A Qualifier-Based Approach to Modeling Biolink Associations in Translator

A Qualifier-Based Association Proposal **1**

I. Background / Problem Statement 2

Summary of Challenges posed by a Predicate-Based Approach: 4

II. A Qualifier-Based Modeling Paradigm 4

III. Translator Modeling Principles (for a Qualifier-Based Approach) 6

IV. A Biolink Qualifier-Based Model (for Chemical-Gene Associations) 7

Overview 7

The Base ‘Association’ Schema 8

The Association Slot Hierarchy 9

Association Qualifier Slot Dictionary: 10

V. ‘Reading’ a Qualifier-Based Statement 12

Example: Composing Qualifier Semantics 12

Example: Ordering Node Qualifier Semantics 13

V. Statement Examples 14

"Fenofibrate binds to PPARA protein" (CTD) 14

"Cyclophosphamide affects the hydroxylation of CYP2B6" (CTD) 14

"Bisphenol A results in decreased degradation of ESR1 protein" (CTD) 15

"Bisphenol A is associated with decreased degradation of ESR1 protein" (CTD) 15

“Methionine deficiency results in increased expression of ADRB2” (CTD) 15

“Progesterone metabolites cause decreased methylation of APP promoter mutant forms” (CTD) 16

“Hexachlorobenzene analog causes increased methylation of CDKN2A enhancer alternative form” (CTD) 16

"The protein ser/thr kinase activator activity of Ras85D in the plasma membrane directly positively regulates MAPKKK activity of Raf in the cytoplasm within the EGFR signaling pathway" 17

VI. Application of the Qualifier Model in Translator 18

General Considerations 18

Summary: When to Use Qualifiers for Translator Knowledge Representation 19

VII. Considerations for Specific Translator Components 20

Biolink Model 20

LinkML Modeling Language 20

TRAPI 21

Knowledge Providers 21

Autonomous Relay Agents 22

ARS 22

User Interface 23
MetaKG endpoints 23

Appendix I: Chemical Gene-Association Schema and Enums 25
Association Type Schema 25
ChemicalGeneInteractionAssociation 25
ChemicalAffectsGeneAssociation 25
Definitions of Chem-Gene Association Statement Qualifier Slot 26
Predicates supporting Chem-Gene Associations 26
Enumerations supporting Chem-Gene Associations 27

I. Background / Problem Statement

The core currency of knowledge representation in Translator is the Biolink:Association object. An Association minimally includes a **subject-predicate-object (S-P-O) triple**, which expresses a **Statement** of knowledge about the domain. In Translator knowledge graphs (KGs), the subjects and objects of Associations are foundational domain concepts (e.g. genes, diseases, chemicals, phenotypes), whose IRIs (Internationalized Resource Identifiers) come from community standard ontologies (e.g. HGNC, MONDO, ChEBI, HPO). The predicate is a Biolink slot that represents the relationship between the subject and object concepts.

As Translator grows, we are seeing **increasingly diverse and complex types of Statements** that KPs want to be able to express using the Association model (**Figure 1**, left column).

Complex Statements Requiring Overloaded Predicates (in a Predicate-Based Approach)

Plain Language Statement	S-P-O Representation
"bisphenol A results in decreased degradation of ESR1 protein"	**bisphenol A** - *decreases degradation of* **ESR1 protein**
“methionine deficiency results in increased expression of ADRB2”	**methionine** - *deficiency of which causes increased expression of ->* **ADRB2**
"HER2 amplification is associated with sensitivity to neratinib in treatment of breast cancer"	**HER2** - *amplification of which is associated with sensitivity to ->* **neratinib**
"Daily PM2.5 exposure is associated with ED visits for asthma"	**PM2.5** - *daily exposure to which is associated with ED visits for ->* **asthma**
"Increased blood glucose is risk factor for diabetes in timeframe of two years"	**glucose** - *increased blood level of which is 2 year risk factor for ->* **diabetes**
"Alcohol dehydrogenase activity in the microbiome is correlated with insulin levels in the blood (of patients in cohort X)"	**ADH** - *activity of which in microbiome is correlated with blood level of ->* **insulin**

"Digenic germline mutations in Gene X and Gene Y cause Disease Z can cause Disease Z"	***Gene X*** *- is causal germline mutation partially giving rise to -* ***Disease Z*** (lossy)
"The protein ser/thr kinase activator activity of Ras85D in the plasma membrane directly positively regulates MAPKKK activity of Raf in the cytoplasm within the EGFR signaling pathway"	**Ras85D** *- has activity X in compartment Y that directly positively regulates activity A in compartment B of ->* **Raf**

Figure 1*: Examples of the kinds of overloaded predicates that <u>would be needed</u> to represent complex/nuanced statements in a strict (unqualified) predicate-based approach. The right-hand column shows a S-P-O representation - where without qualifiers, and considering the need to keep node concepts simple - semantics would have to be captured in overloaded predicates.*

These Statements may have subject or object concepts that are more nuanced and **lack standard community identifiers** (e.g. *'Methionine Deficiency'*, '*HER2 amplification'*), or assert **complex relationships** between these concept that **rely on additional information or context** (e.g. '*increases transport of (in hepatocytes)*', *'associated with susceptibility to (in covid patients)*'. Such Statements **strain our ability to fit into a simple S-P-O triple.**

To date, in the absence of a mechanism to load additional semantics into subject/object concepts, we have resorted to **pushing them into predicates** (**Figure 1**, right column) - and begun to see a **proliferation of overloaded predicates** bloat the Biolink predicate hierarchy to >300 terms. This is most evident in the 'affects' hierarchy created to support CTD and Drug-Target interaction data sources like DrugBank, DGIdb, and ChEMBL. **Figure 2** shows a small subset of these predicates, which number over 60 in total.

Proliferation of Predicates (needed to represent Chemical effects on Gene/Products)

affects ❖ affects abundance of ● affects synthesis of ○ increases synthesis of ○ decreases synthesis of ● affects expression of ○ decreases expression of ○ increases expression of ● affects degradation of ○ decreases degradation of ○ increases degradation of ● decreases abundance of (mixin) ○ decreases synthesis of ○ decreases expression of ○ increases degradation of ● increases abundance of (mixin) ○ increases synthesis of ○ increases expression of ○ decreases degradation of	◆ affects activity of ● decreases activity of ○ inactivates ○ neutralizes ○ inhibits ■ antagonist_of ■ blocks molecular channel ■ antibody inhibitor of ■ inverse agonist of ■ negative allosteric modulator of ● increases activity of ○ agonist of ○ opens molecular channel ○ positive allosteric modulator of ○ potentiates

***Figure 2**: Examples of proliferating / overloaded predicates currently defined in the 'affects' hierarchy of Biolink.*

But this is just the tip of the iceberg. Reliance on a strict predicate-based approach for emerging use cases would **lead to predicates numbering in the thousands**, many of which would be **overloaded with semantics** required to express complex and nuanced Statements (Figure 2).

Summary of Challenges posed by a Predicate-Based Approach:

- Predicates required to capture semantics of more complex statements can be awkward/overloaded
- Proliferation of predicates poses challenges to maintaining a principled and consistent hierarchy in the Biolink Model
- Managing multiple inheritance hierarchies along different dimensions of classification would inevitably result in the need to represent many of the predicates as slot mixins, further complicating the modeling
- The size and complexity of the predicate hierarchy can make it difficult for KPs to find the most appropriate predicate they need when creating data, introducing the potential for inaccurate, lossy, inconsistent, or duplicated data representations.
- It will be hard for users to find predicates to query the data (can lead to misleading / incomplete query results)

II. A Qualifier-Based Modeling Paradigm

To eliminate over-reliance on predicates to accommodate more complex and nuanced Statements, we are pursuing a qualifier-based modeling paradigm. This approach uses Associations slots called qualifiers that let us layer additional semantics onto a simpler core triple statement.

