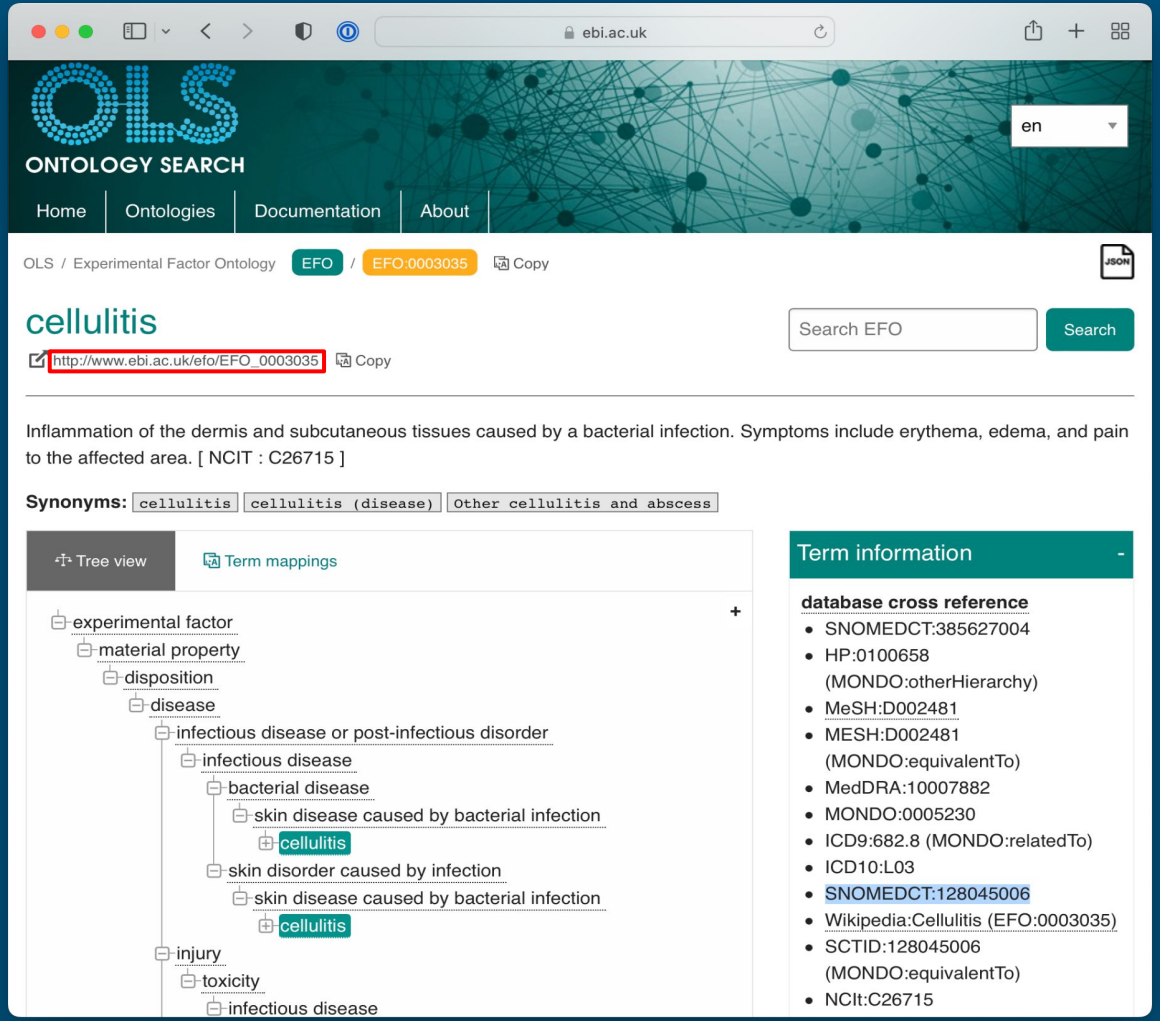
Concept IRIs in FHIR

Gaurav Vaidya

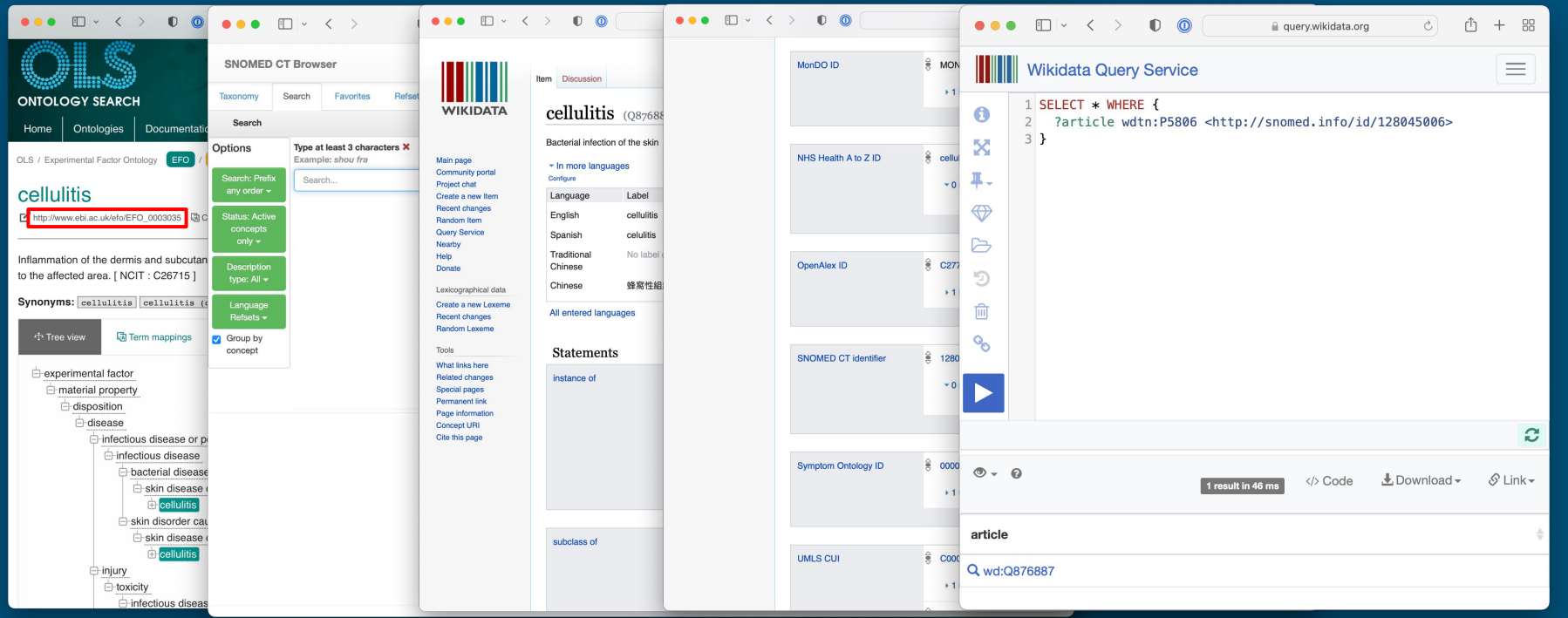
Jul 18, 2022

Slides: https://bit.ly/fhir-rdf-concept-iris-july-2022-slides

Full details: https://bit.ly/fhir-rdf-concept-iris-july-2022


ebi.ac.uk
OLS
ONTOLOGY SEARCH
en
Home
Ontologies
Documentation
About
OLS / Experimental Factor Ontology
EFO
EFO:0003035
Copy
JSON
cellulitis
http://www.ebi.ac.uk/efo/EFO_0003035
Copy
Search EFO
Search
Inflammation of the dermis and subcutaneous tissues caused by a bacterial infection. Symptoms include erythema, edema, and pain to the affected area. [NCIT : C26715]
Synonyms:
cellulitis
cellulitis (disease)
Other cellulitis and abscess
Tree view
Term mappings
experimental factor
material property
disposition
disease
infectious disease or post-infectious disorder
infectious disease
bacterial disease
skin disease caused by bacterial infection
cellulitis
skin disorder caused by infection
skin disease caused by bacterial infection
cellulitis
injury
toxicity
infectious disease
Term information
database cross reference
SNOMEDCT:385627004
HP:0100658 (MONDO:otherHierarchy)
MeSH:D002481
MESH:D002481 (MONDO:equivalentTo)
MedDRA:10007882
MONDO:0005230
ICD9:682.8 (MONDO:relatedTo)
ICD10:L03
SNOMEDCT:128045006
Wikipedia:Cellulitis (EFO:0003035)
SCTID:128045006 (MONDO:equivalentTo)
NCIt:C26715

The Linked Data world: cellulitis



http://snomed.info/id/128045006

Concept IRIs are widely used in RDF

Form

URIs for components, based on the corresponding SCTID, take the following form:

http://snomed.info/id/{sctid}

URIs for members of a Reference Set, based on the corresponding UUID, take the following form:

http://snomed.info/id/{uuid}

For simplicity this document refers to either of the above forms as a *component URI*.

Examples

