Analytics using SNOMED CT: results from a genomics pilot

Alejandro Metke

October 2019

AUSTRALIAN E-HEALTH RESEARCH CENTRE
www.csiro.au

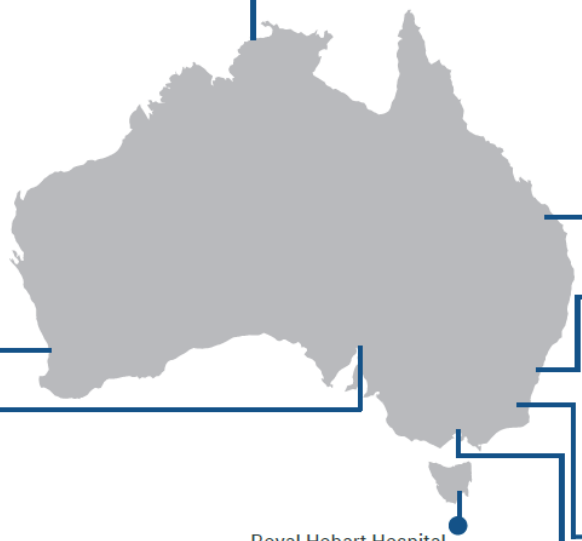
About me

- Australian e-Health Research Centre
  - Largest e-Health program in Australia
  - Aproximately 100 researchers
  - Health Informatics
    - Health Data Interoperability
  - Involved in several genomics initiatives
    - Melbourne, Australian, Queensland Genomics
    - Global Alliance for Genomics and Health

Australian Genomics

BUILT ON NATIONAL COLLABORATION (80 PARTNERS)

Australian Genomics Health Alliance



Harry Perkins Institute of Medical Research
Path West
Genetic Services of Western Australia
Princess Margaret Hospital
Telethon Kids Institute
Royal Perth Hospital
Sir Charles Gairdner Hospital
The University of Western Australia

South Australian Health & Medical Research Institute
SA Pathology
Centre for Cancer Biology
Women's and Children's Hospital
The University of Adelaide
University of South Australia
Royal Adelaide Hospital

The University of Queensland
Lady Cilento Children's Hospital
Institute for Molecular Bioscience
QIMR Berghofer Medical Research Institute
Wesley Hospital
Royal Brisbane and Women's Hospital/Genetic Health Queensland
Princess Alexandra Hospital
Diamantina Institute
Pathology Queensland
Queensland University of Technology
Queensland Genomics Health Alliance

Sydney Childrens Hospitals Network
Royal North Shore Hospital
Garvan Institute of Medical Research & KCCG
Kinghorn Cancer Centre
Genome.One
NSW Health Pathology
Children's Cancer Institute Australia
The University of Sydney
Children's Medical Research Institute
University of New South Wales
Centre for Genetics Education
Macquarie University/AIHI

The Australian National University

Murdoch Children's Research Institute
Melbourne Bioinformatics
Victorian Clinical Genetics Services
Melbourne Health / Royal Melbourne Hospital
The University of Melbourne
Walter and Eliza Hall Institute of Medical Research
Peter MacCallum Cancer Centre
The Royal Children's Hospital
Austin Health

Melbourne Genomics Health Alliance
Monash Health
Victorian Comprehensive Cancer Centre
Florey Institute of Neuroscience and Mental Health

Peak Professional Bodies

Royal College of Pathologists of Australasia
Human Genetics Society of Australasia

National Partners

Bioplatforms Australia
Australian Genome Research Facility
BioGrid Australia
National Computational Infrastructure
CSIRO
Rare Voices Australia
Rare Cancers Australia
Australian Mitochondrial Disease Foundation