At the highest level we distinguish two kinds of qualifiers that contribute to an Association Statement:

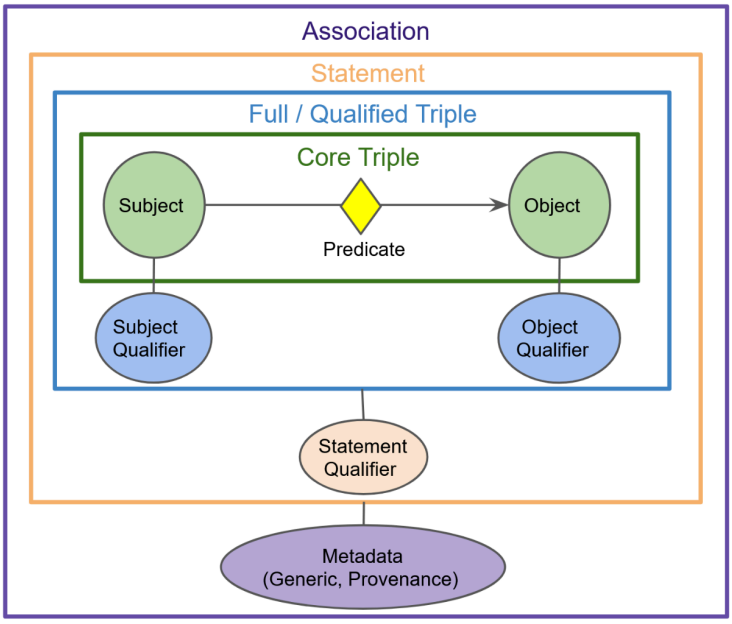
> (1) **node qualifiers** (aka **subject / object qualifiers**) extend or refine the meaning of an Association subject or object concept;
>
> (2) **statement qualifiers** refine or extend the meaning of the core S-P-O triple as a whole.

Together, the subject, predicate, object, and optional qualifier(s) comprise the full semantics of the statement that an Association puts forth as true (i.e. its 'S-P-O-Q' semantics).

Association objects may also include slots to hold **Metadata** about this core statement - primarily information about the **provenance** and **evidence** supporting it - but unlike qualifiers, this metadata does not contribute to the meaning of the core Statement itself.

Using these qualifier and metadata elements together, we can build Associations with many possible 'layers' of complexity (**Figure 3**).

Onion Diagram: Conceptual Model of a Qualifier-Based Biolink Association

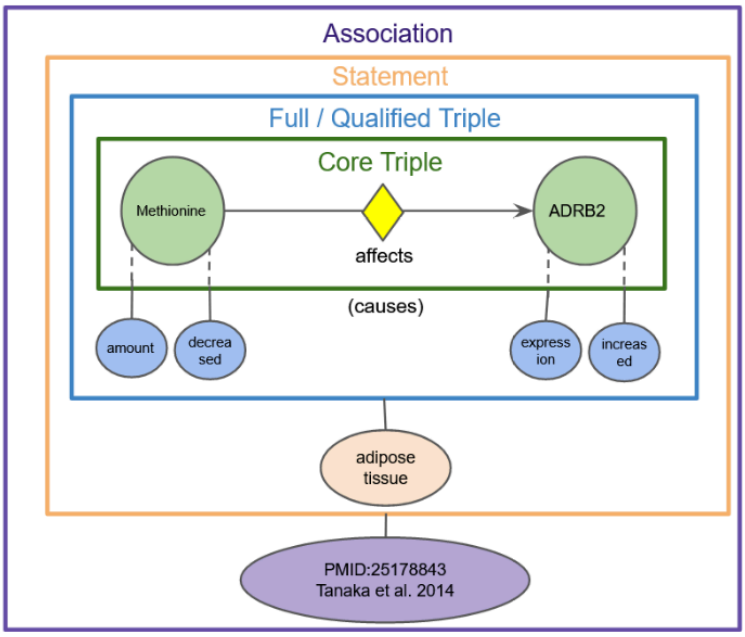


Four 'Layers' of an Association	
Core Triple	Just the claim expressed in the subject and object nodes plus the predicate connecting them. The minimum required to produce a valid Association.
Full Triple	The claim made when considering any qualifiers of subject or object semantics (still a 'triple' because it is comprised of three concepts (a subject, predicate and object) - but the subject or object concepts might be 'composed' from >1 term)
Statement	The full claim being expressed in the association, by the qualified triple plus any statement qualifiers
Association	The Statement plus supporting Metadata / Provenance information

***Figure 3**: 'Onion-layer' diagram of the Association conceptual model*

Notably, some Associations may simply consist of an S-P-O triple. Others may represent more complex statements that employ multiple qualifiers, and may be supported by rich evidence and provenance metadata. **Figure 4** provides a layered view of an example complex Association Statement which asserts that

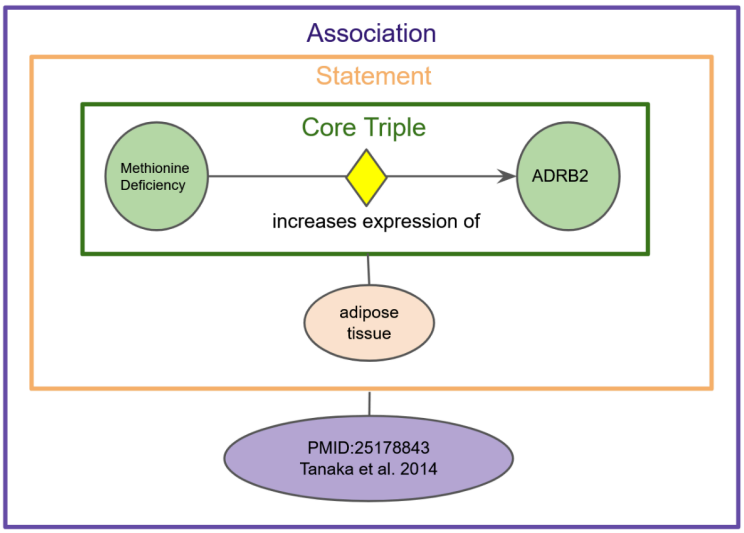
"Methionine deficiency results in increased expression of ADRB2, in adipose tissue".



Plain-language semantics of each 'layer' of the Association	
Core Triple	*"Methionine affects ADRB2"*
Full Triple	*"Methionine deficiency affects (causes) increased expression of ADRB2"*
Statement	*"Methionine deficiency affects (causes) increased expression of ADRB2 in adipose tissue"*
Association (Statement + Metadata)	*"Methionine deficiency affects (causes) increased expression of ADRB2 in adipose tissue" . . . and we know this from PMID:25178843 Tanaka et al. 2014*

***Figure 4**:Representation of an Association expressing the statement "Methionine deficiency results in increased expression of ADRB2, in adipose tissue", using a decomposed qualifier-based approach. Note that the blue subject and object qualifiers refine/extend the meaning of the green subject and object concepts in the context of a given Association. Accordingly, the model will implement slots for subject and object qualifiers as Edge properties , not Node properties.*

Contrast this qualifier-based representation with how we might capture this knowledge under the presently supported predicate-based approach (with no subject/object qualifiers):



Plain-language semantics of each 'layer' of the Association	
Core Triple	*"Methionine Deficiency increases expression ADRB2"*
Statement	*"Methionine Deficiency increases expression ADRB2 in adipose tissue"*
Association (Statement + Metadata)	*"Methionine Deficiency increases expression ADRB2 in adipose tissue" . . . and we know this from PMID:25178843 Tanaka et al. 2014*

***Figure 5**:Representation of an Association expressing the statement "Methionine deficiency results in increased expression of ADRB2, in adipose tissue", using the currently supported predicate-based model.*

Here we must create a phenotype term to represent the concept of 'methionine deficiency, and capture gene aspect (expression) and direction (increased) semantics in the overloaded predicate (increases expression of).

