

Contents

1	Axiomatic Data Types	2
1.1	Atomic Data Types	2
1.2	Composite Data Types	2
2	The Substrate	3
2.1	Substrate Components	3
2.2	Substrate	4
3	Result	5
4	Interpretation of Intermediate Constructs	5
4.1	expressionConstraint	6
4.1.1	Interpretation	6
4.1.2	unrefinedExpressionConstraint	6
4.1.3	refinedExpressionConstraint	7
4.2	simpleExpressionConstraint	7
4.3	compoundExpressionConstraint	7
4.4	conjunctionExpressionConstraint	8
4.5	disjunctionExpressionConstraint	9
4.6	exclusionExpressionConstraint	9
4.7	subExpressionConstraint	9
4.8	refinement	10
4.9	expressionConstraintValue	11
4.10	Attributes and Attribute Groups	12
4.11	AttributeName and AttributeSubject	12
4.12	Attribute	12
4.13	AttributeSet and AttributeGroup	14
4.14	Compound attribute evaluation	15
4.15	Group Cardinality	16
4.16	Cardinality	17
5	Substrate Interpretations	18
5.1	AttributeExpressionConstraint	18
5.2	ConcreteAttributeConstraint	19
5.3	ConstraintOperator	20
5.4	FocusConcept	20
5.5	ConceptReference	21
6	Helper Functions	21

1 Axiomatic Data Types

1.1 Atomic Data Types

This section identifies the atomic data types that are assumed for the rest of this specification, specifically:

- **SCTID** – a SNOMED CT identifier
- **TERM** – a fully specified name, preferred term or synonym for a SNOMED CT Concept
- **REAL** – a real number
- **STRING** – a string literal
- **GROUP** – a role group identifier
- \mathbb{N} – a non-negative integer
- \mathbb{Z} – an integer

We also introduce several synonyms for *SCTID*:

- **SUBJECT** – a *SCTID* that appears in the *sourceId* position of a relationship.
- **ATTRIBUTE** – a *SCTID* that appears in the *typeId* position of a relationship.
- **OBJECT** – a *SCTID* that appears in the *destinationId* position of a relationship.
- **REFSETID** – a *SCTID* that identifies a reference set

[*SCTID*, *TERM*, *REAL*, *STRING*, *GROUP*]

SUBJECT == *SCTID*

ATTRIBUTE == *SCTID*

OBJECT == *SCTID*

REFSETID == *SCTID*

1.2 Composite Data Types

While we can't fully specify the behavior of the concrete data types portion of the specification at this point, it is still useful to spell out the anticipated behavior on an abstract level.

- **CONCRETEVALUE** – a string, integer or real literal
- **TARGET** – the target of a relationship that is either an **OBJECT** or a **CONCRETEVALUE**

CONCRETEVALUE ::= *string*⟨⟨*STRING*⟩⟩ | *integer*⟨⟨ \mathbb{Z} ⟩⟩ | *real*⟨⟨*REAL*⟩⟩

TARGET ::= *object*⟨⟨*OBJECT*⟩⟩ | *concrete*⟨⟨*CONCRETEVALUE*⟩⟩

2 The Substrate

A substrate represents the context of an interpretation. A substrate consists of:

- **c** The set of *SCTIDs* (concepts) that are considered valid in the context of the substrate. **References to any *SCTID* that is not a member of this set MUST be treated as an error.**
- **a** The set of *SCTIDs* that are considered to be valid attributes in the context of the substrate. **Reference to any *ATTRIBUTE* that is not a member of this set MUST be treated as an error.**
- **r** A set of relationship quads (subject, attribute, target, group)
- **descs** The subsumption (ISA) closure from general to specific (descendants)
- **ancs** The subsumption (ISA) closure from specific to general (ancestors)
- **refset** The reference sets within the context of the substrate whose members are members are concept identifiers (i.e. are in *c*). While not formally spelled out in this specification, it is assumed that the typical reference set function would be returning a subset of the *refsetId/referencedComponentId* tuples represented in one or more RF2 Refset Distribution tables.

Open questions: While the Core will not contain cycles, and probably NRC Extensions will not, arbitrary extensions (e.g. as a result of handling Expression libraries) may involve cycles (or at least equivalent concepts).

- **Resolution:** A reflexive, symmetric equivalence relationship (*equiv*) was added to the substrate. How it is determined is left unspecified, but a rule was added saying that an equivalent concept can not be a descendant of its equivalence.
- Is it correct to interpret the transitive closure directly against the substrate relationships (*r*) set or is there an implication of a reasoner being invoked somewhere, which could potentially render the transitive closure as logically derived from (*r*). **Answer:** If substrate includes postcoordinated expressions then subsumption would need to be calculated. **Followup:** How does one know whether it includes postcoordinated expressions? How do we express this decision formally?
- Do we want to assert that it is an error to use a non-attribute concept in a attribute position? If not, should we explicitly include a formal definition of the attribute set? **Answer:** Because the ECL doesn't declare this as an error condition, we shouldn't assert in the Z spec. **Resolution:** *a = descs attribute_concept* assertion pulled from the substrate declaration.

2.1 Substrate Components

Quad Relationships in the substrate are represented a 4 element tuples or "Quads", which consist of a subject, attribute, target and role group identifier.

Quad

s : *SUBJECT*

a : *ATTRIBUTE*

t : *TARGET*

g : *GROUP*

We will also need to recognize some "well known" identifiers: the *is_a* attribute, the *zero_group* and *attribute_concept*, the parent of all attributes

is_a : *ATTRIBUTE*

zero_group : *GROUP*

attribute_concept : *SCTID*

refset_concept : *SCTID*

2.2 Substrate

The formal definition of substrate follows, where *c* and *r* are given and the remainder are derived. The expressions below assert that:

1. All subjects and attributes and targets of type *object* in *r* must be members of the set of concepts, *c*. All attributes are also members of *a*.
2. The *is_a* relation is irreflexive.
3. *childrenOf* includes the function from a concept in *c* to the set of concepts that are the source of *c* in an *is_a* relationship in the *zero_group*. Note that substrates that include post-coordinated expressions may include additional entries in the *childrenOf* function that are not in the relationships table and, instead, are derived through the application of a DL reasoner.
4. *parentsOf* includes the function from a concept *c* to the set of concepts that are the target of *c* in an *is_a* relationship in the *zero_group*. Note that substrates that include post-coordinated expressions may include additional entries in the *parentsOf* function that are not in the relationships table and, instead, are derived through the application of a DL reasoner.
5. *parentsOf* and *childrenOf* are irreflexive.
6. The *equiv* relationship is symmetric and reflexive.
7. The ancestors function, *ancs*, is the *irreflexive* transitive closure of the *childrenOf* relationship.
8. The descendants function, *descs*, is the *irreflexive* transitive closure of the inverse of the *parentsOf* relationship.
9. The descendants relationship must not contain any concepts that are declared as equivalent to the root.