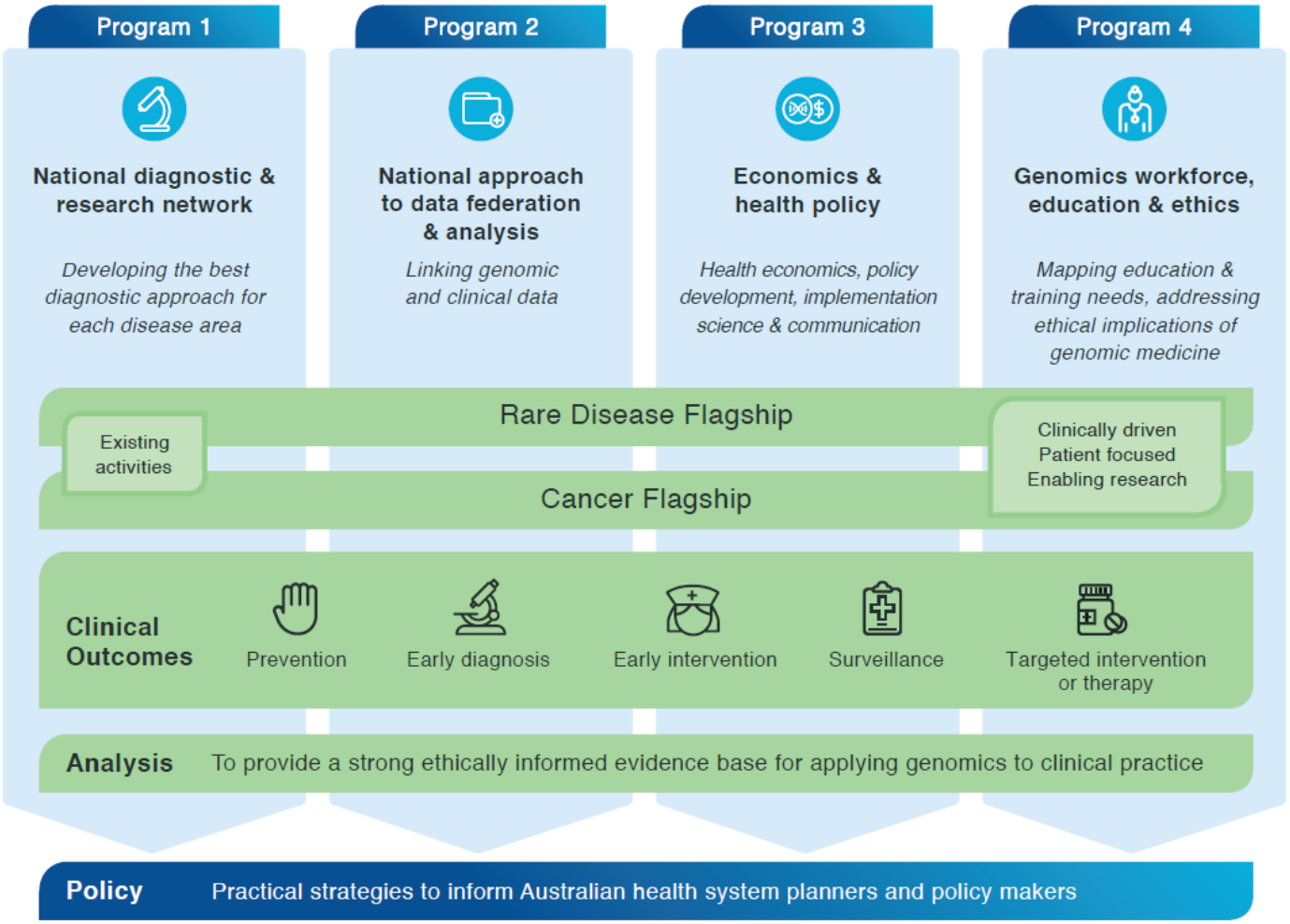
International Partners

Broad Institute of MIT and Harvard
Baylor College of Medicine
UCL Great Ormond Street Institute of Child Health
Global Alliance for Genomics and Health
Global Genomic Medicine Collaborative
Genomics England