- This representation poses two key challenges:
    - (1) a class representing 'decreased amount of Methionine' as a phenotype is not present in HPO at the moment (and classes for deficiency of every known metabolite is likely not a path HPO wants to go down); and
    - (2) the KG does not have a direct link between the core concepts of Methionine and ABRD2 - which we would like to have in the graph.
        - We could achieve this by creating a predicate that directly links Methionine and ABRD2, such as '*decreased levels of which increases expression of'* - but this is even more overloaded, and would lead to potential for significant explosion of predicates.

III. Translator Modeling Principles (for a Qualifier-Based Approach)

A comprehensive assessment of diverse use cases helped us define a coherent set of Translator modeling principles that promote clear and consistent knowledge representations, in a way that supports graph-based operations/reasoning Translator needs to perform (a more detailed description of these principles can be found here). These modeling principles/desiderata have guided modeling choices throughout the qualifier-based approach we are defining in Biolink.

1. **Nodes should represent core domain concepts:** If possible, IRIs for KG nodes should represent fundamental domain concepts (genes, chemicals, phenotypes, diseases, etc.) This facilitates

connections between primary entities of interest with fewer hops, and avoids the need to create/maintain/resolve new IRIs.

   a. **Corollary**: **Limit dependencies on term creation by external ontologies:** We don't want a scenario where we are waiting for external, unpredictable ontologies to add terms we need, e.g., addition of terms like ''severe bleeding" to HP, "early onset Alzheimers" to MONDO, or ''exposure to PM2.5" to ECTO.

2. **Use qualifiers to compose full node semantics:** When an identifier/IRI does not exist for a node concept in a standard, Translator-approved ontology, use qualifiers to post-compose their meaning. This is preferred over minting new ontology terms at a more granular level than is practical, or using structured data objects as Statement subject/objects.
3. **The 'core triple' should remain true if qualifiers are ignored:** When using qualifiers, ensure that the core SPO triple remains true when qualifiers are ignored. (However, note that there may be one predicate used for the core triple and a different predicate for the qualified assertion.) If certain necessary qualifiers may violate this rule (e.g., 'negated'), these should be flagged and NEVER ignored.
4. **Control predicate proliferation:** When deciding where to place Statement semantics, choose modeling approaches that avoid a potential for an explosion of predicates. Pushing semantics into qualifiers is one way to achieve this.
5. **Represent information consistently:** Where possible, a given type of semantics (e.g., gene aspect, direction of effect) should be represented using the same pattern across Statement types and components. This will facilitate clear and consistent creation of data by KPs, and simplify query construction and answering.
6. **Separate the Tasks of Knowledge Ingest from Inference:**
   a. Ingest of Assertions made by a source: KPs capture what sources assert to be true (with as little interpretation/inference as possible)
   b. Inference of Predictions based on such assertions: leave the job of interpretation/inference to CM prediction tools - which can make stronger 'treats' claims that are very clearly flagged/softened as 'predictions

IV. A Biolink Qualifier-Based Model (for Chemical-Gene Associations)

Overview

Guided by these principles, and a comprehensive collection of Association examples, we implemented a qualifier-based Association model in Biolink that will best support Translator data and use cases (see PR#1028).

The initial model was based primarily on use cases from the domain of **Chemical-Gene Associations.** This area was chosen because it requires representation of relatively complex node concepts and statement semantics, and would surface diverse requirements and considerations. We were careful to evaluate modeling patterns

defined for these Associations for their support of other domains and statement types so as to ensure generalizability of the approach.

The model we established supports representation of simpler Statements using the existing SPO pattern already in use across Translator, e.g., *"Methionine affects ADRB2".*

```
subject: Methionine
predicate: affects
object: ADRB2
```

But it also provides qualifiers to layer additional semantics onto a core SPO triple to express more complex and nuanced statements, e.g., *"Methionine deficiency results in increased expression of ADRB2 in adipose tissue"*

```
subject: Methionine
subject_aspect: abundance
subject_direction: decreased
predicate: affects
qualified_predicate: causes
object: ADRB2
object_aspect: expression
object_direction: increased

anatomical_context_qualifier: adipose tissue
```

Box 1: Comparing Predicate- vs Qualifier- Based Representations

Compare with the current predicate-based approach, where we would have to represent the complex Statement above in one of two ways:

```
# stuff all semantics into an unwieldy predicate
 subject: Methionine
 predicate: deficiency_of_which_increases_expression_of
 object: ADRB2
 anatomical_context_qualifier: adipose tissue
```

or

```
# capture some semantics in the subject by casting as a phenotype
 subject: Methionine Deficiency (phenotype)
 predicate: results_in_increased_expression_of
 object: ADRB2
 anatomical_context_qualifier: adipose tissue
```

In the top approach, we are forced to use a heavily overloaded predicate and deal with the unwieldy predicate proliferation that would result.

In the bottom approach, we have a slightly less overloaded predicate, but we lose the direct association between the chemical and gene that the MolePro KP desired.

Below we describe additions and changes to the Biolink Model required to support this type of qualifier-based representational approach across Translator use cases.

The Base 'Association' Schema

A 'dual predicate', qualifier-based model for representing statement semantics in Biolink Associations (only Association slots holding statement semantics are shown). Slots used to define the subject entity/concept are in **blue**, the object concept in **purple**, the relationship holding between them in **green**, and additional detail about this relationship in **red**.

```
subject: Named Thing  [1..1]
subject_part:  enum [0..1]
subject_form_or_variant:  enum [0..1]
subject_derivative:  enum [0..1]
subject_aspect:  enum [0..1]
subject_direction: enum [0..1]
subject_context:  enum [0..1]

predicate: PredicateType [1..1]
qualified_predicate: PredicateType [0..1]

object: Named Thing  [1..1]
object_part:  enum [0..1]
object_form_or_variant:  enum [0..1]
object_derivative:  enum [0..1]
object_aspect:  enum [0..1]
object_direction: enum [0..1]
object_context:  enum [0..1]

statement_qualifier:   enum [0..m]
statement_quantifier: decimal [0..1]

negated: boolean [0..1]

epc properties . . .
```

Notes:

- Our use case and requirements analysis for chemical-gene associations led to the identification of these six node qualifier types (part, form, derivative, aspect, direction, and context to represent the kinds of complex concepts that are subjects and objects of Translator Statements.
- These qualifier slots are named/defined to indicate the specific role the qualifier plays, and how it combines with other qualifiers, in composing the semantics of the full subject or object concept.
- These subject/object qualifier slots are needed only when a standard, Translator-approved identifier/IRI for a subject/object concept does not exist to enable composition of node concept semantics from >1 term.
- Regarding the statement-level qualifier and quantifier slots - more specific qualifiers than the generic slots shown here will be defined as relevant for particular Association types.