10. The reference set function is a function from *subset of c* to set of SCTID's in *c*. A SCTID that is not in the domain of *refset* cannot appear as the target of a *memberOf* function

Substrate
$c : \mathbb{P} \text{ SCTID}$ $r : \mathbb{P} \text{ Quad}$ $a : \mathbb{P} \text{ SCTID}$ $childrenOf : \text{SCTID} \rightarrow \mathbb{P} \text{ SCTID}$ $parentsOf : \text{SCTID} \rightarrow \mathbb{P} \text{ SCTID}$ $descs : \text{SCTID} \rightarrow \mathbb{P} \text{ SCTID}$ $ancs : \text{SCTID} \rightarrow \mathbb{P} \text{ SCTID}$ $equiv : \mathbb{P}(\text{SCTID} \times \text{SCTID})$ $refset : \text{REFSETID} \rightarrow \mathbb{P} \text{ SCTID}$
$\forall rel : r \bullet rel.s \in c \wedge rel.a \in c \wedge (object \sim rel.t \in \text{SCTID} \Rightarrow object \sim rel.t \in c)$ $\forall q : r \bullet q.a \in a \wedge q.a = is_a \Rightarrow q.s \neq object \sim q.t$ $\forall q : r \bullet q.a \in descs \text{ attribute_concept}$ $\text{dom } childrenOf \subseteq c \wedge \bigcup(\text{ran } childrenOf) \subseteq c$ $\forall s : c \bullet parentsOf s = \{q : r \mid q.s = s \wedge q.a = is_a \wedge q.g = zero_group \bullet object \sim q.t\}$ $\forall t : c \bullet childrenOf t = \{q : r \mid q.a = is_a \wedge q.t = object t \wedge q.g = zero_group \bullet q.s\}$ $\forall c1, c2 : c \bullet c1 \in parentsOf c2 \Leftrightarrow c2 \in childrenOf c1$ $\forall x : c \bullet (x \mapsto x) \in equiv$ $\forall x_1, x_2 : c \bullet x_1 \mapsto x_2 \in equiv \Rightarrow x_2 \mapsto x_1 \in equiv$ $\forall s : c \bullet descs s = childrenOf s \cup \bigcup\{t : childrenOf s \bullet descs t\}$ $\forall t : c \bullet ancs t = parentsOf t \cup \bigcup\{s : parentsOf t \bullet ancs s\}$ $\forall r : equiv \bullet first r \in \text{dom } descs \Rightarrow second r \notin (descs (first r))$ $\text{dom } refset \subseteq descs \text{ refset_concept} \wedge \bigcup(\text{ran } refset) \subseteq c$

3 Result

The result of applying a query against a substrate is either a (possibly empty) set of SCTID's or an *ERROR*.

$$\begin{aligned}
 \text{ERROR} ::= & \text{unknownOperation} \mid \text{unknownConceptReference} \mid \\
 & \text{unknownAttribute} \mid \text{unknownRefsetId} \\
 \text{Result} ::= & ok \langle \langle \mathbb{P} \text{ SCTID} \rangle \rangle \mid error \langle \langle \text{ERROR} \rangle \rangle
 \end{aligned}$$

4 Interpretation of Intermediate Constructs

This section carries the interpretation of the intermediate constructs – the various forms of expressions and their combinations.

4.1 expressionConstraint

$\text{expressionConstraint} =$ $(\text{refinedExpressionConstraint} / \text{unrefinedExpressionConstraint})$

4.1.1 Interpretation

The interpretation of `expressionConstraint` the interpretation of the `refinedExpressionConstraint` or the `unrefinedExpressionConstraint`.

$$\text{expressionConstraint} ::=$$

$$\text{ecrec}\langle\langle\text{refinedExpressionConstraint}\rangle\rangle \mid$$

$$\text{ecurec}\langle\langle\text{unrefinedExpressionConstraint}\rangle\rangle$$

$i_expressionConstraint :$ $Substrate \rightarrow expressionConstraint \rightarrow Result$ <hr/> $\forall ss : Substrate; ec : expressionConstraint \bullet i_expressionConstraint ss ec =$ $\text{if } \text{ecrec} \sim ec \in \text{refinedExpressionConstraint}$ $\quad \text{then } i_refinedExpressionConstraint ss (\text{ecrec} \sim ec)$ $\text{else if } \text{ecurec} \sim ec \in \text{unrefinedExpressionConstraint}$ $\quad \text{then } i_unrefinedExpressionConstraint ss (\text{ecurec} \sim ec)$ $\text{else error unknownOperation}$

4.1.2 unrefinedExpressionConstraint

An `unrefinedExpressionConstraint` is either a `compoundExpressionConstraint` or a `simpleExpressionConstraint`

$$\text{unrefinedExpressionConstraint} =$$

$$\text{compoundExpressionConstraint} / \text{simpleExpressionConstraint}$$

$$\text{unrefinedExpressionConstraint} ::=$$

$$\text{ucec}\langle\langle\text{compoundExpressionConstraint}\rangle\rangle \mid$$

$$\text{usec}\langle\langle\text{simpleExpressionConstraint}\rangle\rangle$$

$i_unrefinedExpressionConstraint :$ $Substrate \rightarrow unrefinedExpressionConstraint \rightarrow Result$ <hr/> $\forall ss : Substrate; uec : unrefinedExpressionConstraint \bullet$ $i_unrefinedExpressionConstraint ss uec =$ $\text{if } \text{ucec} \sim uec \in \text{compoundExpressionConstraint}$ $\quad \text{then } i_compoundExpressionConstraint ss (\text{ucec} \sim uec)$ $\text{else if } \text{usec} \sim uec \in \text{simpleExpressionConstraint}$ $\quad \text{then } i_simpleExpressionConstraint ss (\text{usec} \sim uec)$ $\text{else error unknownOperation}$

4.1.3 refinedExpressionConstraint

```
refinedExpressionConstraint =
  unrefinedExpressionConstraint ":" refinement /
  "(" refinedExpressionConstraint ")"
```

The interpretation of `refinedExpressionConstraint` is the intersection of the interpretation of the `unrefinedExpressionConstraint` and the `refinement`, both of which return a set of SCTID's or an error.