CSIRO

Australian Genomics

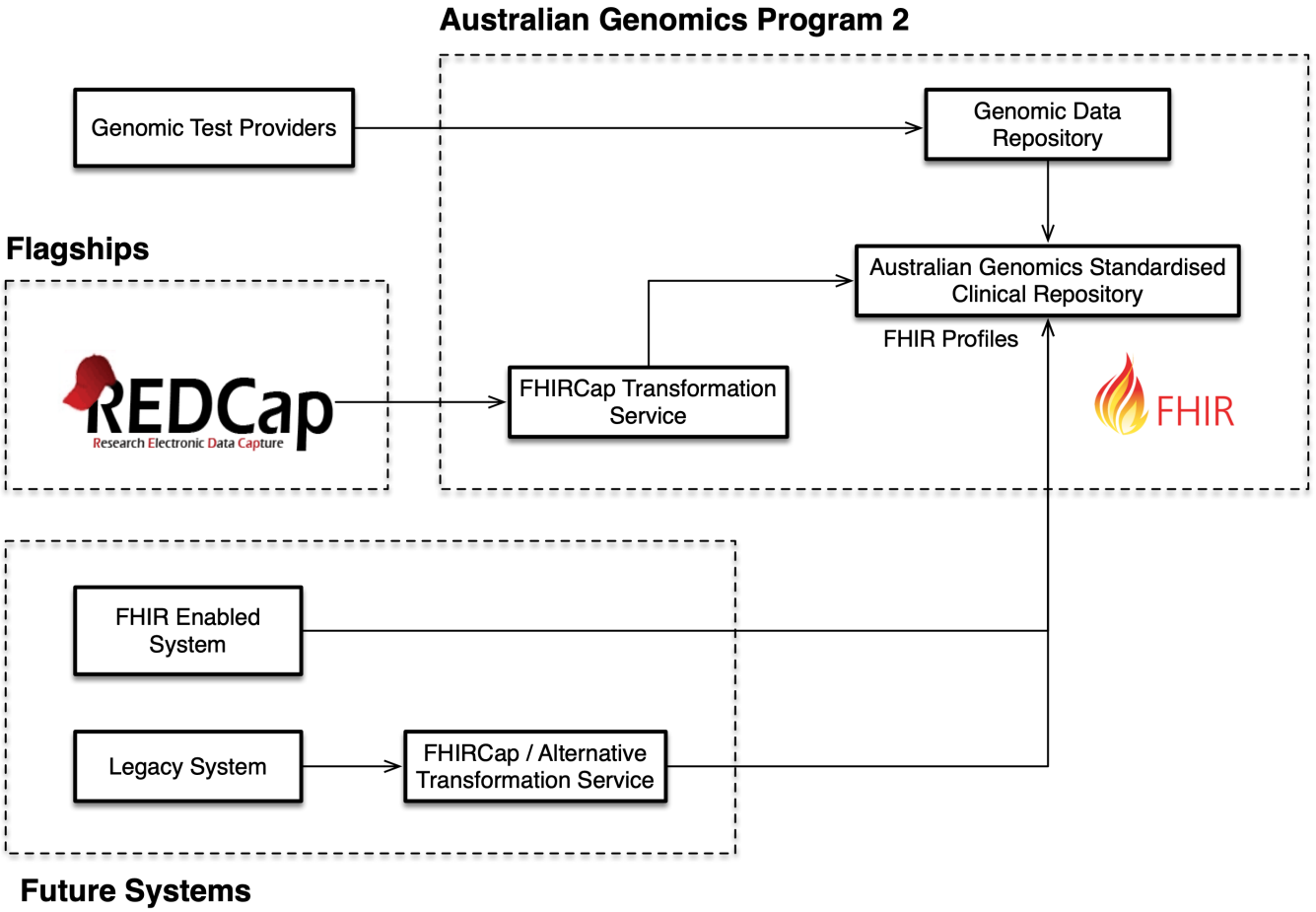
PROGRAMS, FLAGSHIPS AND PROJECTS



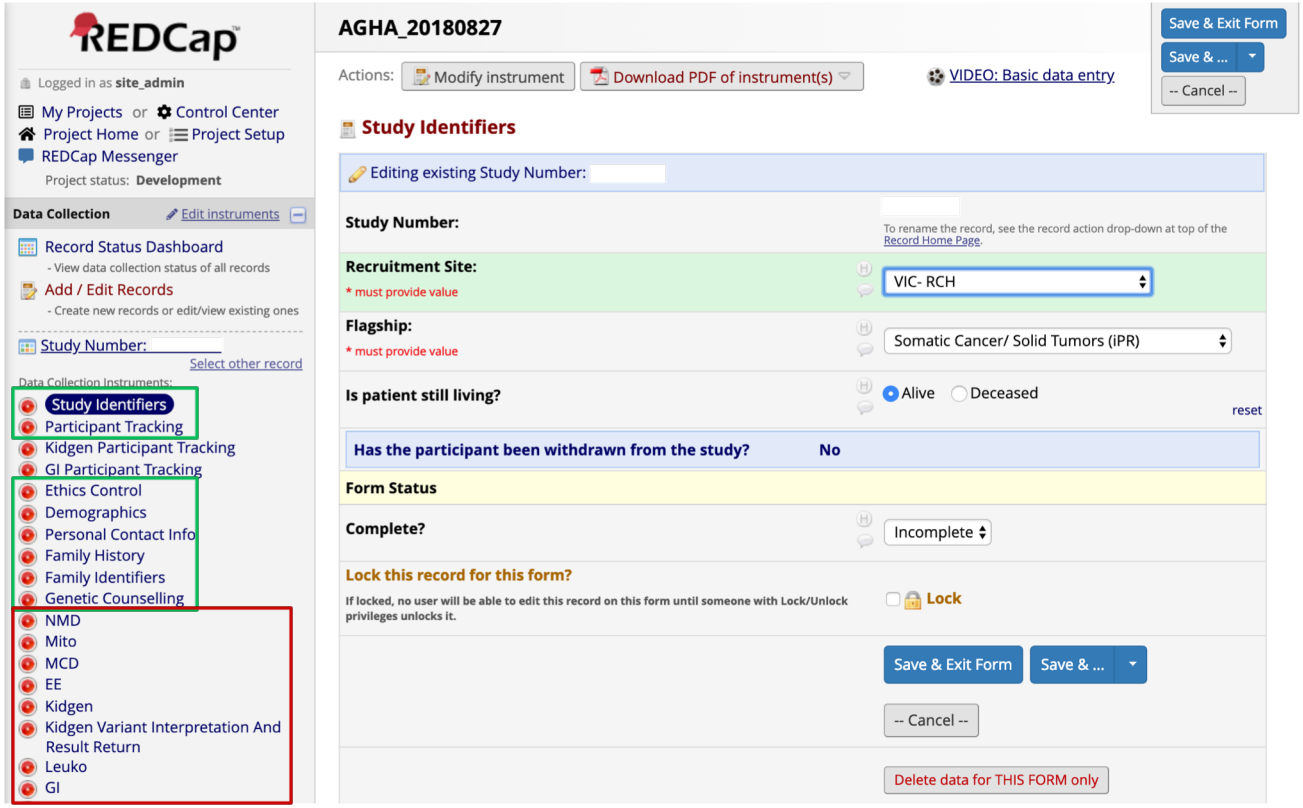
A National Gen-Phen Database

- A national approach for data federation and analysis
  - Linking genomic and clinical data
    - Standardised
    - Secondary use
    - De-identified, pseudonymised
  - Initial sources are Flagships - most use REDCap
  - Will need to support other sources

