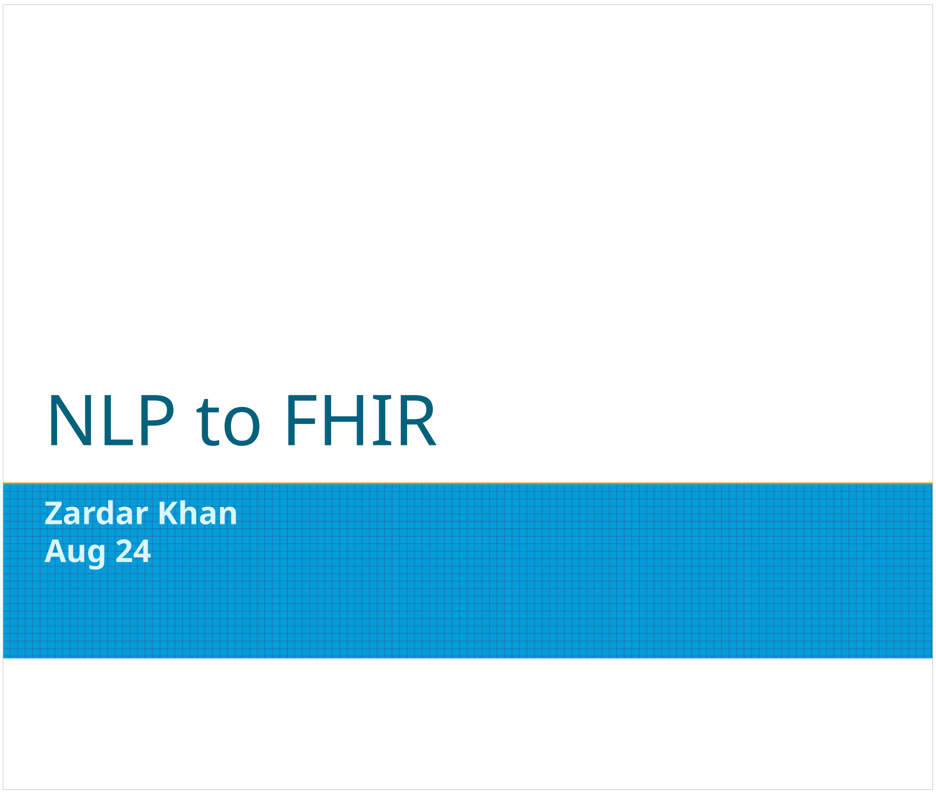
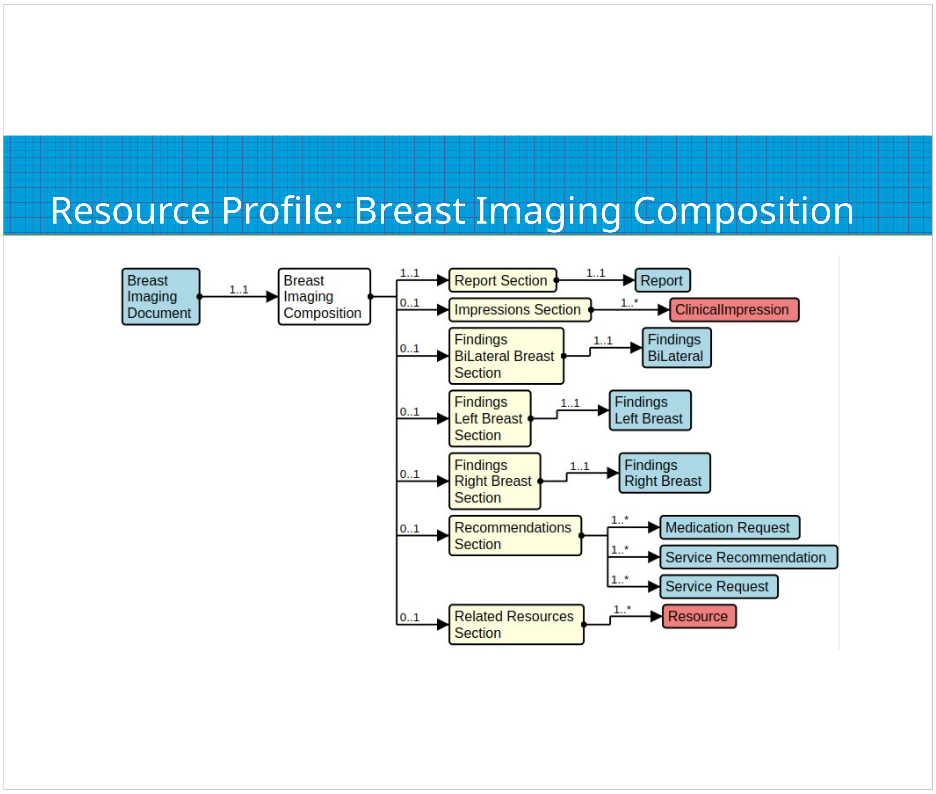