The interpretation of the second option adds no value.

$$\text{refinedExpressionConstraint} == \text{unrefinedExpressionConstraint} \times \text{refinement}$$

$i_refinedExpressionConstraint :$ $Substrate \rightarrow refinedExpressionConstraint \rightarrow Result$
$\forall ss : Substrate; rec : refinedExpressionConstraint \bullet$ $i_refinedExpressionConstraint ss rec =$ $intersect (i_unrefinedExpressionConstraint ss (first rec)) (i_refinement ss (second rec))$

4.2 simpleExpressionConstraint

The interpretation of `simpleExpressionConstraint` is the application of an optional constraint operator to the interpretation of `focusConcept`, which returns a set of SCTID's or an error. The interpretation of an error is the error.

```
simpleExpressionConstraint =
  [constraintOperator ] focusConcept
```

$$\begin{aligned} \text{constraintOperator} &::= \text{descendantOrSelfOf} \mid \text{descendantOf} \mid \\ &\quad \text{ancestorOrSelfOf} \mid \text{ancestorOf} \\ \text{simpleExpressionConstraint} &== \text{constraintOperator}[0..1] \times \text{focusConcept} \end{aligned}$$

$i_simpleExpressionConstraint :$ $Substrate \rightarrow simpleExpressionConstraint \rightarrow Result$
$\forall ss : Substrate; sec : simpleExpressionConstraint \bullet$ $i_simpleExpressionConstraint ss sec =$ $i_constraintOperator ss (first sec) (i_focusConcept ss (second sec))$

4.3 compoundExpressionConstraint

The interpretation of a `compoundExpressionConstraint` is the interpretation of its corresponding component.

```

compoundExpressionConstraint = conjunctionExpressionConstraint |
    disjunctionExpressionConstraint | exclusionExpressionConstraint |
    "(" compoundExpressionConstraint ")"

```

```

compoundExpressionConstraint ::=
    cecConj⟨⟨conjunctionExpressionConstraint⟩⟩ |
    cecDisj⟨⟨disjunctionExpressionConstraint⟩⟩ |
    cecExc⟨⟨exclusionExpressionConstraint⟩⟩

```

$i_compoundExpressionConstraint :$ $Substrate \rightarrow compoundExpressionConstraint \rightarrow Result$
$\forall ss : Substrate; cec : compoundExpressionConstraint \bullet$ $i_compoundExpressionConstraint ss cec =$ if $cecConj \sim cec \in conjunctionExpressionConstraint$ then $i_conjunctionExpressionConstraint ss (cecConj \sim cec)$ else if $cecDisj \sim cec \in disjunctionExpressionConstraint$ then $i_disjunctionExpressionConstraint ss (cecDisj \sim cec)$ else if $cecExc \sim cec \in exclusionExpressionConstraint$ then $i_exclusionExpressionConstraint ss (cecExc \sim cec)$ else $error unknownOperation$

$i_compoundExpressionConstraint' :$ $Substrate \rightarrow compoundExpressionConstraint \rightarrow Result$

4.4 conjunctionExpressionConstraint

conjunctionExpressionConstraint is interpreted the conjunction (intersection) of the interpretation of two or more *subExpressionConstraints*

```

conjunctionExpressionConstraint =
    subExpressionConstraint 1*(conjunction subExpressionConstraint)

```

$conjunctionExpressionConstraint == subExpressionConstraint \times seq_1(subExpressionConstraint)$

$i_conjunctionExpressionConstraint :$ $Substrate \rightarrow conjunctionExpressionConstraint \rightarrow Result$
$\forall ss : Substrate; cecr : conjunctionExpressionConstraint \bullet$ $i_conjunctionExpressionConstraint ss cecr =$ if $tail(second cecr) = \langle \rangle$ then $intersect (i_subExpressionConstraint ss (first cecr))(i_subExpressionConstraint ss (head (second cecr)))$ else $intersect (i_subExpressionConstraint ss (first cecr))(i_conjunctionExpressionConstraint ss (head (second cecr)))$

4.5 disjunctionExpressionConstraint

disjunctionExpressionConstraint is interpreted the disjunction (union) of the interpretation of two or more *subExpressionConstraints*

```
disjunctionExpressionConstraint =
  subExpressionConstraint 1*(disjunction subExpressionConstraint)
```

```
disjunctionExpressionConstraint ==
  subExpressionConstraint × seq1(subExpressionConstraint)
```

$i_disjunctionExpressionConstraint :$ $Substrate \rightarrow disjunctionExpressionConstraint \rightarrow Result$
$\forall ss : Substrate; cecr : disjunctionExpressionConstraint \bullet$ $i_disjunctionExpressionConstraint ss cecr =$ if $tail(second\ cecr) = \langle \rangle$ then $union\ (i_subExpressionConstraint\ ss\ (first\ cecr))(i_subExpressionConstraint\ ss\ (head\ (second\ cecr)))$ else $union\ (i_subExpressionConstraint\ ss\ (first\ cecr))(i_disjunctionExpressionConstraint\ ss\ ((head\ (second\ cecr))\ tail\ (second\ cecr)))$

4.6 exclusionExpressionConstraint

exclusionExpressionConstraint is interpreted the result of removing the members of the second *subExpressionConstraint* from the first

```
exclusionExpressionConstraint =
  subExpressionConstraint exclusion subExpressionConstraint
```

```
exclusionExpressionConstraint ==
  subExpressionConstraint × subExpressionConstraint
```