    - For example, a Statement with the core triple 'Variant X' - causes - 'Phenotype Y' might include a `developmental_stage_qualifier` slot that describes the life stage during which the variant-phenotype relationship occurs (e.g., 'juvenile stage').

- Regarding enums bound to qualifier slots, specific enumerations will be defined for specific Association subtypes to hold values relevant for the type of Statement expressed.
    - For example, in a `ChemicalAffectsGeneAssociation`, the enum bound to the `object_part` qualifier will contain terms representing parts of genes or gene products.

Full examples of schema for a `ChemicalGeneInteractionAssociation` and `ChemicalAffectsGeneAssociation` are shown in Appendix I, which include statement qualifiers and enum bindings specific to these kinds of statements.

The Association Slot Hierarchy

```
association slot
    association qualifier
        subject or object qualifier
            part_qualifier
                subject_part_qualifier
                object_part_qualifier
            form_or_variant_qualifier
                subject_form_or_variant_qualifier
                object_form_or_variant_qualifier
            derivative_qualifier
                subject_derivative_qualifier
                object_derivative_qualifier
            aspect_qualifier
                subject_aspect_qualifier
                object_aspect_qualifier
            direction_qualifier
                subject_direction_qualifier
                object_direction_qualifier
            context_qualifier
                subject_context_qualifier
                object_context_qualifier

        statement qualifier
            affect_mechanism_qualifier
            interaction_mechanism_qualifier
            anatomical_context_qualifier
            species_context_qualifier
```

Association Qualifier Slot Definitions:

1. **form_or_variant_qualifier**: A qualifier that composes with a core subject/object concept to define a specific type, variant, alternative version of this concept.

a. The composed concept remains a subtype or instance of the core concept, and the high level category of the concept does not change.
b. Examples:
    i. the qualifier ‘mutation’ combines with the core concept ‘Gene X’ to express the compose concept ‘Mutated forms of Gene X’.
    ii. the qualifier ‘Late Stage’ combines with a core concept of ‘Disease X’ to express the more specific concept ‘Late Stage forms of Disease X’
    iii. the qualifier ‘Severe’ combines with a core concept of ‘Bleeding’ to express the more specific concept ‘Severe forms of Bleeding’
    iv. the qualifier ‘Liver’ combines with a core concept of ‘Surgical Transplant’ to express the composed concept ‘Liver Transplant’
    v. the qualifier ‘Recombinant’ combines with a core concept of ‘FLT1 Gene’ to express the composed concept ‘Recombinant forms of the FLT1 gene’
    vi. the qualifier ‘chemical analog’ combines with a core concept of ‘Ditiocarb’ to express the composed concept ‘analog forms of Ditiocarb’
c. Notes concerning use of ‘form’ and ‘aspect’ qualifiers:
    i. The purpose of the aspect slot is to indicate what aspect is being affected in an 'affects' association. The purpose of the form slot is to describe the form of the concept bearing the aspect that is affected.
    ii. There may be concepts that can fit into either slot, depending on the purpose of the concept in a Statement. For example,consider the concept of being ‘methylated’ in the two scenarios below, and where it gets captured in the Statement model:
        1. There may be cases where a Statement asserts, e.g., that some Chemical X increases degradation of methylated forms of a Protein Y. Here, we would use the *form_or_variant_qualifier* slot to hold the value 'methylated form' (indicating that it is methylated forms of Protein Y specifically whose degradation is being affected).

            subject: Chemical X
            predicate: affects
            qualified_predicate: causes
            object: Protein Y
            object form or variant: methylated form
            object_aspect: degradation
            object_direction: increased

        2. There may be other cases where a Statement asserts, e. g., that some Chemical X increases methylation of mutant forms of Protein Y. Here, we use the *aspect_qualifier* slot to hold the value 'methylation' (indicating that it is the methylation of mutant forms of Protein Y that is being affected).

            subject: Chemical X
            predicate: affects
            qualified_predicate: causes
            object: Protein Y

object_form_or_variant: mutant
object_aspect: methylation
object_direction: increased

2. **derivative_qualifier**: A qualifier that composes with a core subject/object concept to describe something that is derived from the core concept.
    a. Examples:
        i. the qualifier 'metabolite' combines with a core concept 'Chemical X' to express the composed concept 'a (derivative) metabolite of Chemical X'.
3. **part_qualifier**: defines a specific part/component of the core concept
    a. This qualifier is used in cases there the specific part has no IRI we can use to directly represent it
    b. Examples:
        i. the qualifier 'polyA tail' combines with a core concept of 'Gene X' to express the composed concept 'polyA tail of Gene X'
        ii. the qualifier 'finger numbness' combines with a core concept 'Diabetes' to express the composed concept 'finger numbness as a part of Diabetes'.
4. **aspect_qualifier**: Composes with the core concept to describe new concepts of a different ontological type - such as a process in which the core concept participates, a function or role held by the core concept, or a characteristic/quality that inheres in the core concept.
    a. The aspect qualifier is loosely defined, as a catch all for qualifiers that don't fit into the other node qualifier categories
    b. Examples:
        i. the qualifier 'expression' combines with a core concept of 'Gene X' to express the composed concept 'expression of Gene X' (Gene → Biological Process)
        ii. the qualifier 'exposure' combines with a core concept of 'Chemical X' to express the composed concept 'exposure to Chemical X' (Chemical → Exposure Process)
        iii. the qualifier 'Activity' combines with a core concept of 'PPARG' to express the concept 'activity of PPARG' (Gene → function/activity)
        iv. the qualifier 'Emergency Department Visit' combines with a core concept of 'Disease X' to express the concept 'Emergency Department visits for Disease X' (Disease → Clinical Event)
        v. the qualifier 'Infection' combines with a core concept of 'Giardia' to express the concept 'Infection with Giardia' (Taxon → Biological / Pathological Process)
        vi. the qualifier 'Severity' combines with a core concept of 'DILI' to express the concept 'the severity level of DILI' (Disease → (intrinsic) Characteristic/Quality)
        vii. the qualifier 'Abundance' combines with a core concept of 'BRCA2' to express the concept 'abundance of BRCA2' (Gene → (extrinsic) characteristic/quality)

5. **context_qualifier**: Restricts the setting/context/location where the core concept (or qualified core concept) resides or occurs
    a. Examples:
        i. the qualifier 'hippocampus' combines with a core concept of 'neuron' to express the composed concept 'neuron in the hippocampus'
        ii. The species_context_qualifier applies taxonomic context, e.g. species-specific molecular activity (as specified by a GO term?)
6. **direction_qualifier**: Composes with the core concept (+ aspect if provided) to describe a change in its direction or degree
    a. Examples:
        i. the qualifier 'increased' combines with a core concept of 'Gene X' and an aspect of 'expression' to express the composed concept 'increased expression of Gene X'
        ii. the qualifier 'decreased' combines with a core concept of 'Protein X' and an aspect of 'abundance' to express the composed concept 'decreased abundance of Protein X'