The following table shows some examples of URIs for components and reference set members.

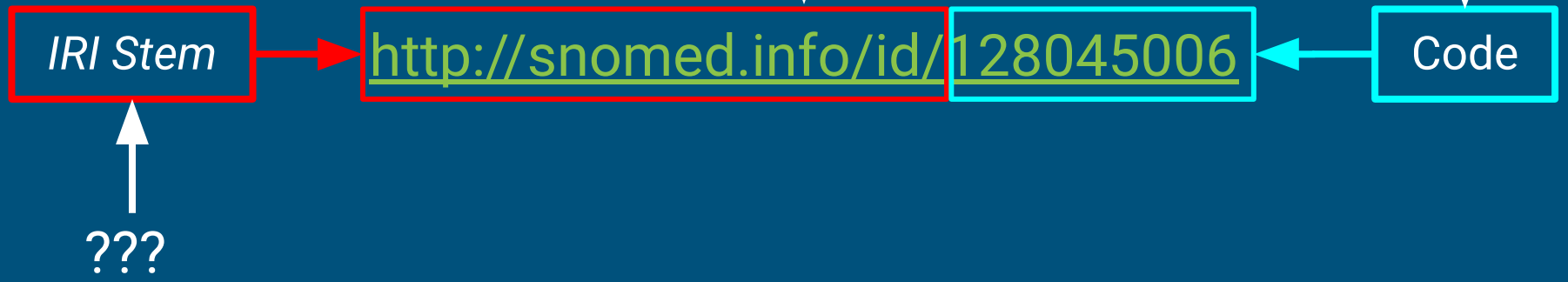
Table 2.2-1: Examples

Resource	URI
The concept 74400008 | Appendicitis|	http://snomed.info/id/74400008
The description "Appendicitis" with id=123558018	http://snomed.info/id/123558018

IRI Stem → http://snomed.info/id/ 128045006 ← Code

Coding.system/code pairs → Concept IRIs

Codesystem/NamingSystem	**Coding.system**	**Coding.code**	**Coding.display**
SNOMED CT https://terminology.hl7.org/2.0.0/CodeSystem-v3-snomed-CT.html	http://snomed.info/sct	128045006	Cellulitis (disorder)



Examples

Proposal: Add to HL7 Terminology website

Coding.system	Coding.code	IRI Stem	Concept IRI
ICD 10: http://hl7.org/fhir/sid/icd-10	G44.1	http://purl.bioontology.org/ontology/ICD10/	http://purl.bioontology.org/ontology/ICD10/G44.1
SNOMED CT: http://snomed.info/sct	128045006	http://snomed.info/id/	http://snomed.info/id/128045006
MeSH: https://www.nlm.nih.gov/mesh	D000305	https://id.nlm.nih.gov/mesh/	https://id.nlm.nih.gov/mesh/D000305
LOINC: http://loinc.org	35217-9	https://loinc.org/rdf/	https://loinc.org/rdf/35217-9
		Regional example	
Example of a code containing non-ASCII characters http://नारायणहृदयालय.example/	हृदय	http://नारायणहृदयालय.example/code/	http://नारायणहृदयालय.example/code/हृदय

IRIs vs URIs vs URLs

IRI	http://नारायणहृदयालय.example/code/हृदय	**Does not allow** characters that are significant in IRIs (such as spaces, ‘#’, ‘/’) but **allows** Unicode characters (including non-Latin characters).	RFC 3987
URI	http://xn--g2bge2acacu4d4bbf6d.example/code/%E0%A4%B9%E0%A5%83%E0%A4%A6%E0%A4%AF	Only allows ASCII characters.	RFC 3986
URL	http://xn--g2bge2acacu4d4bbf6d.example/code/%E0%A4%B9%E0%A5%83%E0%A4%A6%E0%A4%AF	Only allows ASCII characters.	RFC 3986
URN	urn:ietf:rfc:3986	Only allows ASCII characters.	RFC 8141

Who would this benefit?