$i_exclusionExpressionConstraint :$ $Substrate \rightarrow exclusionExpressionConstraint \rightarrow Result$
$\forall ss : Substrate; ecr : exclusionExpressionConstraint \bullet$ $i_exclusionExpressionConstraint ss ecr =$ $minus\ (i_subExpressionConstraint\ ss\ (first\ ecr))\ (i_subExpressionConstraint\ ss\ (second\ ecr))$

4.7 subExpressionConstraint

subExpressionConstraint is interpreted as the interpretation of either a *simpleExpressionConstraint* or a *compoundExpressionConstraint*

```

subExpressionConstraint =
  simpleExpressionConstraint | "(" compoundExpressionConstraint ")"

```

```

subExpressionConstraint ::=
  secSec⟨⟨simpleExpressionConstraint⟩⟩ |
  secCec⟨⟨compoundExpressionConstraint⟩⟩

```

$i_subExpressionConstraint :$ $Substrate \rightarrow subExpressionConstraint \rightarrow Result$
$\forall ss : Substrate; sec : subExpressionConstraint \bullet$ $i_subExpressionConstraint ss sec =$ if $(secSec \sim sec) \in simpleExpressionConstraint$ then $i_simpleExpressionConstraint ss (secSec \sim sec)$ else if $secCec \sim sec \in compoundExpressionConstraint$ then $i_compoundExpressionConstraint' ss (secCec \sim sec)$ else $error\ unknownOperation$

4.8 refinement

refinement is the interpretation of one or more **attribute** or **attributeGroups** joined by **binaryOperators**

```

binaryOperator = conjunction / disjunction / minus
refinement = (attribute / attributeGroup / "(" refinement ")")
             [binaryOperator refinement]

```

```

refTarget ::= att⟨⟨attribute⟩⟩ | attg⟨⟨attributeGroup⟩⟩
refinement == refTarget × seq(binaryOperator × refTarget)

```

$i_refTarget : Substrate \rightarrow refTarget \rightarrow Result$ $i_refinement : Substrate \rightarrow refinement \rightarrow Result$
$\forall ss : Substrate; rt : refTarget \bullet i_refTarget ss rt =$ if $att \sim rt \in attribute$ then $i_attribute ss (att \sim rt)$ else if $attg \sim rt \in attributeGroup$ then $i_attributeGroup ss (attg \sim rt)$ else $error\ unknownOperation$
$\forall ss : Substrate; ref : refinement \bullet$ $i_refinement ss ref =$ $evalRefinement ss (i_refTarget ss (first\ ref)) (second\ ref)$

$evalRefinement :$ $Substrate \rightarrow Result \rightarrow seq(binaryOperator \times refTarget) \rightarrow Result$
$\forall ss : Substrate; r : Result; opt : seq(binaryOperator \times refTarget) \bullet$ $evalRefinement ss r opt =$ $\text{if } opt = \langle \rangle$ $\text{then } r$ $\text{else if } first(head\ opt) = conjunction$ $\text{then } evalRefinement ss (intersect\ r(i_refTarget'\ ss\ (second\ (head\ opt)))) (tail\ opt)$ $\text{else if } first(head\ opt) = disjunction$ $\text{then } evalRefinement ss (union\ r(i_refTarget'\ ss\ (second\ (head\ opt)))) (tail\ opt)$ $\text{else if } first(head\ opt) = exclusion$ $\text{then } evalRefinement ss (minus\ r(i_refTarget'\ ss\ (second\ (head\ opt)))) (tail\ opt)$ $\text{else } error\ unknownOperation$
$i_refTarget' : Substrate \rightarrow refTarget \rightarrow Result$

4.9 expressionConstraintValue

`expressionConstraintValue = simpleExpressionConstraint /`
`"(" (refinedExpressionConstraint /`
`compoundExpressionConstraint) ")"`

$$expressionConstraintValue ::= ecvsec \langle \langle simpleExpressionConstraint \rangle \rangle |$$

$$ecvrec \langle \langle refinedExpressionConstraint' \rangle \rangle |$$

$$ecvcec \langle \langle compoundExpressionConstraint \rangle \rangle$$

$$[refinedExpressionConstraint']$$

$i_expressionConstraintValue : Substrate \rightarrow$ $expressionConstraintValue \rightarrow Result$
$\forall ss : Substrate; ecv : expressionConstraintValue \bullet$ $i_expressionConstraintValue ss ecv =$ $\text{if } ecvsec \sim ecv \in simpleExpressionConstraint$ $\text{then } i_simpleExpressionConstraint ss (ecvsec \sim ecv)$ $\text{else if } ecvrec \sim ecv \in refinedExpressionConstraint'$ $\text{then } i_refinedExpressionConstraint' ss (ecvrec \sim ecv)$ $\text{else if } ecvcec \sim ecv \in compoundExpressionConstraint$ $\text{then } i_compoundExpressionConstraint ss (ecvcec \sim ecv)$ $\text{else } error\ unknownOperation$
$i_refinedExpressionConstraint' :$ $Substrate \rightarrow refinedExpressionConstraint' \rightarrow Result$

4.10 Attributes and Attribute Groups

The interpretation of an attribute. `attributeOperator` and `attributeName` determines the set of possible attributes in the substrate relationship table. `reverseFlag` and `expressionConstraintValue` determine the set of candidate targets (if `reverseFlag` is absent) or subjects (if `reverseFlag` is present).

`cardinality` determines the minimum and maximum matches. In all cases, only a subset of the subjects (targets if `reverseFlag` is present) in the substrate relationship table will be returned in the interpretation.

```
attribute = [cardinality] [reverseFlag] [attributeOperator] attributeName
           (comparisonOperator concreteValue / "=" expressionConstraintValue )
attributeGroup = "{" attributeSet "}"
attributeSet = attribute [binaryOperator attributeSet] / "(" attributeSet ")"
cardinality = "[" nonNegativeIntegerValue to (nonNegativeIntegerValue / many) "]"
constraintOperator = descendantOrSelfOf / descendantOf / ancestorOrSelfOf / ancestorOf
attributeOperator = descendantOrSelfOf / descendantOf
comparisonOperator = "=" / "!=" ws "=" /
                   ("n"/"N") ("o"/"O") ("t"/"T") ws "=" / "<=" / "<" / ">=" / ">"
```