A National Gen-Phen Database

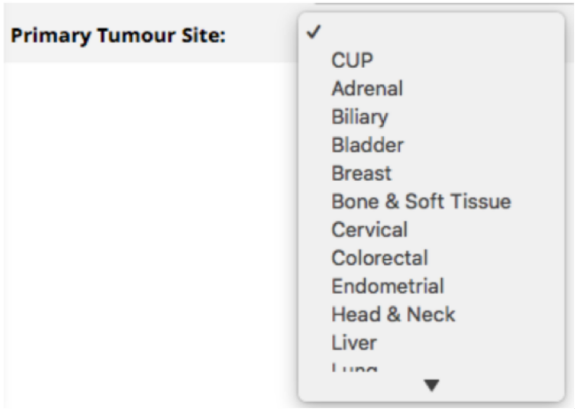


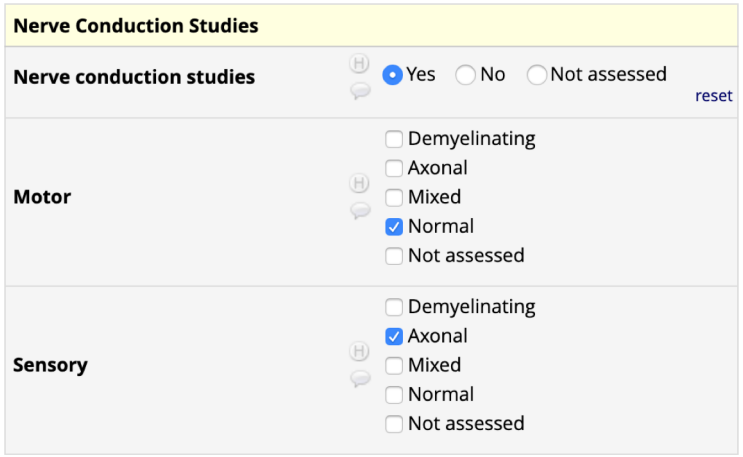
REDCap in the Flagships



REDCap in the Flagships

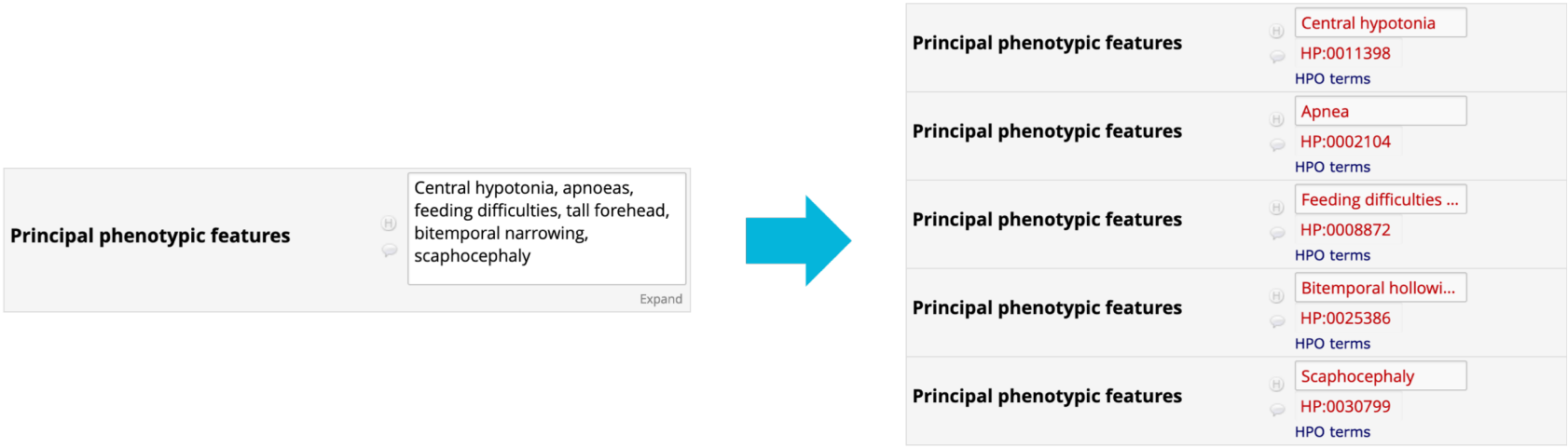
- Clinical data capture is not standardised





REDCap in the Flagships

- REDCap FHIR Terminology Plugin



Get the REDCap plugin here:
https://github.com/aehrc/redcap_fhir_ontology_provider

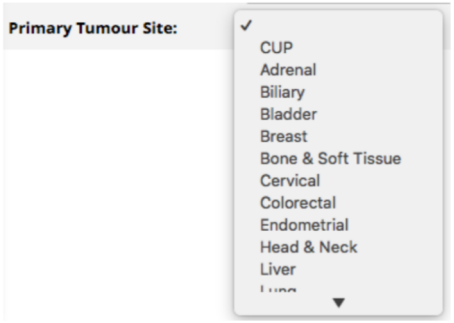
Try Ontoserver here: http://genomics.ontoserver.csiro.au

FHIRCap

- Rules-based transformation engine
  - Usable by non-programmers (but still need to understand FHIR!)
  - Rule structure

```
[condition] :
   [resource_type]<[resource_id]> ->
     ([attribute] = [value])*
```