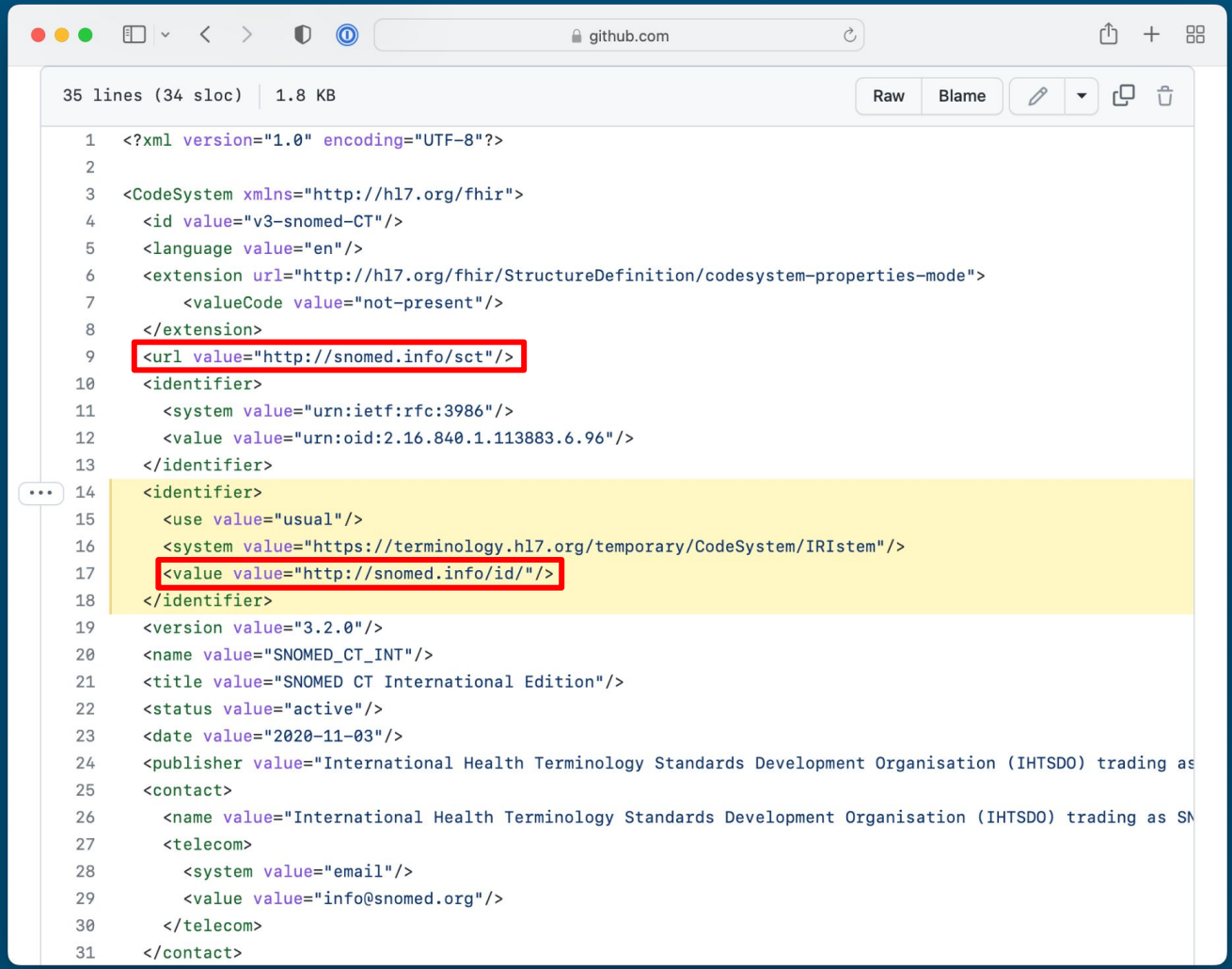
- RDF developers
- Any other applications that prefer to use a single identifier per concept, instead of a System + Code pair
- Anyone wanting to follow Web Architecture:

> ***Good practice: Identify with URIs***
>
> *To benefit from and increase the value of the World Wide Web, agents should provide URIs as identifiers for resources.*