4.11 AttributeName and AttributeSubject

i_attributeName verifies that *conceptReference* is a valid concept in the substrate and interprets it as a set of (exactly one) *ATTRIBUTE*

i_attributeSubject interprets the constraintOperator, if any in the context of the attribute name and interprets it as a set of *ATTRIBUTE*s

Question: Should we add any additional validation checks?

attributeName == *conceptReference*

attributeOperator == {*descendantOrSelfOf*, *descendantOf*}

attributeSubject == *attributeOperator*[0..1] × *attributeName*

<i>i_attributeName</i> : <i>Substrate</i> → <i>attributeName</i> → <i>Result</i>
<i>i_attributeSubject</i> : <i>Substrate</i> → <i>attributeSubject</i> → <i>Result</i>
$\forall ss : \text{Substrate}; an : \text{attributeName} \bullet i_attributeName\ ss\ an =$ $i_conceptReference\ ss\ an$
$\forall ss : \text{Substrate}; as : \text{attributeSubject} \bullet i_attributeSubject\ ss\ as =$ $i_constraintOperator\ ss\ (first\ as)\ (i_attributeName\ ss\ (second\ as))$

4.12 Attribute

`attribute` consists of an optional `cardinality` and an `attributeConstraint`.

unlimitedNat ::= *num*⟨ℕ⟩ | *many*
cardinality == ℕ × *unlimitedNat*

<i>attribute</i>	_____
<i>card</i> :	<i>cardinality</i> [0 .. 1]
<i>attw</i> :	<i>attributeConstraint</i>

attributeConstraint is either an attribute expression constraint or a concrete value constraint.

$$\text{attributeConstraint} ::= \text{aec}\langle\langle\text{attributeExpressionConstraint}\rangle\rangle \mid \text{acvc}\langle\langle\text{attributeConcreteValueConstraint}\rangle\rangle$$

attributeExpressionEonstraint is a combination of a subject constraint (target if reverse flag is true) and an expression constraint value whose interpretation yields a set of SCTID's that filter the target (subject if reverse flag).

[*reverseFlag*]

<i>attributeExpressionConstraint</i>	_____
<i>a</i> :	<i>attributeSubject</i>
<i>rf</i> :	<i>reverseFlag</i> [0 .. 1]
<i>cs</i> :	<i>expressionConstraintValue</i>

A concrete value constraint is the combination of a subject constraint and a comparison operator/concrete value whose interpretation yields a set of subject ids. The intersection of the subject and concrete value interpretation is the interpretation of **attributeConcreteValueConstraint**

$$\text{comparisonOperator} ::= \text{eq} \mid \text{neq} \mid \text{gt} \mid \text{ge} \mid \text{lt} \mid \text{le}$$

<i>attributeConcreteValueConstraint</i>	_____
<i>a</i> :	<i>attributeSubject</i>
<i>op</i> :	<i>comparisonOperator</i>
<i>v</i> :	<i>concreteValue</i>

The interpretation of an attribute is the interpretation of the cardinality applied against the underlying attribute constraint, producing a set of SCTID's or an error condition

The interpretation of an attribute constraint is the interpretation of either the attribute expression constraint or the concrete expression constraint that produces a set of Quads or an error condition.

The actual interpretations if attribute expression and concrete value are in Section 5

<i>i_attribute</i> : <i>Substrate</i> \rightarrow <i>attribute</i> \rightarrow <i>Result</i> <i>i_attributeConstraint</i> : <i>Substrate</i> \rightarrow <i>attributeConstraint</i> \rightarrow <i>Quads</i> <i>i_attributeExpressionConstraint</i> : <i>Substrate</i> \rightarrow <i>attributeExpressionConstraint</i> \rightarrow <i>Quads</i> <i>i_attributeConcreteValueConstraint</i> : <i>Substrate</i> \rightarrow <i>attributeConcreteValueConstraint</i> \rightarrow <i>Quads</i>
$\forall ss : \text{Substrate}; a : \text{attribute} \bullet$ <i>i_attribute</i> <i>ss</i> <i>a</i> = <i>i_cardinality</i> <i>a</i> . <i>card</i> (<i>i_attributeConstraint</i> <i>ss</i> <i>a</i> . <i>attw</i>) $\forall ss : \text{Substrate}; aw : \text{attributeConstraint} \bullet i_attributeConstraint\ ss\ aw =$ if <i>aec</i> \sim <i>aw</i> \in <i>attributeExpressionConstraint</i> then <i>i_attributeExpressionConstraint</i> <i>ss</i> (<i>aec</i> \sim <i>aw</i>) else if <i>acvc</i> \sim <i>aw</i> \in <i>attributeConcreteValueConstraint</i> then <i>i_attributeConcreteValueConstraint</i> <i>ss</i> (<i>acvc</i> \sim <i>aw</i>) else <i>qerror</i> <i>unknownOperation</i> $\forall ss : \text{Substrate}; aec : \text{attributeExpressionConstraint}; eca : \text{expressionConstraintArgs} \mid$ <i>eca</i> . <i>atts</i> = <i>i_attributeSubject</i> <i>ss</i> <i>aec</i> . <i>a</i> \wedge <i>eca</i> . <i>rf</i> = <i>aec</i> . <i>rf</i> \wedge <i>eca</i> . <i>subjOrTarg</i> = <i>i_expressionConstraintValue</i> <i>ss</i> <i>aec</i> . <i>cs</i> \bullet <i>i_attributeExpressionConstraint</i> <i>ss</i> <i>aec</i> = <i>i_attributeExpression</i> <i>ss</i> <i>eca</i> $\forall ss : \text{Substrate}; awc : \text{attributeConcreteValueConstraint}; aca : \text{concreteConstraintArgs} \mid$ <i>aca</i> . <i>atts</i> = <i>i_attributeSubject</i> <i>ss</i> <i>awc</i> . <i>a</i> \wedge <i>aca</i> . <i>op</i> = <i>awc</i> . <i>op</i> \wedge <i>aca</i> . <i>t</i> = <i>awc</i> . <i>v</i> \bullet <i>i_attributeConcreteValueConstraint</i> <i>ss</i> <i>awc</i> = <i>i_concreteAttributeConstraint</i> <i>ss</i> <i>aca</i>

4.13 AttributeSet and AttributeGroup

An **attributeGroup** is an **attributeSet**. An **attributeSet** consists of a sequence of one or more attribute constraints joined by **binaryOperators**.