7. **qualified_predicate**: holds a relationship to be used instead of that expressed by the primary predicate, in a 'full statement' reading of the association, where qualifier-based semantics are included. This is necessary only in cases where the primary predicate does not work in a full statement reading.
    a. e.g. to express the statement that *"Chemical X causes increased expression of Gene Y"*, the core triple is read using the fields **subject**:ChemX, **predicate**:affects, **object**:GeneY . . . and the full statement is read using the fields **subject**:ChemX, **qualified_predicate**:causes, **object**:GeneY, **object_aspect**: expression, **object_direction:**increased. The predicate 'affects' is needed for the core triple reading, but doesnt make sense in the full statement reading (because "Chemical X *affects* increased expression of Gene Y'' is not what we mean to say here . . . it *causes* increased expression of Gene Y)
8. **statement_qualifier:** provides additional information that extends or refines the meaning of the statement triple as a whole
    a. applies to the asserted relationship between subject and object concepts, not either concept on its own.
    b. these can add additional details/precision (e.g. penetrance level of a disease-phenotype association), or describe a context in which statement is true (e.g. a particular cohort or population, a type of cell or tissue, an environmental context in which a genotype-phenotype association applies)
    c. Implementation Note: data creators should not use this slot directly - rather, use one of the more specific statement qualifier subproperties that specify the semantics of the qualifying information (e.g. 'anatomical context qualifier')
9. **statement_quantifier:** a type of statement_qualifier holding a numerical value that quantifies some aspect of the knowledge asserted in the Statement

    a. e.g. a ‘frequency quantifier’ might carry a value representing the specific frequency (0.78) at which a phenotype is observed in a disease

10. **negated**: a boolean slot that can be set to ‘true’ to indicate that the Statement expressed by SPOQ slots is explicitly asserted to be false.

V. ‘Reading’ a Qualifier-Based Statement

Here we provide below a guide for how to ‘read’ the meaning of an Association Statement by composing the semantics encoded in its SPOQ slots. Note that many such Associations will consist only of a **subject-predicate-object** triple - in which case a reading of the statement expressed is straight-forward. When qualifiers are included, this introduces a **second level of meaning** that complements the statement made by the core SPO triple. This idea that there can be **two related statements** at two levels of granularity encoded in a single Association is a critical idea behind the qualifier-based approach:

- The first level is the **Core Triple Reading** of the Association, which considers only the `subject`, `predicate`, and `object` slots. This reading must be clear and true.
- The second level is the **Full Statement Reading** of the Association, which considers the semantics added by `qualifier` slots (subject/object qualifiers, and statement qualifiers). This reading must also be clear and true as well (which often requires use of a `qualified_predicate`).

Note that the **Full Statement Reading** provides additional detail or context for the claim made in the core triple. It should not represent a completely independent or unrelated statement.

Example: Composing Qualifier Semantics

The example below shows a complex Association Statement that utilizes each of these types of node qualifier slots to represent a record from the CTD database that asserts “*A Hexachlorobenzene metabolite increases methylation of a mutant form of the CDKN2A promoter in the nucleus of HeLa cells”*.

```
subject: Hexachlorobenzene
subject_derivative: metabolite
predicate: affects
qualified_predicate: causes
object: CDKN2A
object_part: promoter
object_form_or_variant: mutant form
object_aspect:  methylation
object_direction: increased
object_context: nucleus
experimental_context_qualifier: HeLa cells
```

To extract the assertions encoded in this structure:

1. Read the Core S-P-O Triple: This first layer of meaning is comprised only of the 'core concepts' captured in the `subject` and `object` slots, and the `predicate` which expresses a true, high-level relationship between them.
    a. *"Hexachlorobenzene affects CDKN2A"*
2. Compose the Full Subject and Object Concepts: When there are subject/object qualifiers in the Association, combine the core concepts with their qualifiers to compose their full meaning.
    a. Full/composed subject concept = *"Hexachlorobenzene metabolite"*
    b. Full/composed object concept = *"Increased Methylation of a mutant form of the CDKN2A promoter in the nucleus"*
3. Read the Full Statement: Read the relationship between the composed subject and object concepts, along with additional info/context provided by any statement-level qualifiers or quantifiers. When provided, the `qualified_predicate` should be used in this full statement reading.
    a. *"Hexachlorobenzene metabolite causes Increased methylation of a mutant form of the CDKN2A promoter in the nucleus of HeLa cells"*

Example: Ordered Application of Qualifier Semantics to Compose an Object Node Concept

In cases where there are multiple qualifiers on a single subject or object concept, we can provide conventions for the order in which qualifier semantics are layered onto the core concept. This can facilitate the semantic interpretation of a statement, and remove ambiguity where different qualifier ordering results in subtly different statement meaning (see Box 2).

1. **object**: the simple core concept 'X'
    - e.g. the CDKN2A gene
2. **object_part**: a particular part of 'X'
    - e.g. the promoter of the CDKN2A gene
3. **object_form_or_variant**: a particular form of this part of 'X'
    - e.g. a modified form of the promoter of the CDKN2A gene
4. **object_aspect**: a particular aspect of this form of this part of 'X' (e.g. its abundance, its transport)
    - e.g. methylation of a modified form of the promoter of the CDKN2A gene
5. **object_direction**: an increase or decrease in this particular aspect of 'X'
    - e.g. increased methylation of a modified form of the CDKN2A promoter
6. **object_context**: an inc/dec of a particular aspect of this form of this part of 'X', in a particular context
    - e.g. increased methylation of a modified form of the CDKN2A promoter in the nucleus

Importantly, each new qualifier layer **applies to the concept composed by all qualifiers preceding it**. So in the example above, the 'methylation' aspect applies to the concept composed by the three qualifiers preceding it ('a mutant form of the promoter of the CDKN2A gene'), not directly to the core concept ('CDKN2A').

BOX 2: The order in which qualifier semantics are composed matters.

In cases where multiple qualifiers are attached to a single subject or object concept, the order in which qualifiers are added can result in subtly different composed concept semantics. For example, consider the order in which the two object qualifiers might be applied in the example below:

object: CDKN2A
object_part: promoter
object_form_or_variant: mutant form

Adding the `object_part` semantic *before* the `object_form_or_variant` semantic gives the expression:

`[ [ [CDKN2A] promoter]  mutant form ]`

. . . composing inside to out, this gives the concept '*mutant forms of the CDKN2A promoter*', which suggests the promoter itself is mutated in some way.

Adding the `object_part` semantic *after* the `object_form_or_variant` semantic gives the expression:

`[ [ [CDKN2A] mutant form]  promoter ]`

. . . again composing inside to out, this gives the concept *'the promoter of mutant forms of the CDKN2A gene*'. In this case, the mutation in CDKN2A can reside anywhere within the gene - not necessarily in the promoter.

The difference in meaning between these interpretations is subtle, but may be important to preserve for certain use cases. Therefore, we should consider how to specify the order in which qualifier semantics are composed.

V. Statement Examples

Structured representations of several example Chemical-Gene Association Statements, comparing the proposed qualifier-based approach with a predicate-based approach.

1. *"Fenofibrate binds to PPARA protein" (CTD)*

   A simple Chemical interacts with gene Statement (no qualifiers needed)

Qualifier-Based Approach	Predicate-Based Approach
subject: Fenofibrate **predicate:** physically interacts with **object:** PPARA	**subject:** Fenofibrate **predicate:** physically interacts with **object:** PPARA

2. *"Cyclophosphamide affects the hydroxylation of CYP2B6" (CTD)*

   A simple chemical affects gene (aspect) Statement - no direction to the effect.