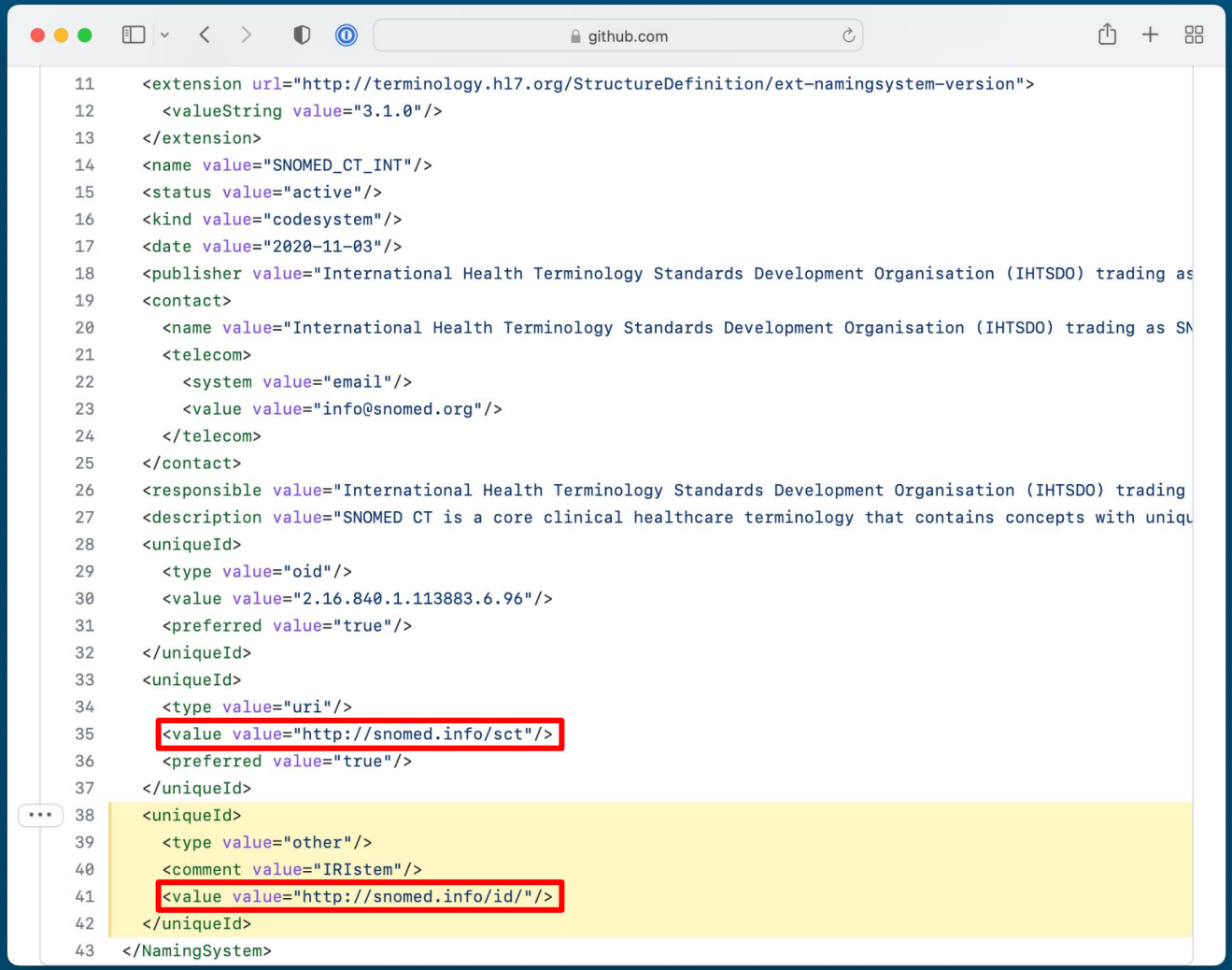
Additions needed

1. Designate an *Identifier.system* value for IRI Stems in CodeSystems (e.g. “https://terminology.hl7.org/IdentifierSystem/IRIstem”).
2. Designate “`urn:ietf:rfc:3987`” as an `Coding.system` value for IRIs in https://build.fhir.org/identifier-registry.html to be used where the `Coding.code` is already an IRI.
3. Add an “`IRIstem`” value to the `NamingSystemIdentifierType`, allowing it to be used as a value in the `NamingSystem.uniqueId.type` field in NamingSystems.

(Full details in https://bit.ly/fhir-rdf-concept-iris-july-2022)