NLP to FHIR

Zardar Khan
Aug 24

- We have a class of NLP models which can classify diagnostic reports in various categories. For clarity in this discussion, I will refer to a singular model for classifying patient's current microcalcification status into Positive, Negative or Not Stated.
- For my own experimentation, I have code that given a FHIR ID for a Diagnostic Report will return the free text report.
- There are some basic pre-processing steps in which we remove excessive spacing in the free text report
- We take the free text and feed it into our NLP model which returns one of three labels; Positive, Negative, Not Stated
- For intraoperability, we need two things; a commonly used vocabulary and some agreed upon format for storing the NLP derived data

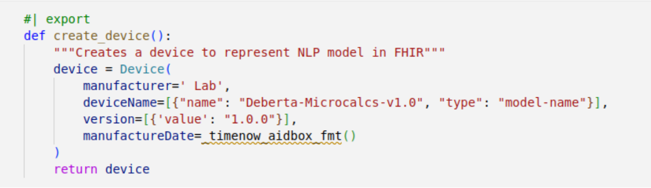


Resource Profile: Breast Imaging Composition

I've used the fhir.resources library to represent these FHIR resource in version R4 and extended it by implementing functions for creating resources based on the IG. Now all this needs to be created in order, so we can have the FHIR IDs to reference in the parent resource. I haven't had a chance to complete nor test my code as there's a difference between data in PROD/STAGING vs DEV. I only read from PROD/STAGING and have only tested writing to DEV.

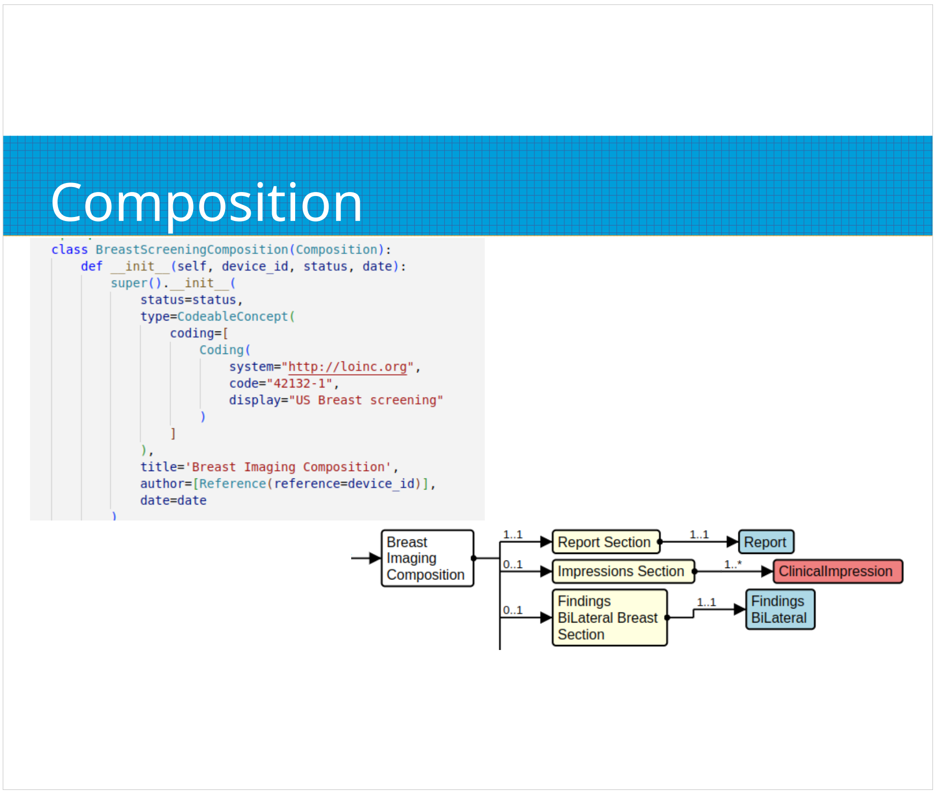
All validations have passed validator step. This is what I've done so far, I don't know if my approach via creating FHIR objects is complete different direction than how