Qualifier-Based Approach	Predicate-Based Approach
subject: Cyclophosphamide **predicate**: affects **object**: CYP2B6 **object_aspect**: hydroxylation	**subject**: Cyclophosphamide **predicate**: affects hydroxylation of **object**: CYP2B6

3. *"Bisphenol A results in decreased degradation of ESR1 protein" (CTD)*

   A Statement where the effect has a direction (decreased)

Qualifier-Based Approach	Predicate-Based Approach
subject: Bisphenol A **predicate**: affects **qualified_predicate**: causes **object**: ESR1 **object_aspect**: degradation **object_direction:** decreased	**subject**: Bisphenol A **predicate**: decreases degradation of **object**: ESR1

4. *"Bisphenol A is associated with decreased degradation of ESR1 protein" (CTD)*

   A (hypothetical) chemical *associated_with* gene (aspect) Statement with same S/O concepts as above:

Qualifier-Based Approach	Predicate-Based Approach
subject: Bisphenol A **predicate**: associated_with **object**: ESR1 **object_aspect**: degradation **object_direction:** decreased	**subject**: Bisphenol A **predicate**: associated with decreased degradation of **object**: ESR1

5. *“Methionine deficiency results in increased expression of ADRB2” (CTD)*

   A more complex statement - shows how aspect and direction semantics represented for S and O nodes

Qualifier-Based Approach	Predicate-Based Approach
subject: Methionine **subject_aspect**: abundance	

subject_direction: decreased **predicate**: affects **qualified_predicate:** causes **object**: ADRB2 **object_aspect:** expression **object_direction**: increased	**subject**: Methionine **predicate**: decreased amount increases expression of **object**: ADRB2

6. *"Progesterone metabolites cause decreased methylation of APP promoter mutant forms" (CTD)*

   A more complex example where a metabolite of the specified chemical is the effector of a heavily qualified Statement object.

Qualifier-Based Approach	Predicate-Based Approach
subject: Progesterone **subject_derivative:** metabolite **predicate**: affects **qualified_predicate:** causes **object**: APP **object_part:** promoter **object_form_or_variant:** mutant **object_aspect**: methylation **object_direction**: decreased	**subject**: Progesterone **predicate**: metabolites decrease methylation of variant promoter of **object**: APP

7. *"Hexachlorobenzene analog causes increased methylation of CDKN2A enhancer alternative form" (CTD)*

   Another complex example where an analog of a specified chemical is the effector of a heavily qualified Statement object.

Qualifier-Based Approach	Predicate-Based Approach
subject: Hexachlorobenzene **subject_form_or_variant:** analog **predicate**: related to **qualified_predicate:** causes **object**: CDKN2A **object_part:** enhancer **object_form_or_variant**: modified form **object_aspect**: methylation **object_direction**: increased	**subject**: Hexachlorobenzene **predicate**: analogs increase methylation of modified enhancer of **object**: CDKN2A

8. *Fenofibrate is an agonist of PPARA protein" * (DrugBank)*
   Chemical increases gene activity, via a specific control mechanism (agonism)

Qualifier-Based Approach	Predicate-Based Approach
subject: Fenofibrate **predicate**: affects **qualified_predicate**: causes **object**: PPARA protein **object_aspect**: activity **object_direction**: increased **mechanism_qualifier**: agonism	**subject**: Fenofibrate **predicate**: is agonist of **object**: PPARA protein

* Note that agonism as a molecular control mechanism implies the existence of a physical interaction between a chemical and gene product - so by convention, we would require that *a second Association is created* to represent this implied knowledge (here, that **Fenofibrate** – *physically_interacts_with* → **PPARA protein**)

9. *"The protein ser/thr kinase activator activity of Ras85D in the plasma membrane directly positively regulates MAPKKK activity of Raf in the cytoplasm within the EGFR signaling pathway"*

   A very complicated GO-CAM example . . .

Qualifier-Based Approach	Predicate-Based Approach
subject: Dmel Ras85D **subject_aspect**: protein ser/thr kinase activator activity **subject_specialization**: plasma membrane **predicate**: regulates **qualified_predicate**: causes **object**: Dmel Raf **object_aspect**: MAPKKK activity **object_specialization**: cytoplasm **object_direction**: increased **pathway_context_qualifier**: EGFR pathway	**subject**: Dmel Ras85D **predicate**: . . . (don't even know how to start here) **object**: Dmel Raf

A library of additional structured Statement examples can be found in the file here, along with a set of example TRAPI queries posed against these data.

A TRAPI API serving CTD data using this qualifier-based representation can be found here (LINK)
Code used to generate this dasta/API is here.

VI. Application of the Qualifier Model in Translator

General Considerations

At the end of the day, qualifiers are simply a means to represent the semantics of Association Statements in a more expressive, consistent, flexible, and computable way. This approach will require Translator developers who are accustomed to a simple SPO representation *to shift their thinking* in allowing qualifiers to carry certain types of Statement semantics. And learn to compose the meaning expressed in a Statement from a potentially large number of terms. That said, it may not be as hard as it sounds, if we consider the following:

1. **The Biolink Team will provide clear and comprehensive guidance on how to use qualifiers**
    a. Primers and tutorials in various formats about the general conceptual model and approach
    b. Mappings between predicate and qualifier-based representations
    c. A corpus of diverse data examples
    d. Computable schema for Association subtypes that provide specific guidance for how to apply qualifiers to represent a particular type of Statement
    e. Validation tooling to test compliance with formal model semantics, and informal data representation conventions

2. **Changes to existing Translator data will be limited***
    a. The majority of existing Statement representations in Translator data will remain unchanged - as they are simple enough to be accommodated by a SPO triple with a simple predicate (e.g. 'X treats Y', 'A affects B', 'N expressed in M')
    b. Another large proportion of Statements may require only minor changes - such as removing direction from the predicate and putting it into a node qualifier. e.g.:
        i. 'Gene X - negatively_regulates - Gene Y' —>
        'Gene X - regulates - Gene Y (object_direction: negative)'
    c. Only a small percentage of Statements will require more significant restructuring - such as those involving effects of chemicals on aspects of gene function, where we will be deprecating overloaded predicate created to support this use case

    * An exception may be Clinical Feature Variable Associations, where qualifiers will be used to increase the precision and information content of existing Statements across the board.

3. **The need for more complex qualifier-based representations primarily applies to new and emerging use cases:**
    a. Qualifier-based modeling is primarily needed for new use cases that the current Biolink Model does not support.
    b. The Biolink Team will work with KPs providing these use cases to define qualifier-based models for their data
    c. Over time, we will document and codify modeling patterns and conventions such that data providers are able to draft and submit compliant models on their own.

4. **We do have the option of adopting qualifiers in a more limited way:**
    a. The modeling team proposals here explore what an strict qualifier-based approach would look like - as a counterpoint to the present strict predicate-based approach.
    b. We may decide that a more middle ground approach is best - where certain semantics can move from qualifiers back into predicates, in a principled way.
        i. e.g. maybe we decide that direction of biological effect is important to represent in predicates so we can more directly navigate/reason over direction semantics in a KG edge. We could move these semantics into predicates, and keep other semantic types (aspects, parts, forms, context, etc.) in qualifiers.
    c. Ultimately we need to strike a pragmatic balance between maximal expressivity/flexibility afforded by qualifiers, with operational requirements of Translator use cases that may benefit from certain types of semantics living in predicate.

5. **Some details accommodated by qualifiers may be deemed out of scope**
    a. We may decide that the level of detail some qualifier patterns are defined to capture are not necessary to represent (at least not in a structured/computable way) . . . in which case our models may be simplified.
    b. We should consider this as we assess use cases form KPs, and develop models to support them

When to Use Qualifiers for Translator Knowledge Representation

When it comes to representing knowledge in Translator, there are three main areas to think about when applying qualifier modeling patterns to capture or query Statement semantics.