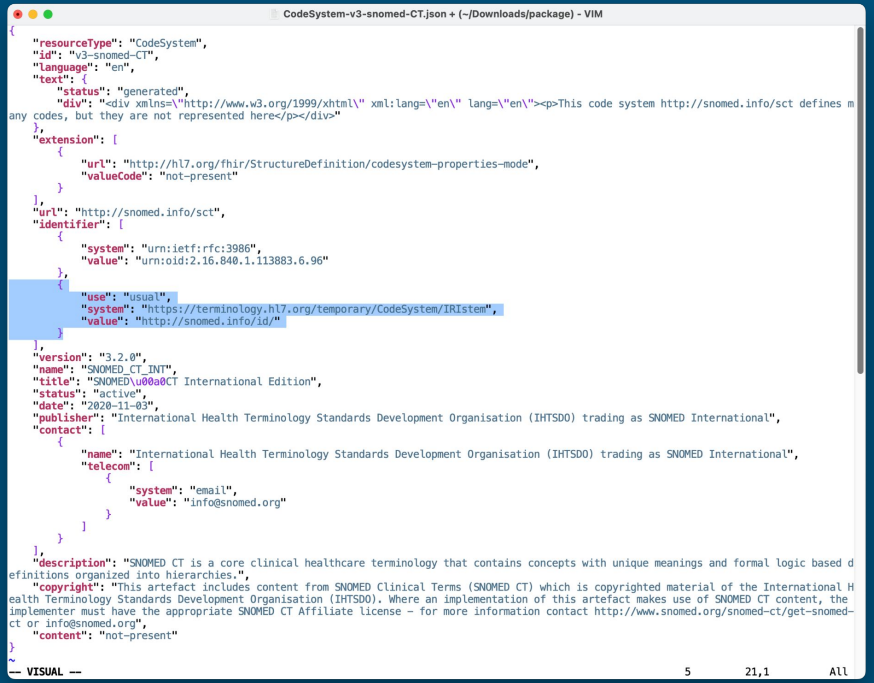


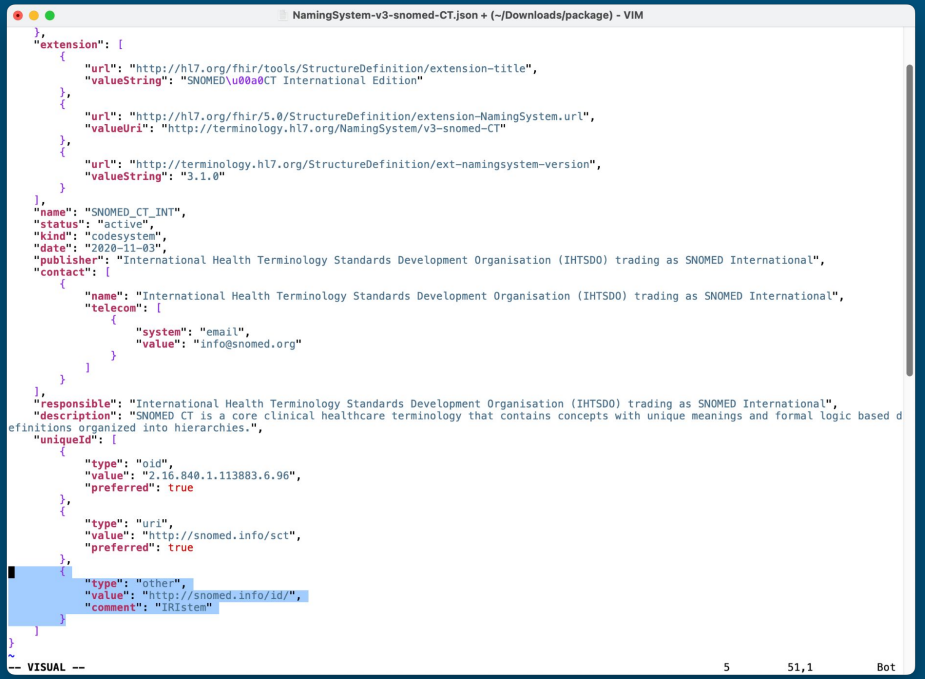
CodeSystem (SNOMED)



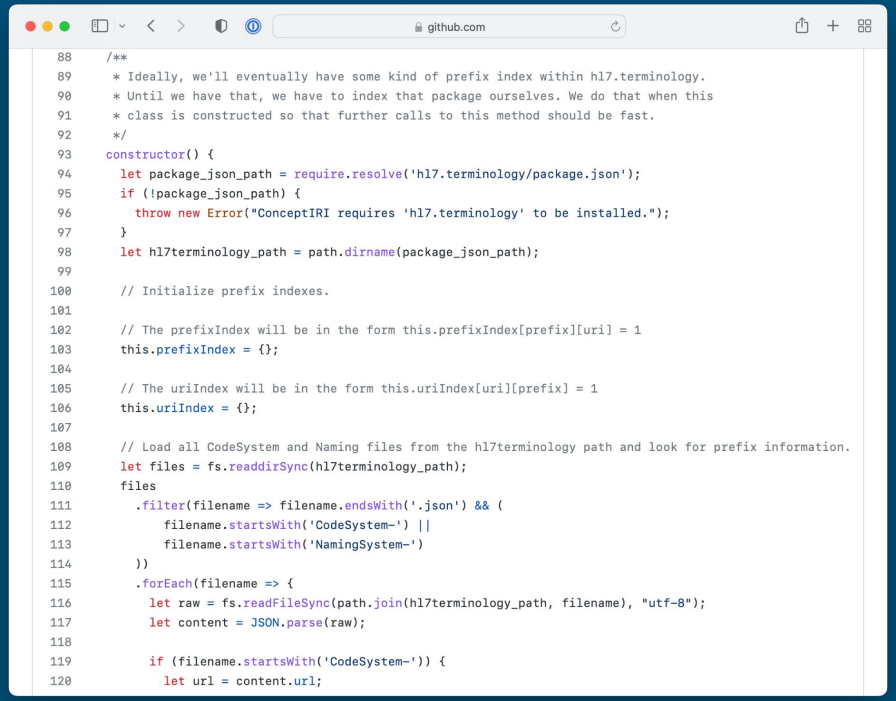
NamingSystem (SNOMED)

Also accessible through the NPM package!