binaryOperator ::= *conjunction* | *disjunction* | *exclusion*
attributeGroup == *attributeSet*
attributeSet == *attribute* \times seq(*binaryOperator* \times *attribute*)

$i_attributeGroup : Substrate \rightarrow attributeGroup \rightarrow Result$
$i_attributeSet : Substrate \rightarrow attributeSet \rightarrow Result$
$\forall ss : Substrate; ag : attributeGroup \bullet i_attributeGroup ss ag =$ $i_attributeSet ss ag$
$\forall ss : Substrate; a : attributeSet \bullet i_attributeSet ss a =$ $result (evalCmpndAtt ss (i_groupedAttribute ss (first a)) (second a))$

4.14 Compound attribute evaluation

The left-to-right evaluation of **attributeSet**. The interpretation takes a *lhs* as a function from *SCTID* to *GROUP* and a *rhs* which is sequence of operator/attribute tuples and recursively interprets the *lhs* and interpretation of the head of the *rhs*.

$ \begin{array}{l} \text{evalCmpndAtt} : \\ \quad \text{Substrate} \rightarrow \text{IDGroups} \rightarrow \text{seq}(\text{binaryOperator} \times \text{attribute}) \rightarrow \text{IDGroups} \\ \text{gintersect}, \text{gunion}, \text{gminus}, \text{gfirstError} : \\ \quad \text{IDGroups} \rightarrow \text{IDGroups} \rightarrow \text{IDGroups} \end{array} $
$ \begin{array}{l} \forall ss : \text{Substrate}; lhs : \text{IDGroups}; rhs : \text{seq}(\text{binaryOperator} \times \text{attribute}) \bullet \\ \text{evalCmpndAtt } ss \text{ lhs } rhs = \\ \text{if } rhs = \langle \rangle \\ \quad \text{then } lhs \\ \text{else if } \text{first}(\text{head } rhs) = \text{conjunction} \\ \quad \text{then } \text{evalCmpndAtt } ss (\text{gintersect } lhs (\text{i_groupedAttribute } ss (\text{second}(\text{head } rhs)))) (\text{tail } rhs) \\ \text{else if } \text{first}(\text{head } rhs) = \text{disjunction} \\ \quad \text{then } \text{evalCmpndAtt } ss (\text{gunion } lhs (\text{i_groupedAttribute } ss (\text{second}(\text{head } rhs)))) (\text{tail } rhs) \\ \text{else if } \text{first}(\text{head } rhs) = \text{exclusion} \\ \quad \text{then } \text{evalCmpndAtt } ss (\text{gminus } lhs (\text{i_groupedAttribute } ss (\text{second}(\text{head } rhs)))) (\text{tail } rhs) \\ \quad \text{else } \text{gerror unknownOperation} \\ \\ \forall a, b, r : \text{IDGroups} \mid \\ \quad r = \text{if } \text{gerror} \sim a \in \text{ERROR} \vee \text{gerror} \sim b \in \text{ERROR} \text{ then } \text{gfirstError } a \text{ } b \\ \quad \text{else } gv (gv \sim a \cap gv \sim b) \bullet \\ \quad \text{gintersect } a \text{ } b = r \\ \\ \forall a, b, r : \text{IDGroups} \mid \\ \quad r = \text{if } \text{gerror} \sim a \in \text{ERROR} \vee \text{gerror} \sim b \in \text{ERROR} \text{ then } \text{gfirstError } a \text{ } b \\ \quad \text{else } gv (gv \sim a \cup gv \sim b) \bullet \\ \quad \text{gunion } a \text{ } b = r \\ \\ \forall a, b, r : \text{IDGroups} \mid \\ \quad r = \text{if } \text{gerror} \sim a \in \text{ERROR} \vee \text{gerror} \sim b \in \text{ERROR} \text{ then } \text{gfirstError } a \text{ } b \\ \quad \text{else } gv (gv \sim a \setminus gv \sim b) \bullet \\ \quad \text{gminus } a \text{ } b = r \end{array} $
$ \begin{array}{l} \text{i_groupedAttribute} : \\ \quad \text{Substrate} \rightarrow \text{attribute} \rightarrow \text{IDGroups} \\ \\ \forall ss : \text{Substrate}; a : \text{attribute} \bullet \\ \quad \text{i_groupedAttribute } ss \text{ } a = \text{i_groupCardinality } (\text{i_attributeConstraint } ss \text{ } a.\text{attw}) \text{ } a.\text{card} \end{array} $

4.15 Group Cardinality

The interpretation of cardinality within a group impose additional constraints:

- $[0..n]$ – the set of all substrate concept codes that have at least one group (entry) in the substrate relationships and, at most n matching entries in the same group

- $[0..0]$ – the set of all substrate concept codes that have at least one group (entry) in the substrate relationships and *no* matching entries
- $[1..*]$ – (default) at least one matching entry in the substrate relationships
- $[m_1..n_1]op[m_2..n_2]...$ – set of substrate concept codes where there exists at least one group where all conditions are simultaneously true

The interpretation of a grouped cardinality is a function from a set of SCTID's to the groups in which they were qualified.

The algorithm below partitions the input set of Quads by group and validates the cardinality on a per-group basis. Groups that pass are returned

TODO: This assumes that *q.t* is always type object. It doesn't say what to do if it is concrete **TODO:** the *gresult* function seems to express what is described below more simply

i_groupCardinality :

$Quads \rightarrow cardinality[0..1] \rightarrow IDGroups$

$\forall quads : Quads; oc : cardinality[0..1]; uniqueGroups : \mathbb{P} GROUP;$
 $quadsByGroup : GROUP \rightarrow \mathbb{P} Quad \mid$
 $uniqueGroups = \{q : qids\ quads \bullet q.g\} \wedge$
 $quadsByGroup = \{g : uniqueGroups; q : \mathbb{P} Quad \mid$
 $q = \{e : qids\ quads \mid e.g = g\} \bullet g \mapsto (evalCardinality\ oc\ q)\} \bullet$
 $i_groupCardinality\ quads\ oc =$
 $gv \{sctid : SCTID; groups : \mathbb{P} GROUP \mid sctid \in \{q : \bigcup(\text{ran}\ quadsByGroup)\} \bullet$
 $\text{if } qidd\ quads = subjects \text{ then } q.s \text{ else } object \sim q.t\} \wedge$
 $groups = \{g : \text{dom}\ quadsByGroup \mid (\exists q : quadsByGroup\ g \bullet$
 $sctid = \text{if } qidd\ quads = subjects \text{ then } q.s \text{ else } object \sim q.t)\} \bullet$
 $sctid \mapsto groups\}$