1. **Adding Nuance to an S/O Node Concept** : Node Qualifiers can provide a way to compose representations of more complex concepts that lack standard community identifiers, as subject/object nodes in Associations.
    a. relevant qualifier slots: part, form_or_variant, derivative, context
    b. example use cases:
        i. Representing clinical feature variables
        ii. Representing specific molecular entities

    Note here that the qualifier pattern allows us to represent these additional, important nuances in a way that maintains direct linkages between core domain concepts in the primary graph - which is important for display, navigation, and reasoning over the data. The additional detail is there to use as needed, but 'stays out of the way' the connections in the primary graph.

2. **Refactoring an Overloaded Predicate**: Other qualifiers can provide a place to move semantics out of overloaded predicates, which will no longer be supported by Biolink Model
    a. relevant qualifier slots: aspect, direction (these are things that at present are captured mainly in predicates)
    b. example use case: Specifying a precise effect of a chemical on gene function

3. **Extending or Constraining Semantics of an SPO triple as a whole:** Statement Qualifiers . . .
    a. **relevant qualifier slots:** statement qualifiers defined as needed for specific Statement types.
    b. **example use case**: Specifying a particular developmental stage during which a variant-phenotype association holds.

These three areas of utility will be important for all Translator components and stakeholders to consider, including **KPs** when creating data, **ARAs** when implementing reasoning tools that that traverse and collect data from KG edges, the **UI** in designing workflows and templates guiding query creation, and the **TRAPI** standard in defining message and query formats that must access qualifier-based semantics.

Below we consider specific implications of a qualifier-based model for these and other components of the Translator architecture.

VII. Considerations for Specific Translator Components

How will components/tasks across the Translator architecture accommodate a qualifier-based modeling approach.

1. Biolink Model

- **a. Roles / Responsibilities**
    - i. Define semantic data models that support a qualifier-based approach
- **b. Changes /Considerations**
    - i. Trim predicate hierarchy to remove those where semantics have been moved into qualifiers
    - ii. Add node qualifier slots to Association model
    - iii. Refactor/add to set of existing statement qualifier slots
    - iv. Create linkML enumerations to support constraints on permissible values for qualifier slots
    - v. Define/refactor Association types, which will include relevant qualifier slots and constraints
- **c. Needs:**
    - i. Feedback from all components about the utility, feasibility, and challenges posed by the proposed qualifier-based approach - so we can decide if the approach is tenable, and to what extent we will leverage qualifiers.
    - ii. Improved access to summary level descriptions of data provided by KPs.
        1. We need to know what is in the data to inform and help prioritize modeling efforts. And to help KPs migrate their data to new structures.
        2. MetaKG provides starting point - but need additional functionality to slide/dice/filter these metaKG summaries in more flexible way

2. LinkML Modeling Language

- **a. Roles / Responsibilities**
    - i. Provide modeling language with expressivity to support Translator use case
- **b. Changes / Considerations**

i. Improve meta-language to support new enum requirements (e.g. Hierarchical value sets, etc.)

c. Needs

i. Requirements from Translator Biolink team

3. TRAPI

a. Roles / Responsibilities

i. Provide a standard structure for passing queries and responses between Translator systems (using Biolink semantics)

b. Changes /Considerations

i. Relatively insulated from major disruptions/changes - given built in flexibility and generality of the TRAPI model

ii. Will need to extend schema with the following features to support qualifier (see PR#330)

1. Add a 'qualifiers' property to the Edge object
2. Define a new Qualifier object
    a. will be a generic key-value structure like an Attribute, but specific to holding info that contributes to Statement semantics (as opposed to EPC)
3. Extend QEdge model to include support for qualifier-constraints

. . . these will be included in the next TRAPI release (this summer?)

c. Needs

i. Requirements from other Translator teams/components (data modeling, UI, . . .)

4. Knowledge Providers

a. Roles / Responsibilities

i. Generate data compliant with Biolink Semantics

ii. Provide TRAPI compliant API endpoints that serves this data to ARS systems

iii. Implement query logic needed things like transitive closures on classes/predicates (to enable hierarchical query expansion)

b. Changes /Considerations

i. Refactor ETL/data creation code to produce structures compliant with new modeling patterns

ii. Implement query logic to support closures of qualifier values as defined in hierarchical Biolink enumerations - to support hierarchical query expansion (previously only had to worry about closure in class and predicate hierarchies)

1. Simple use case: abundance > expression, synthesis
2. Harder use case: increased abundance> increased expression, increased synthesis, decreased degradation
    a. This was easier to support when aspect+direction semantics were captured in pre-composed predicates - requiring a simple closure / expansion down the hierarchy.

b. With aspect and direction semantics split across separate fields, we need more complex rules to do this type of hierarchical inference (e.g. to infer that ‘decreased degradation’ should be returned on queries for ‘increased abundance’)

c. Needs/Questions

i. WIll KPs continue to be responsible for implementing hierarchical closures /query expansion independently?

5. Autonomous Relay Agents

a. Roles / Responsibilities

i. Receive TRAPI queries from ARS
ii. Query MetaKG endpoints to determine which KPs can answer a query passed to them
iii. Send TRAPI queries to retrieve data from relevant KPs
iv. Annotate Edges with additional information
v. Perform scoring/ranking of results
vi. Assemble and merge Edges/Results according to Translator rules/specifications
vii. Perform custom Creative Mode Inferencing operations (and return these results)
1. Typically this means taking a simple one hop query, and applying rules specific to that type of query to expand into multi-hop query(ies) that get issued to KPs, and then assembling and scoring results. But often there are other inferencing approaches such as Machine Learning, etc.

b. Changes /Considerations

i. Queries of MetaKG will be more complex to write, as 'SPO' will no longer be sufficient to uniquely identify an 'Edge Type' / 'MetaEdge', and tell the ARA if a given KP can answer a specific type of question.
1. Queries will have to now interrogate qualifiers to know if a given KP is relevant. See questions/notes about MetaKG considerations below.
ii. Creative Mode algorithms/methods will need to be modified to handle qualifier-based representation of Edges that they take as input.
1. multi-hop queries will have to look inside qualifiers for some traversed edges, to determine if a path supports the desired inference (because predicates won't hold all the required semantics for this).
2. e.g. instead of X - increases_activity_of-> Y, triple will read X - affects-> Y, and qualifiers will hold aspect (activity) and direction (increased) semantics.

6. ARS

a. Roles / Responsibilities

i. Send TRAPI query to ARAs

ii. Does ARS use MetaKG endpoints for ARS/KPS to decide which to send query to? Or is this done only by ARSs?

b. Changes /Considerations

i. …

c. Needs

i. . . .

7. User Interface

a. Roles / Responsibilities

i. Provide interface / templates to guide users in posing questions
ii. Take user input and generate TRAPI query
iii. Send query to ARS
iv. Receive results back form ARS
v. Display results to user
vi. Enable further exploration of answers, refinement of queries, etc

b. Features /Considerations

i. Query building tools and templates and workflows may need to be more complex to handle more complex modeling structures.
    1. When/how will the interface allow users to query based on qualifier values?
    2. How will the UI deal with / handle exposing the 'two levels of meaning' in a qualifier-based Statement?
        a. Leverage the higher level core triple for a first pass query, then refine? Or just query the full Statement directly by providing access to qualifiers up front?
ii. TRAPI-compliant queries generated from user input will be more complex, often requiring several qualifier-constraints to specify the desired question.
iii. See examples in Query Catalog here

c. Needs:

i. . .