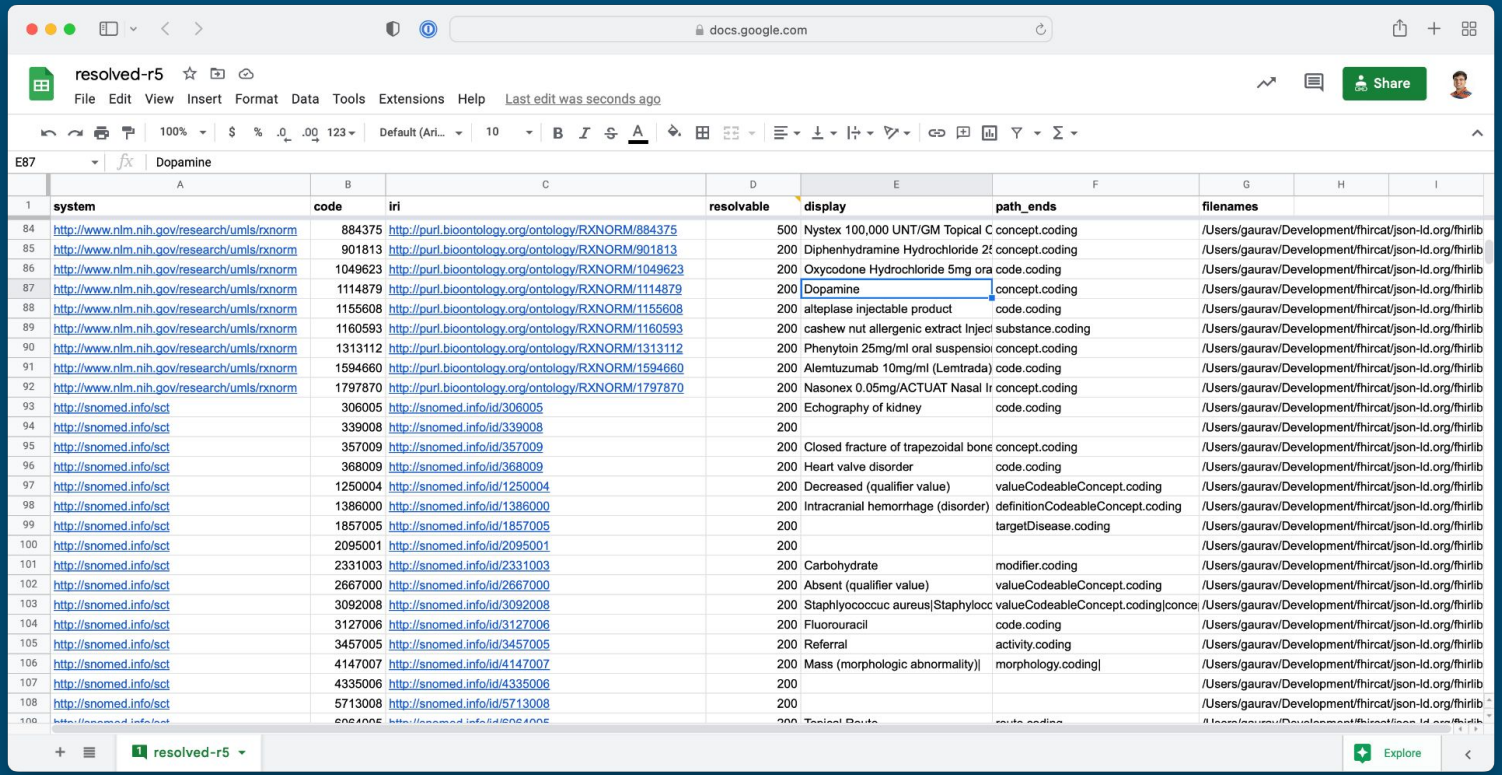


The NPM package can be used in JavaScript



https://github.com/fhircat/fhir-rdf-playground/pull/12

Tested our approach with FHIR examples



	system	code	iri	resolvable	display	path_ends	filenames
84	http://www.nlm.nih.gov/research/umls/rxnorm	884375	http://purl.bioontology.org/ontology/RXNORM/884375	500	Nystex 100,000 UNT/GM Topical C	concept.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
85	http://www.nlm.nih.gov/research/umls/rxnorm	901813	http://purl.bioontology.org/ontology/RXNORM/901813	200	Diphenhydramine Hydrochloride 25	concept.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
86	http://www.nlm.nih.gov/research/umls/rxnorm	1049623	http://purl.bioontology.org/ontology/RXNORM/1049623	200	Oxycodone Hydrochloride 5mg ora	code.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
87	http://www.nlm.nih.gov/research/umls/rxnorm	1114879	http://purl.bioontology.org/ontology/RXNORM/1114879	200	Dopamine	concept.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
88	http://www.nlm.nih.gov/research/umls/rxnorm	1155608	http://purl.bioontology.org/ontology/RXNORM/1155608	200	alteplase injectable product	code.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
89	http://www.nlm.nih.gov/research/umls/rxnorm	1160593	http://purl.bioontology.org/ontology/RXNORM/1160593	200	cashew nut allergenic extract Injec	substance.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
90	http://www.nlm.nih.gov/research/umls/rxnorm	1313112	http://purl.bioontology.org/ontology/RXNORM/1313112	200	Phenytoin 25mg/ml oral suspensio	concept.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
91	http://www.nlm.nih.gov/research/umls/rxnorm	1594660	http://purl.bioontology.org/ontology/RXNORM/1594660	200	Alemtuzumab 10mg/ml (Lemtrada)	code.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
92	http://www.nlm.nih.gov/research/umls/rxnorm	1797870	http://purl.bioontology.org/ontology/RXNORM/1797870	200	Nasonex 0.05mg/ACTUAT Nasal Ir	concept.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
93	http://snomed.info/sct	306005	http://snomed.info/id/306005	200	Echography of kidney	code.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
94	http://snomed.info/sct	339008	http://snomed.info/id/339008	200			/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
95	http://snomed.info/sct	357009	http://snomed.info/id/357009	200	Closed fracture of trapezoidal bone	concept.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
96	http://snomed.info/sct	368009	http://snomed.info/id/368009	200	Heart valve disorder	code.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
97	http://snomed.info/sct	1250004	http://snomed.info/id/1250004	200	Decreased (qualifier value)	valueCodeableConcept.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
98	http://snomed.info/sct	1386000	http://snomed.info/id/1386000	200	Intracranial hemorrhage (disorder)	definitionCodeableConcept.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
99	http://snomed.info/sct	1857005	http://snomed.info/id/1857005	200		targetDisease.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
100	http://snomed.info/sct	2095001	http://snomed.info/id/2095001	200			/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
101	http://snomed.info/sct	2331003	http://snomed.info/id/2331003	200	Carbohydrate	modifier.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
102	http://snomed.info/sct	2667000	http://snomed.info/id/2667000	200	Absent (qualifier value)	valueCodeableConcept.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
103	http://snomed.info/sct	3092008	http://snomed.info/id/3092008	200	Staphlyococcuc aureus|Staphylocc	valueCodeableConcept.coding|conce	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
104	http://snomed.info/sct	3127006	http://snomed.info/id/3127006	200	Fluorouracil	code.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
105	http://snomed.info/sct	3457005	http://snomed.info/id/3457005	200	Referral	activity.coding	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
106	http://snomed.info/sct	4147007	http://snomed.info/id/4147007	200	Mass (morphologic abnormality)|	morphology.coding|	/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
107	http://snomed.info/sct	4335006	http://snomed.info/id/4335006	200			/Users/gaurav/Development/fhircat/json-ld.org/fhirlib
108	http://snomed.info/sct	5713008	http://snomed.info/id/5713008	200			/Users/gaurav/Development/fhircat/json-ld.org/fhirlib

https://docs.google.com/spreadsheets/d/1g-TKGzKA2YdV4gELLkjL0Ku5cZN0rz-VPNXgMLaJGhk/edit#gid=1303917068

Summary

- Adding IRI stems to HL7 Terminology would be a simple change that would make interconversion between FHIR and the Linked Data world much, much easier.
- HL7 Terminology's NPM package is a powerful mechanism for extracting and using data from the HL7 Terminology records.