4.16 Cardinality

Interpretation: *cardinality* is tested against a set of quads with the following rules:

1. Errors are propagated
2. No cardinality or passing cardinality returns the subjects / targets of the set of quads
3. Otherwise return an empty set

i_cardinality :

$cardinality[0..1] \rightarrow Quads \rightarrow Result$

$\forall oc : cardinality[0..1]; qr : Quads \bullet$
 $i_cardinality\ oc\ qr =$
 $\text{if } qerror \sim qr \in ERROR \text{ then } result\ (gresult\ qr)$
 $\text{else } result\ (gresult\ (qv\ (evalCardinality\ oc\ (qids\ qr),\ qidd\ qr)))$

evalCardinality Evaluate the cardinality of a set, returning the set if it meets the constraints, otherwise return the empty set.

$evalCardinality : cardinality[0..1] \rightarrow \mathbb{P} Quad \rightarrow \mathbb{P} Quad$
$\forall oc : cardinality[0..1]; s : \mathbb{P} Quad \bullet evalCardinality oc s =$ if $\#oc = 0$ then s else if $(\#s \geq first(head\ oc)) \wedge$ $(second(head\ oc) = many \vee num^\sim(second(head\ oc)) \leq \#s)$ then s else \emptyset

5 Substrate Interpretations

This section defines the interpretations that are realized against the substrate.

5.1 AttributeExpressionConstraint

expressionConstraintArgs carries the arguments necessary to evaluate an attribute expression

- **rf** – If present, return subjects matching attribute/target subjects
- **subjOrTarg** – Set of subject or target SCTIDS (or error)
- **atts** – Set of attributes to evaluate

$expressionConstraintArgs$
$rf : reverseFlag[0..1]$ $subjOrTarg : Result$ $atts : Result$

Evaluate an attribute expression, returning the set of quads in the substrate relationship table with the supplied subject / attribute if *rf* is absent and attribute / target if *rf* is present. The interpretation returns an error there is an error in either the subject/targets or attributes. **TODO:** We don't say what to do if the target (t) is concrete

$i_attributeExpression :$ $Substrate \rightarrow expressionConstraintArgs \rightarrow Quads$
$\forall ss : Substrate; args : expressionConstraintArgs;$ $sti : \mathbb{P} TARGET; atts : \mathbb{P} ATTRIBUTE \mid$ $sti = ok \sim args.subjectOrTarg \wedge atts = ok \sim args.atts \bullet$ $i_attributeExpression ss args =$ if $error \sim (bigunion\{args.subjectOrTarg, args.atts\}) \in ERROR$ then $qfirstError\{args.subjectOrTarg, args.atts\}$ else if $\neg (sti \cup atts) \subseteq ss.c$ then $qerror\ unknownConceptReference$ else if $\#args.rf = 0$ then $qv(\{s : SUBJECT; t : sti; a : atts; g : GROUP; re : ss.r \mid$ $re.t = t \wedge re.a = a \bullet re\}, subjects)$ else $qv(\{s : sti; a : atts; t : TARGET; g : GROUP; re : ss.r \mid$ $re.s = object \sim t \wedge re.a = a \bullet re\}, targets)$

5.2 ConcreteAttributeConstraint

```

concreteValue = QM stringValue QM / "#" numericValue
stringValue = 1*(anyNonEscapedChar / escapedChar)
numericValue = decimalValue / integerValue
integerValue = ( ["-"/"+"] digitNonZero *digit ) / zero

```

The mechanism for interpreting *concreteValue* is not fully specified at this point. Much of the rest of the machinery is focused around interpreting constraints in the context of quads - subject, attribute, target and group, so the function is currently defined as returning a *Quads*, but another type or interpretation may be in order.

$concreteValue ::= stringValue \mid integerValue \mid realValue$

concreteConstraintArgs The arguments to the concrete value function.

- **atts** – list of attributes to test
- **op** – operator
- **t** – value to test against

Note: *reverseFlag* has no meaning for concrete constraints.

$concreteConstraintArgs$
$atts : Result$ $op : comparisonOperator$ $t : concreteValue$

i_concreteAttributeConstraint :
 $Substrate \rightarrow concreteConstraintArgs \rightarrow Quads$

5.3 ConstraintOperator

Issue: What is the behavior on a transitive closure that has a cycle? At the moment we always remove the focus concept on *descendantsOf* and *ancestorsOf*

Interpretation: Apply the substrate forward (*descs*) or reverse (*ancs*) functions to a set of SCTID's in the supplied *Result*. Error conditions are propagated.

i_constraintOperator :
 $Substrate \leftrightarrow constraintOperator[0..1] \leftrightarrow Result \leftrightarrow Result$

completeFun : $(SCTID \leftrightarrow \mathbb{P} SCTID) \rightarrow SCTID \rightarrow \mathbb{P} SCTID$

$\forall ss : Substrate; oco : constraintOperator[0..1]; refset : Result \bullet$
i_constraintOperator *ss oco refset* =
if *error* ~ *refset* $\in ERROR \vee \#oco = 0$
 then *refset*
else if *head oco* = *descendantOrSelfOf*
 then *ok*($\bigcup \{id : ids \text{ refset} \bullet completeFun \text{ ss.descs } id \cup \{id\}\}$)
else if *head oco* = *descendantOf*
 then *ok*($\bigcup \{id : ids \text{ refset} \bullet completeFun \text{ ss.descs } id \setminus \{id\}\}$)
else if *head oco* = *ancestorOrSelfOf*
 then *ok*($\bigcup \{id : ids \text{ refset} \bullet completeFun \text{ ss.ancs } id \cup \{id\}\}$)
else if *head oco* = *ancestorOf*
 then *ok*($\bigcup \{id : ids \text{ refset} \bullet completeFun \text{ ss.ancs } id \setminus \{id\}\}$)
else *error unknownOperation*