Questions for UI Team

a. what types of queries does UI need to support for the September MVP?

8. MetaKG endpoints

a. **Roles / Responsibilities**

i. Enumerates/indexes the types of Edges/Statements provided by a given KP - based on unique pattern of S(cat) - predicate - O(cat)

ii. Provides programmatic query interface for ARS/ARA determine if the KP is relevant to a user query

b. **Changes /Considerations**

i. With so many key statement semantics in qualifiers, identifying unique Edge types based on unique 'S(cat)-P-O(cat)' combinations will no longer be sufficient' - and tell an ARA if a given KP can answer their question. How will MetaKG need to evolve to handle this?

ii. MetaKGs may have to advertise unique combinations of SPOQs - which, given how many possible Q slots there can be. This seems like a significant challenges our architecture/services would have to evolve to deal with.

iii. e.g. when user queries for Small Molecules that increase expression of Gene X, ARA currently looks for this record in MetaKGs

Source	subject cat	predicate	object cat
https://translator.broadinstitute.org/molepro/trapi/v1.2/meta_knowledge_graph	biolink:SmallMolecule	biolink:increases_expression_of	biolink:Gene

iv. With qualifier based representation, MetaKG will have to index unique SPOQ combinations . . .

Source	subject cat	predicate	object cat	object_aspect	object_ direction
https://translator.broadinstitute.org/molepro/trapi/v1.2/meta_knowledge_graph	biolink:SmallMolecule	biolink:causes	biolink:Gene	expression	increased

c. Needs

9. SRI Services

a. Roles / Responsibilities

i. . . .

b. Changes /Considerations

c. Needs

Appendix I: Chemical Gene-Association Schema and Enums

Association Type Schema

1. ChemicalGeneInteractionAssociation

Definition: describes a physical interaction between a chemical entity and a gene or gene product. Any biological or chemical effect resulting from such an interaction are out of scope, and covered by the ChemicalAffectsGeneAssociation type (e.g. impact of a chemical on the abundance, activity, structure, etc, of either participant in the interaction)

Schema:

subject: ChemicalEntity (1..1)
subject_form_or_variant: enum:chemical_entitiy_or_gene_or_gene_product_form (0..1)
subject_part: enum:chemical_entity_part (0..1)
subject_derivative: enum:chemical_entity_derivative (0..1)

predicate: physically interacts with (1..1)

object: GeneorGeneProduct (1..1)
object_form_or_variant: enum: gene_or_gene_product_form (0..1)
object_part: enum: genomic_entity_part (0..1)
object_derivative: enum:gene_or_gene_product_derivative (0..1)

interaction_mechanism_qualifier: enum:molecular_interaction_mechanism (0..1)
anatomical_context_qualifier: enum:anatomical_context (0..1)

2. ChemicalAffectsGeneAssociation

Definition: describes an effect that a chemical has on a gene or gene product (e.g. an impact of on its abundance, activity, localization, processing, expression, etc.)

Schema:

subject: Chemical Entity (1..1)
subject_form_or_variant: enum:chemical_entitiy_or_gene_or_gene_product_form (0..1)
subject_part: enum:chemical_entity_part (0..1)
subject_derivative: enum:chemical_entity_derivative (0..1)

subject_aspect: enum:chemical_entity_or_gene_or_gene_product_aspect (0..1)
subject_context: enum:anatomical_location (0..1)
subject_direction: enum:direction_of_change (0..1)

predicate: affects | related_to (1..1)
qualified_predicate: causes (0..1)

object: Gene or Gene Product (1..1)
object_form_or_variant: enum: chemical_entitiy_orgene_or_gene_product_form (0..1)
object_part: enum: genomic_entity_part (0..1)
object_aspect: enum: chemical_entity_or_gene_or_gene_product_aspect (0..1)
object_context: enum:anatomical_clocation (0..1)
object_direction: enum:direction_of_change (0..1)

effect_mechanism_qualifier: enum:molecular_control_mechanism (0..1)
anatomical_context_qualifier: enum:anatomical_context (0..1)

Definitions of Chem-Gene Association Statement Qualifier Slot

1. **affect_mechanism_qualifier:** A statement qualifier representing a type of molecular control mechanism through which an effect of a chemical on a gene or gene product is mediated (e.g. 'agonism', 'inhibition', 'allosteric modulation', 'channel blocker')
2. **interaction_mechanism_qualifier:** A statement qualifier representing a type of molecular interaction mechanism through which a physical interaction between chemical entities, genes, or gene products occurred.
3. **anatomical_context_qualifier:** A statement qualifier representing an anatomical location where an relationship expressed in an association took place (can be a tissue, cell type, or subcellular location).
4. **species_context_qualifier:** A statement qualifier representing a taxonomic category of species in which a relationship expressed in an association took place.

Predicates supporting Chem-Gene Associations

Predicates needed for the proposed qualifier-based representation of Chemical-Gene Associations (down from > 60 in predicate based approach)

related to
    affects
    interacts with
        physically interacts with
    contributes to
        causes

Enumerations supporting Chem-Gene Associations

Enumerations needed for the proposed qualifier-based representation of Chemical-Gene Associations

1. **enum: chemical_entity_or_gene_or_gene_product_aspect**

abundance
- expression
- synthesis
- degradation
  - cleavage
  - hydrolysis

activity
- *(. . . potentially more specific activities)*

folding
metabolic processing
mutation rate
splicing
stability
localization
transport
- secretion
- uptake

molecular modification
- acetylation
- acylation
- alkylation
- amination
- carbamoylation
- ethylation
- glutathionylation
- glycation
- glycosylation
- glucuronidation
- N-linked glycosylation
- O-linked glycosylation
- hydroxylation
- lipidation

farnesylation
geranoylation
myristoylation
palmitoylation
prenylation
methylation
nitrosation
nucleotidylation
phosphorylation
ribosylation
ADP-ribosylation
sulfation
sumoylation
ubiquitination
oxidation
reduction
carboxylation

2. **enum: chemical_entity_form_or_variant**
   analog
   . . . (potentially more as we address use cases besides CTD)

3. **enum: chemical_entity_derivative**
   metabolite
   . . . (potentially more as we address use cases besides CTD)

4. **enum: gene_or_gene_product_form**
   modified form
   mutant form
   polymorphism
   SNP
   . . . (potentially more as we address use cases besides CTD)

5. **enum: genomic_entity_part**
   3' UTR
   5' UTR
   polyA tail
   promoter
   enhancer
   exon
   intron

. . . (potentially more as we address use cases besides CTD)

6. **enum: chemical_entity_part**
   (n/a for CTD use case, but may be relevant for others. might contain values like 'epitope', or methyl-group, etc.)

7. **enum:direction_of_change**
   - increased
   - decreased

8. **enum:molecular_control_mechanism** *(hierarchy below needs work, but n/a for CTD so can do later)*
   - binding
     - inactivation
     - neutralization
     - inhibition
       - antagonism
       - molecular channel blockage
       - antibody inhibitor
       - inverse agonism
       - negative allosteric modulation
     - agonism
     - molecular channel opening
     - positive allosteric modulation
     - potentiation

9. **enum:molecular_interaction_mechanism**
   (n/a for CTD use case, but might be relevant for others)

10. **enum:anatomical_context**
    - any term from Ubreon, GO-CC, CL, . . .?

11. **enum:taxonomic_species_context**
    - any term from NCITaxonomy ontology