Feedback from Vocab WG

1. Who gets to pick IRI stems and with what criteria?
    a. Vocab/TSMG will pick IRI stems for HL7 terminologies.
    b. RDF subgroup can help find and maintain IRI stems for external terminologies.
        i. Gaurav will be in charge of figuring out how to make changes through the UTG.
2. How do we deal with versioning?
    a. Not necessary at the moment, but we will plan for this going forward.
3. How do we deal with characters that aren't ASCII?
    a. We believe we can fully support translating back and forth from FHIR strings and codes.
4. Should FHIR ValueSets be changed to better support concept IRIs?
    a. We don't anticipate needing to change this (apart from supporting concept IRIs via the "`urn:ietf:rfc:3987`" as an `Coding.system` value for IRIs)

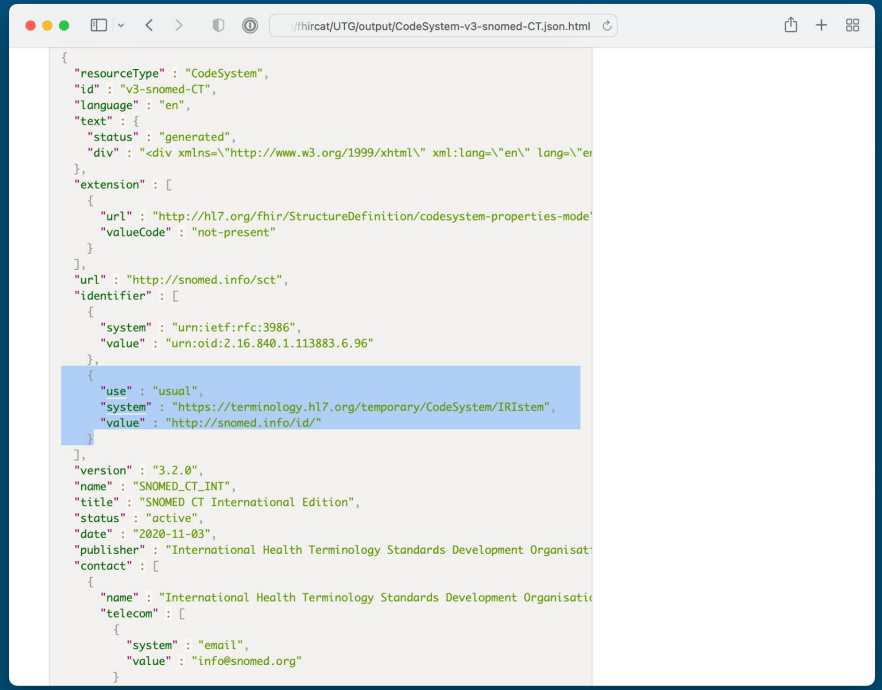
Credits

The RDF subgroup of the HL7 Implementable Technology Standards (ITS) working group

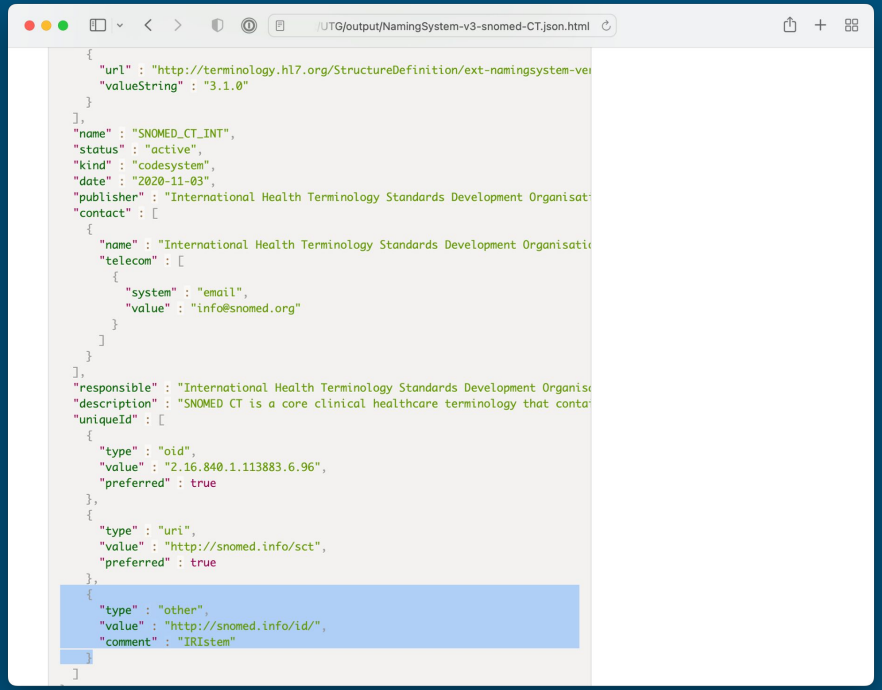
FHIRCat NIH grant #5-R01-EB030529-02

Additional slides

Would be accessible on terminology.hl7.org



CodeSystem (SNOMED)



NamingSystem (SNOMED)

IRIs vs URIs vs URLs

- RDF systems generally use *IRIs*, which allow you to include Unicode characters (allowing non-Latin characters). Defined by RFC 3987.
  - E.g. http://नारायणहृदयालय.example/code/हृदय
- All *IRIs* can be converted to *URIs*, which don't allow Unicode characters, but use percent-encoding to encode them. Defined by RFC 3986.
  - E.g. http://xn--g2bge2acacu4d4bbf6d.example/code/%E0%A4%B9%E0%A5%83%E0%A4%A6%E0%A4%AF
- URIs may be *URLs* (which are *resolvable*, like the link above) or *URNs* (which are "persistent, location-independent resource identifier"). Defined by RFC 8141.
  - URNs are in the format `urn:[namespace]:[identifier]`

IRIs vs URIs vs URLs

- We propose using IRIs, since they are what RDF systems are generally designed to use, and since FHIR already has good support for Unicode in its string type.
- These already work well with `Coding.code` values, which are defined as codes, which are strings restricted to having no leading or trailing whitespace and no internal whitespace apart from single spaces.
- If needed, concept IRIs could be converted into URIs and stored as uris or urls, but we don't think there is a need for concept IRIs to replace `Codings` in FHIR.

An algorithm for conversion

Given:

- a FHIR Coding.system, *s*, that identifies a terminology *t*; and
- a Coding.code, *c*, that is defined within *t*;

a Concept IRI, *conceptIRI*, corresponding to *s* and *c* is computed as follows:

1. If no IRI Stem is defined for *s* in the HL7 Terminology website, then *conceptIRI* is undefined. Halt.
2. Let *iStem* be an IRI Stem that is defined for *s* in the HL7 Terminology website.
3. As a special case, if *iStem* equals `urn:ietf:rfc:3987`, then *conceptIRI* is *c*, and *c* MUST be a syntactically valid absolute-IRI as defined by RFC 3987. Halt.
   (Non-normative comments: The purpose of this special case is to permit System.codes that are already IRIs to be used directly as Concept IRIs, without any transformation. Note that an absolute-IRI may also be a URL or a URN.)