$\forall f : (SCTID \leftrightarrow \mathbb{P} SCTID); id : SCTID \bullet completeFun \text{ f } id =$
if *id* $\in \text{dom } f$ **then** *f id* **else** \emptyset

5.4 FocusConcept

focusConcept = [memberOf] *conceptReference*

focusConcept ::= *mo*⟨⟨*conceptReference*⟩⟩ | *cr*⟨⟨*conceptReference*⟩⟩

$i_focusConcept : Substrate \rightarrow focusConcept \rightarrow Result$
$\forall ss : Substrate; fc : focusConcept \bullet$ $i_focusConcept ss fc =$ if $cr \sim fc \in conceptReference$ then $i_conceptReference ss (cr \sim fc)$ else if $mo \sim fc \in conceptReference$ then $i_memberOf ss (mo \sim fc)$ else $error unknownOperation$

$i_memberOf : Substrate \rightarrow conceptReference \rightarrow Result$
$\forall ss : Substrate; cr : conceptReference \bullet$ $i_memberOf ss cr =$ (let $sctid == toSCTID cr \bullet$ if $sctid \notin \text{dom } ss.refset$ then $error unknownRefsetId$ else $ok (ss.refset sctid)$

5.5 ConceptReference

conceptReference = conceptId ["|" Term "|"]
conceptId = sctId

Interpretation: conceptReference is interpreted as SCTID if it is in the list of concepts in the substrate, otherwise as an error.

[conceptReference]

$toSCTID : conceptReference \rightarrow SCTID$
$i_conceptReference : Substrate \rightarrow conceptReference \rightarrow Result$
$\forall ss : Substrate; c : conceptReference \bullet i_conceptReference ss c =$ if $(toSCTID c) \in ss.c$ then $ok\{(toSCTID c)\}$ else $error unknownConceptReference$

6 Helper Functions

- *Quads* is either a collection of *Quads* or an error condition. If *Quads*, it also carries a direction indicator that determines whether it represents a set of subjects or targets.
- *IDGroups* is a map from SCTID's to their passing groups or an error condition

$$\begin{aligned}
direction &::= subjects \mid targets \\
Quads &::= qv\langle\mathbb{P} Quad \times direction\rangle \mid qerror\langle ERROR\rangle \\
IDGroups &::= gv\langle SCTID \rightarrow \mathbb{P} GROUP\rangle \mid gerror\langle ERROR\rangle
\end{aligned}$$

Results functions

- *ids* – return the set of sctids in a result or the empty set if there is an error
- *quads* – return the set of quads in a quads result or an empty set if there is an error
- *quidd* – return the direction of a quads result. Undefined if error
- *quids* – return the sctid to group map in an id group or an empty map if there is error
- *gresult* – convert a set of quads into a set of id groups
- *result* – convert a set of id groups into a simple result.

$ids : Result \rightarrow \mathbb{P} SCTID$ $qids : Quads \rightarrow \mathbb{P} Quad$ $qidd : Quads \rightarrow direction$ $gids : IDGroups \rightarrow SCTID \rightarrow \mathbb{P} GROUP$ $gresult : Quads \rightarrow IDGroups$ $result : IDGroups \rightarrow Result$
$\forall r : Result \bullet ids\ r =$ if $error \sim r \in ERROR$ then \emptyset else $ok \sim r$ $\forall q : Quads \bullet qids\ q =$ if $qerror \sim q \in ERROR$ then \emptyset else $first(qv \sim q)$ $\forall q : Quads \bullet qidd\ q =$ $second(qv \sim q)$ $\forall g : IDGroups \bullet gids\ g =$ if $gerror \sim g \in ERROR$ then \emptyset else $gv \sim g$ $\forall q : Quads \bullet gresult\ q =$ if $qerror \sim q \in ERROR$ then $gerror(qerror \sim q)$ else if $qidd\ q = subjects$ then $gv\ \{s : SCTID \mid (\exists qr : qids\ q \bullet s = qr.s) \bullet$ $s \mapsto \{qr : qids\ q \bullet qr.g\}\}$ else $gv\ \{t : SCTID \mid (\exists qr : qids\ q \bullet t = object \sim qr.t) \bullet$ $t \mapsto \{qr : qids\ q \bullet qr.g\}\}$ $\forall g : IDGroups \bullet result\ g =$ if $gerror \sim g \in ERROR$ then $error(gerror \sim g)$ else $ok(dom(gv \sim g))$

Definition of the various functions that are performed on the result type.
 $firstError$ is not fully defined – it takes a set of results with one or more errors
and returns an aggregation of all of the errors.

$union, intersect, minus : Result \rightarrow Result \rightarrow Result$ $bigunion, bigintersect : \mathbb{P} Result \rightarrow Result$ $firstError : \mathbb{P} Result \rightarrow Result$ $qfirstError : \mathbb{P} Result \rightarrow Quads$
$\forall x, y : Result \bullet union\ x\ y =$ if $error \sim x \in ERROR$ then x else if $error \sim y \in ERROR$ then y else $ok\ ((ok \sim x) \cup (ok \sim y))$ $\forall x, y : Result \bullet intersect\ x\ y =$ if $error \sim x \in ERROR$ then x else if $error \sim y \in ERROR$ then y else $ok\ ((ok \sim x) \cap (ok \sim y))$ $\forall x, y : Result \bullet minus\ x\ y =$ if $error \sim x \in ERROR$ then x else if $error \sim y \in ERROR$ then y else $ok\ ((ok \sim x) \setminus (ok \sim y))$ $\forall refset : \mathbb{P} Result \bullet bigunion\ refset =$ if $\exists r : refset \bullet error \sim r \in ERROR$ then $firstError\ refset$ else $ok\ (\bigcup \{r : refset \bullet ids\ r\})$ $\forall refset : \mathbb{P} Result \bullet bigintersect\ refset =$ if $\exists r : refset \bullet error \sim r \in ERROR$ then $firstError\ refset$ else $ok\ (\bigcap \{r : refset \bullet ids\ r\})$

Representing optional elements of type T . Representing it as a sequence allows us to determine absence by $\#T = 0$ and the value by $head\ T$.

$$T[0..1] == \{s : seq\ T \mid \#s \leq 1\}$$