_report
    - clinical_event
    - ce_blob
- Made mappings from local prefix_group entries to standardized Procedure domain entries in the SNOMED vocabulary. Mostly along this axis: Procedure | Labratory Procedure (procedure) | Anatomic Pathology Procedure. See spreadsheet.
- Need to to map local Cerner case_specimen specimen codes to the Specimen domain entries in the SNOMED vocabulary.

Challenge:
Include all pathology reports sections in the OMOP instance.

- Validate that pathology reports sections were being included in the current OMOP build ***as separate entries in the NOTE table with the section name populating the note_title field.***
- Pathology reports sections at NM are tracked in Cerner Pathnet Anatomic Pathology.
- Pathology reports are written in ‘sections’. Each section having a dedicated purpose. For example: ‘Final Diagnosis’,’Microscopic Description’, ‘Specimen/Gross Description’ and ‘Clinical Information’.
- Most often the data points desired to be extracted from a pathology report reside in the ‘Final Diagnosis’ section.
- Other ‘sections’ can often be the source of false positives for NLP pipelines. For example, historical diagnoses mentioned in the ‘Clinical Information’ section.
- Simple Determinism is better than Clever NLP.
    - If a source system splits a pathology reports into discrete labeled sections to support a clinical workflow (like Cerner does for pathology reports), prefer this canonical representation instead of a multi-section conglomerated representation.
- Don’t use the sectionizing component of your NLP pipeline if your source system sectionizes for you.

Challenge:
Preserve and represent references between surgeries, pathology procedures and pathology report sections.

- Validate that references between surgeries, pathology procedures and pathology report sections were ***NOT*** being included in the current OMOP build.
- The conventional advice to tie OMOP clinical events by joining to VISIT_OCCURRENCE ***is insufficient***.
    - Possible for one VISIT_OCCURRENCE to span multiple surgery entries and multiple pathology procedures in PROCEDURE_OCCURRENCE.
    - Possible for one VISIT_OCCURRENCE entry to span multiple pathology reports.
- Make explicit references between surgical PROCEDURE_OCCURRENCE entires and pathology PROCEDURE_OCCURRENCE entires via FACT_RELATIONSHIP.
    - Asked the NEMEDW OMOP data architect team to build a join table within the NMEDW integrated data structures associating Cerner pathology procedures and Epic surgeries. Match on patient and surgery date to pathology accession date/case collection date. ETL the join table into the OMOP FACT_RELATIONSHIP table
- Make explicit references between pathology report section NOTE entires and pathology PROCEDURE_OCCURRENCE entires via FACT_RELATIONSHIP (we are not on OMOP CDM version yet where this can be done directly within the NOTE table with note_event_id and note_event_field_concept_id).
    - Asked the NMEDW OMOP data architect team to build a join table within the NMEDW integrated data structures associating Cerner pathology procedures and Cerner pathology report sections. ETL the join table into the OMOP FACT_RELATIONSHIP table

```
SELECT note.note_id
     , note.note_date
     , note_stable_identifier.id
     , note_stable_identifier.stable_identifier_path
     , note_stable_identifier.stable_identifier_value
     , note.note_title
     , note.note_text
     , procedure_occurrence.procedure_occurrence_id
     , procedure_occurrence.procedure_concept_id
     , concept.concept_code
     , procedure_occurrence.procedure_date
     , procedure_occurrence_stable_identifier.id
     , procedure_occurrence_stable_identifier.stable_identifier_path
     , procedure_occurrence_stable_identifier.stable_identifier_value_1
     , prov1.provider_name
     , prov2.provider_name
     , posi2.id
     , posi2.stable_identifier_path
     , posi2.stable_identifier_value_1
FROM note_stable_identifier JOIN note                                    ON stable_identifier_value_1.note_id = note.note_id
                            JOIN fact_relationship                       ON fact_relationship.domain_concept_id_1 = 5085 AND fact_relationship.fact_id_1 = note.note_id AND fact_relationship.relationship_concept_id = 44818790
                            JOIN procedure_occurrence                    ON fact_relationship.domain_concept_id_2 = 10 AND fact_relationship.fact_id_2 = procedure_occurrence.procedure_occurrence_id AND procedure_occurrence.procedure_concept_id = 4213297
                            JOIN procedure_occurrence_stable_identifier  ON procedure_occurrence.procedure_occurrence_id = procedure_occurrence_stable_identifier.procedure_occurrence_id
                            JOIN concept                                 ON procedure_occurrence.procedure_concept_id = concept.concept_id
                            JOIN fact_relationship AS fr2                ON fr2.domain_concept_id_1 = 10 AND fr2.fact_id_1 = procedure_occurrence.procedure_occurrence_id AND fr2.relationship_concept_id = 44818888
                            JOIN procedure_occurrence pr2                ON fr2.domain_concept_id_2 = 10 AND fr2.fact_id_2 = pr2.procedure_occurrence_id
                            JOIN procedure_occurrence_stable_identifier posi2 ON pr2.procedure_occurrence_id = posi2.procedure_occurrence_id
                            JOIN provider prov1                          ON procedure_occurrence.provider_id = prov1.provider_id
                            JOIN provider prov2                          ON pr2.provider_id = prov2.provider_id
WHERE note.note_title = 'Final Diagnosis'
```

FACT_RELATIONSHIP

- Entry between pathology report section in NOTE and pathology procedure in PROCEDURE_OCCURRENCE: relationship_concept_id = 44818790 ‘Has procedure context (SNOMED).’ Plus converse entry.

- Entry between pathology procedure in PROCEDURE_OCCURRENCE and surgery in PROCEDURE_OCCURRENCE: relationship_concept_id = 44818888 ‘Procedure context of (SNOMED)’ Plus converse entry.

Challenge:
De-duping surgeries and pathology procedures

- Entries in FACT_RELATIONSHIP allow us to surface these clinical events from the morass of charge-related representations for the same clinical events. Enabling de-duplication.

- Would be nice if OMOP natively contained some kind of way of designating entries in the PROCEDURE_OCCURRENCE table as canonical or first-class versus entries that are financial echoes of the same clinical events

Future

- Change and Challenges:
    - Extract and open source the mashup of ‘abstractor’ and ‘OMOP’ as “OMOP Abstractor”. Coming Soon!
        - https://github.com/NUARIG/omop-abstractor
    - Integrate the OMOP vocabulary tables.
    - Need to incorporate into our OMOP instance legacy surgeries from Cerner Surginet.
    - Need to incorporate into our OMOP instance pathology procedures from a Cerner Co-Path instance to be loaded into our NMEDW.
    - Need to incorporate into our OMOP instance pathology procedures from a pending migration to Epic Beaker.
    - Improving our NLP algorithms. Not an NLP programmer. NLP pipeline has a RESTful interface that can delegate the generation of suggestions for a document and namespace to an endpoint and receive back suggestions via a endpoint. So better NLP can be used.

Thanks!