An algorithm for conversion (2)

2. Let *cSafe* be the IRI-safe version of *c*, as defined by the algorithm in section 7.3 of R2RML: RDB to RDF Mapping Language (W3C Recommendation 27 September 2012), non-normatively quoted here for convenience:

> *"The **IRI-safe version** of a string is obtained by applying the following transformation to any character that is not in the iunreserved production in [RFC3987]:*
>
> *1. Convert the character to a sequence of one or more octets using UTF-8 [RFC3629]*
>
> *2. Percent-encode each octet [RFC3986]"*

An algorithm for conversion (3)

The `iunreserved` production defined in RFC 3987, section 2.2 using ABNF is also non-normatively quoted here for convenience:

```
iunreserved  = ALPHA / DIGIT / "-" / "." / "_" / "~" / ucschar
```

The ucschar production defined in RFC 3987, section 2.2 is also non-normatively quoted here for convenience. (*Non-normative comment: The ucschar production defines international character ranges that are valid unicode characters within the intersection of path components (ipath), query strings (iquery) and fragment identifiers (ifragment). They do not include any reserved characters involved in parsing apart the various components of an IRI.*)

```
ucschar     = %xA0-D7FF / %xF900-FDCF / %xFDF0-FFEF
              / %x10000-1FFFD / %x20000-2FFFD / %x30000-3FFFD
              / %x40000-4FFFD / %x50000-5FFFD / %x60000-6FFFD
              / %x70000-7FFFD / %x80000-8FFFD / %x90000-9FFFD
              / %xA0000-AFFFD / %xB0000-BFFFD / %xC0000-CFFFD
              / %xD0000-DFFFD / %xE1000-EFFFD
```

2. *conceptIRI* is the result of concatenating *iStem* and *cSafe*.

Security issues (1)

Hostile agents may try to introduce incorrect stem IRIs into the HL7 Terminology records, e.g. trying to set “`http://malicious.actor.org/`” as the stem IRI for SNOMED, so that cellulitis (128045006) would result in the concept IRI `http://malicious.actor.org/128045006`.

- 1. A healthcare application trying to dereference this URL could download malicious data to the application.

Security issues (2)

- 2. A healthcare application trying to dereference this URL could leak information to the malicious actor.
  - E.g. the malicious actor tricks a patient into accessing their healthcare records from a healthcare application at a particular time.
    - The application dereferences the concept IRI for all concepts related to their health information.
    - The malicious actor records can now deduce information about the patient.
  - Even if the previous attack does not work, the malicious actor could still perform a frequency analysis of the concepts being looked up within the healthcare application.
- 3. If some systems use the malicious IRI Stem and others use the correct IRI Stem, interoperability between those systems would be impaired, potentially resulting in denial of service.

What would this cost?

- Additional identifiers in HL7 Terminology records
  - Should be very stable in the long term.
  - Unlikely that users will confuse IRI stems with `Coding.system` values.
- *Possible security issues*

Who gets to pick IRI stems?

- Many vocabularies (SNOMED CT, LOINC) publish their own guidelines on how to create concept IRIs for their concepts.
  - All the vocabularies we've looked at use IRI stems, but someone might want to use more complex concept IRIs in the future.
- Extra property to indicate whether it's "official" from the organization or not?
- However, if concept IRIs are needed, several online databases of concept IRI patterns are available:
  - https://bioregistry.io/ (e.g. https://bioregistry.io/registry/loinc)
  - https://registry.identifiers.org (e.g. https://registry.identifiers.org/registry/uberon)
  - Wikidata (e.g. https://www.wikidata.org/wiki/Property:P5806)

Choosing IRI stems for HL7 Terminologies

- FHIR CodeSystems (https://terminology.hl7.org/codesystems.html)
  - (Check working build, there are some new changes being made to this list)
- External terminologies (https://terminology.hl7.org/external_terminologies.html)
  - 1. Official IRI stem, if one exists.
  - 2. Look through repositories to find any in community use:
    - https://bioregistry.io/ (e.g. https://bioregistry.io/registry/loinc)
    - https://registry.identifiers.org (e.g. https://registry.identifiers.org/registry/uberon)
    - Wikidata (e.g. https://www.wikidata.org/wiki/Property:P5806)
  - 3. Assign an HL7 IRI stem in the hl7.org domain/namespace
    - If the organization comes up with one in the future, we can change the FHIR IRI stem to that.
- FHIR RDF group can be responsible for coming up with IRI stems for all FHIR CodeSystems and external terminologies.
- What if branding changes in the future and an organization wants to change their concept IRIs in the future?

Concept IRIs and versioning?

- Each FHIR `Coding` can specify the version of the terminology from which the term was taken in the `Coding.version` field.
- Concept IRIs generally don't require versions, since obsolete terms are supposed to be marked as deprecated while a new term is created for the updated concept (see e.g. OBO Foundry's URI automated check).
  - E.g. UBERON:0006852 ("obsolete glomerular visceral epithelium") has been deprecated and replaced with UBERON:0005751 ("glomerular visceral epithelium").
- If a use-case for versions becomes necessary, we would want to add an "IRI stem template" that allows both `Coding.code` and `Coding.version` to be inserted separately.

Examples

Coding.system	Coding.code	IRI Stem	Concept IRI
ICD 10: http://hl7.org/fhir/sid/icd-10	G44.1	http://purl.bioontology.org/ontology/ICD10/	http://purl.bioontology.org/ontology/ICD10/G44.1
SNOMED CT: http://snomed.info/sct	128045006	http://snomed.info/id/	http://snomed.info/id/128045006
MeSH: https://www.nlm.nih.gov/mesh	D000305	https://id.nlm.nih.gov/mesh/	https://id.nlm.nih.gov/mesh/D000305
LOINC: http://loinc.org	35217-9	https://loinc.org/rdf/	https://loinc.org/rdf/35217-9
Regional example			
Example of a code containing non-ASCII characters http://नारायणहृदयालय.example/	हृदय	http://नारायणहृदयालय.example/code/	http://नारायणहृदयालय.example/code/हृदय