Device Resource



```python
#| export
def create_device():
    """Creates a device to represent NLP model in FHIR"""
    device = Device(
        manufacturer=' Lab',
        deviceName=[{"name": "Deberta-Microcalcs-v1.0", "type": "model-name"}],
        version=[{'value': "1.0.0"}],
        manufactureDate=_timenow_aidbox_fmt()
    )
    return device
```

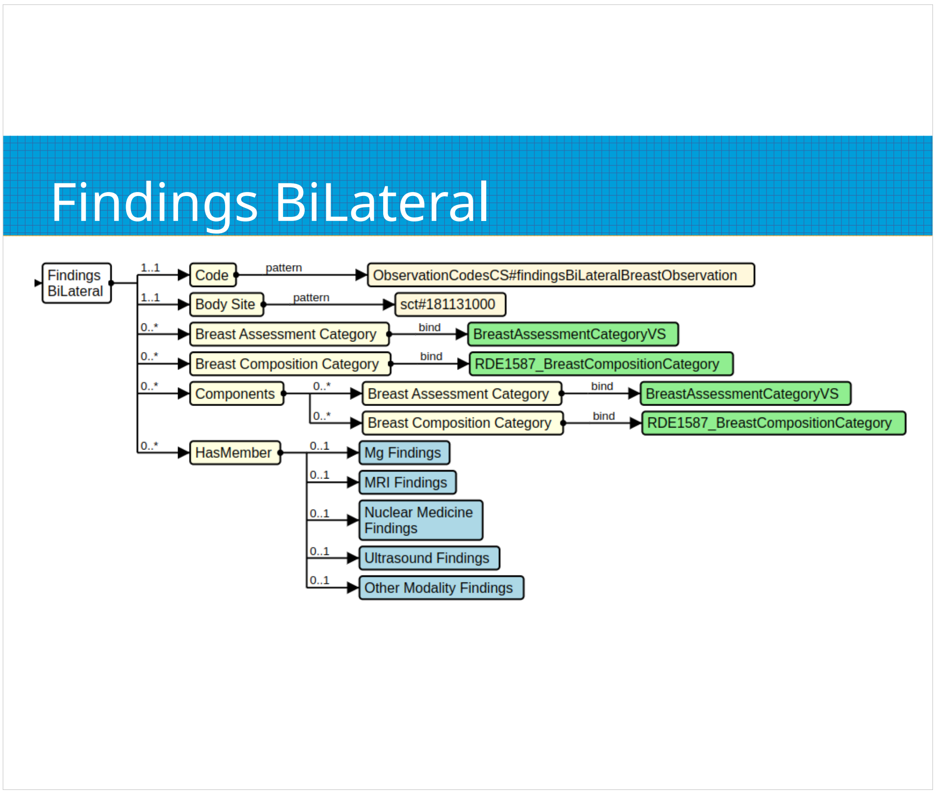
- We start off with a Device resource; this allows us to represent our NLP model in the FHIR



Composition

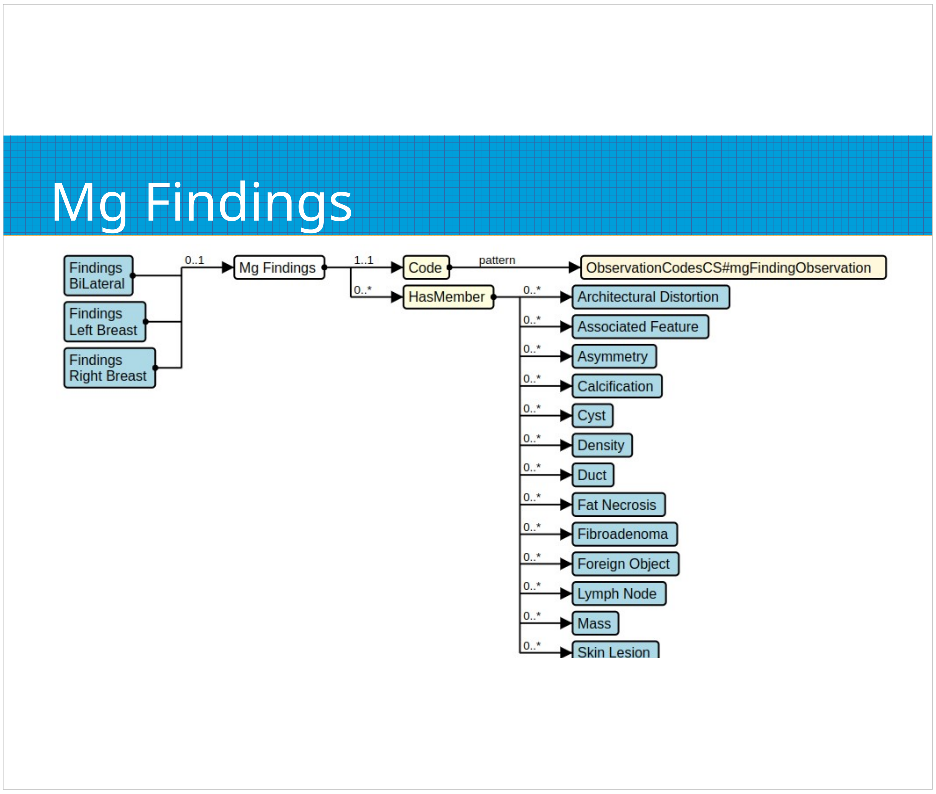
```
class BreastScreeningComposition(Composition):
    def __init__(self, device_id, status, date):
        super().__init__(
            status=status,
            type=CodeableConcept(
                coding=[
                    Coding(
                        system="http://loinc.org",
                        code="42132-1",
                        display="US Breast screening"
                    )
                ]
            ),
            title='Breast Imaging Composition',
            author=[Reference(reference=device_id)],
            date=date
        )
```

- 2nd; we have a Composition resource; by default a Composition must have a type, title, author (the NLP device). The Composition for our usage; has a Report Section and Findings BiLateral Breast Section. These are of type `Composition.section`



Findings BiLateral

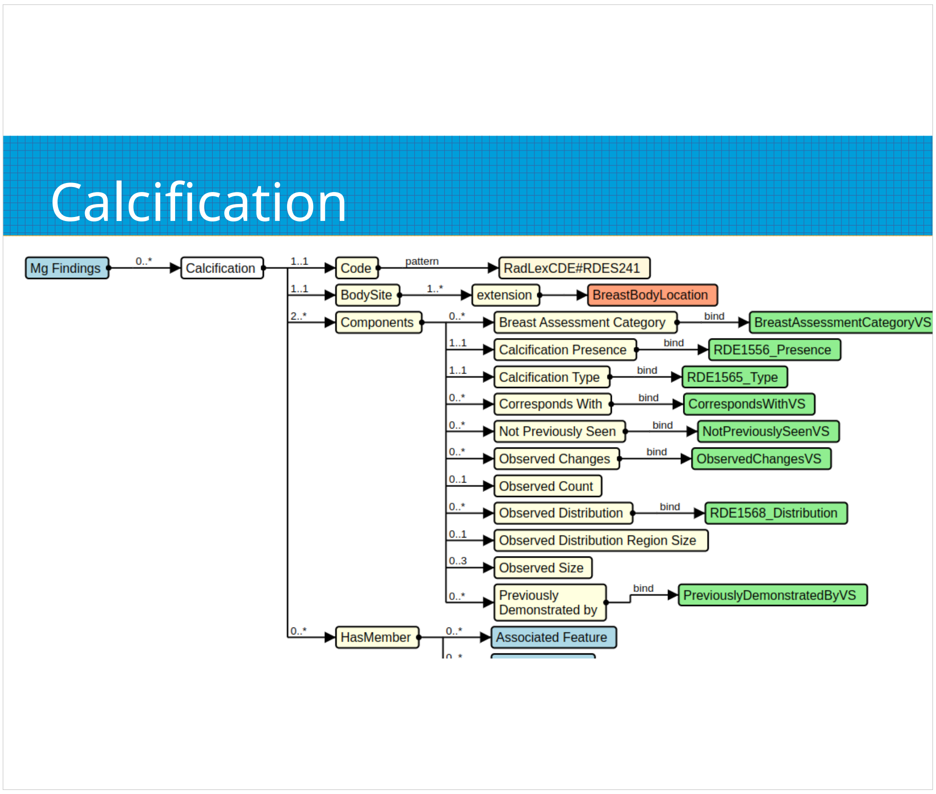
- If we zoom into Findings BiLateral, which is of type Observation, we have the following structure. For our current use case of microcalcification and we'll use Mammography as the source imaging modality for the report, we have to jump to `hasMember -> Mg Findings`.



Mg Findings

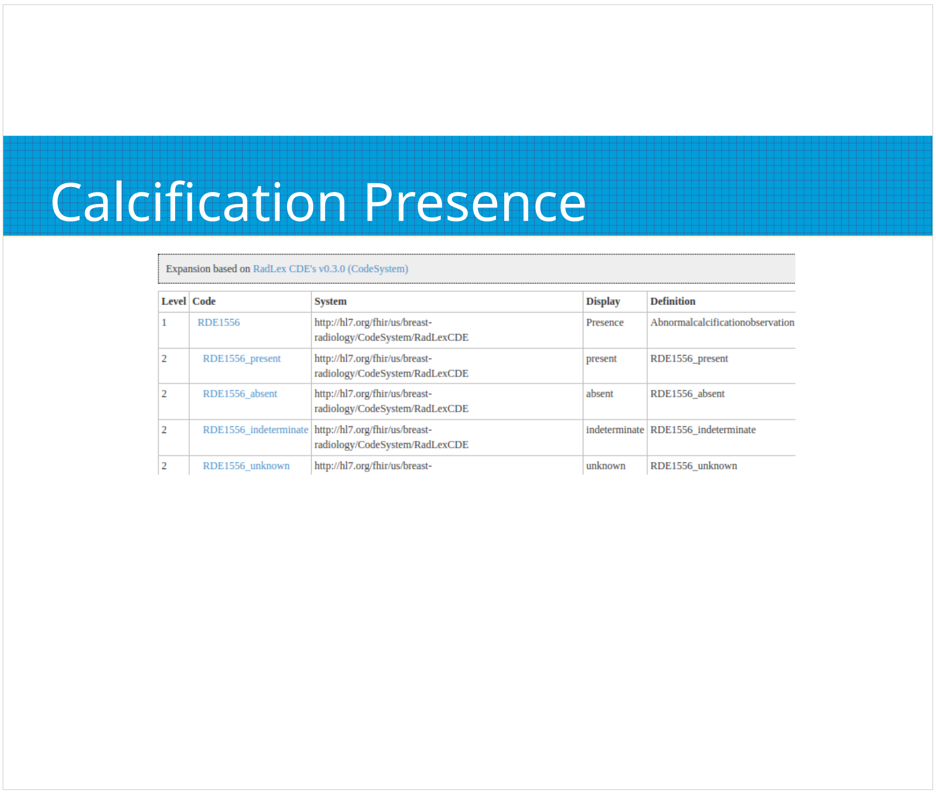
Findings BiLateral
Findings Left Breast
Findings Right Breast
0..1
Mg Findings
1..1
Code
pattern
ObservationCodesCS#mgFindingObservation
0..*
HasMember
0..* Architectural Distortion
0..* Associated Feature
0..* Asymmetry
0..* Calcification
0..* Cyst
0..* Density
0..* Duct
0..* Fat Necrosis
0..* Fibroadenoma
0..* Foreign Object
0..* Lymph Node
0..* Mass
0..* Skin Lesion

Mg Findings is also of type Observation



Calcification

- Inside MgFindings we have Calcification, which is also of type Observation. Here we have a Component `Calcification Presence`. This is where we save the mapped NLP label.



Calcification Presence

Expansion based on RadLex CDE's v0.3.0 (CodeSystem)

Level	Code	System	Display	Definition
1	RDE1556	http://hl7.org/fhir/us/breast-radiology/CodeSystem/RadLexCDE	Presence	Abnormalcalcificationobservation
2	RDE1556_present	http://hl7.org/fhir/us/breast-radiology/CodeSystem/RadLexCDE	present	RDE1556_present
2	RDE1556_absent	http://hl7.org/fhir/us/breast-radiology/CodeSystem/RadLexCDE	absent	RDE1556_absent
2	RDE1556_indeterminate	http://hl7.org/fhir/us/breast-radiology/CodeSystem/RadLexCDE	indeterminate	RDE1556_indeterminate
2	RDE1556_unknown	http://hl7.org/fhir/us/breast-	unknown	RDE1556_unknown

The agreed upon terminology using RadLex however for our OMOP work, we need to use the SNOMED terminology as well. We'll call this element `calcification_abnormality`.

Creating Mg Findings

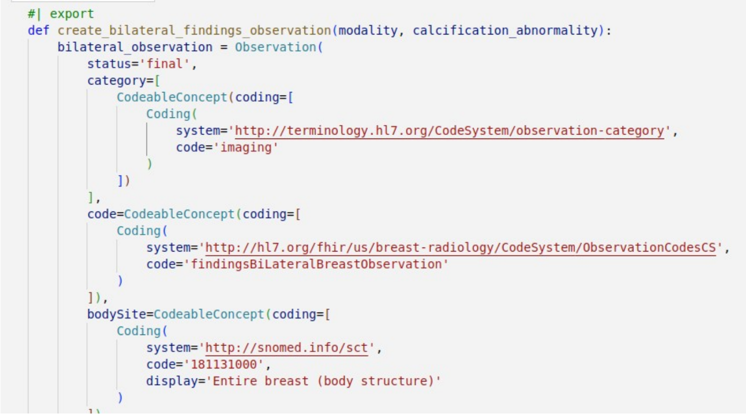
```python
def create_mg_findings_observations(calcification_abnormality):
    mg_findings = Observation(
        status='final',
        category=[CodeableConcept(coding=[
            Coding(system='http://terminology.hl7.org/CodeSystem/observation-category',
                   code='imaging')
        ])],
        code=CodeableConcept(coding=[
            Coding(
                system='http://hl7.org/fhir/us/breast-radiology/CodeSystem/ObservationCodesCS',
                code='mgFindingObservation'
            )
        ]),
    )

    mg_findings.hasMember = [Reference(
        reference=f"Observation"/{calcification_abnormality.id},
        type='Observation'
    )]

    return mg_findings
```

- Once, calcification abnormality is created, we can create a Mg Finding Observation which references the calcification abnormality.id

Creating BiLateral Findings



```python
#| export
def create_bilateral_findings_observation(modality, calcification_abnormality):
    bilateral_observation = Observation(
        status='final',
        category=[
            CodeableConcept(coding=[
                Coding(
                    system='http://terminology.hl7.org/CodeSystem/observation-category',
                    code='imaging'
                )
            ])
        ],
        code=CodeableConcept(coding=[
            Coding(
                system='http://hl7.org/fhir/us/breast-radiology/CodeSystem/ObservationCodesCS',
                code='findingsBiLateralBreastObservation'
            )
        ]),
        bodySite=CodeableConcept(coding=[
            Coding(
                system='http://snomed.info/sct',
                code='181131000',
                display='Entire breast (body structure)'
            )
        ])
```

- Next we create our BiLateral Findings Observation which references the ID for our Mg Finding Observation.