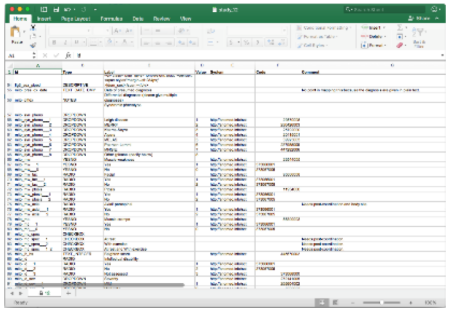
FHIRCap Basics



Rules:

```
VALUE (pts) != null :
  Condition<c1> ->
    bodySite[0] = CONCEPT_SELECTED (ts),
    code = CONCEPT_VALUE (
      http://snomed.info/sct|108369006
    );
```

Mappings:

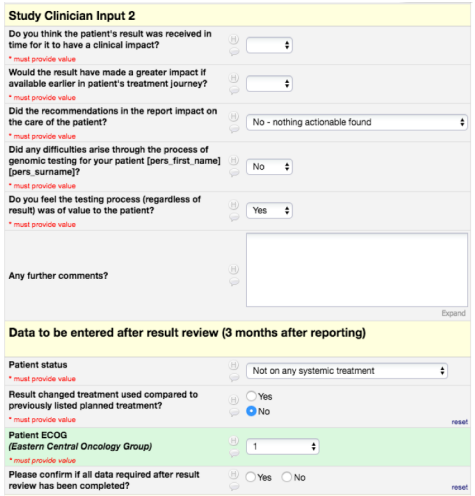


FHIR

CSIRO

FHIRCap Transformation Example

Somatic Cancer Flagship REDCap Forms

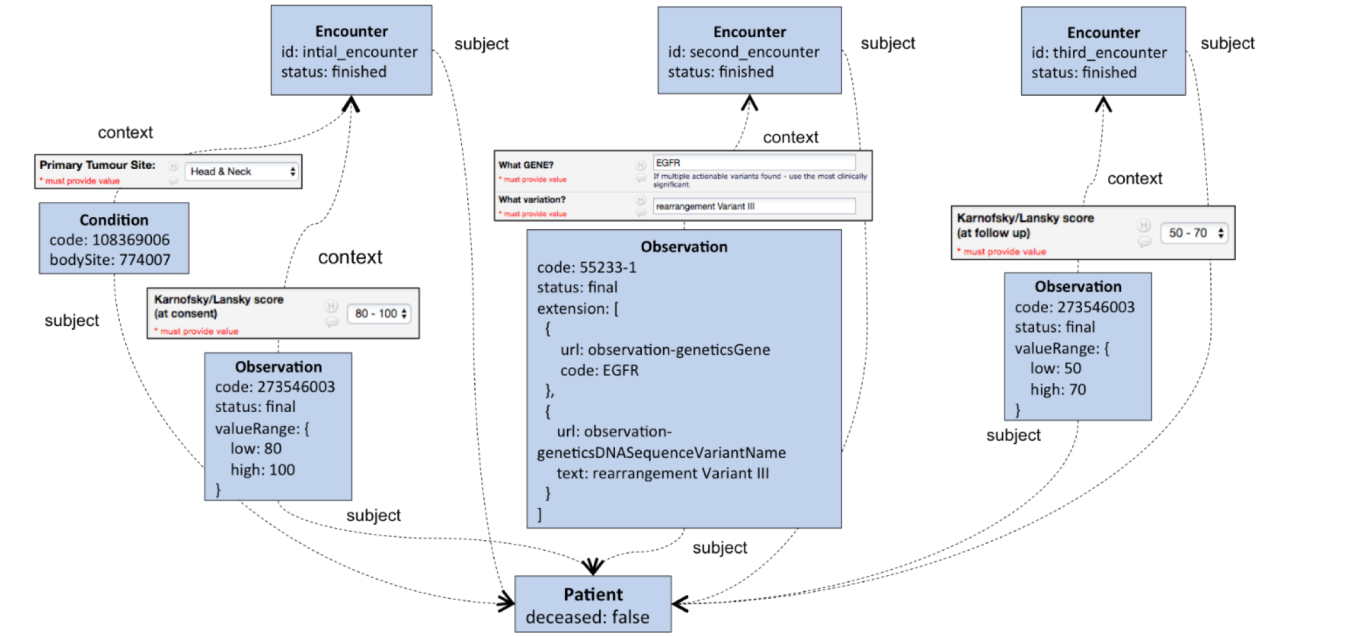


FHIR Representation using SNOMED CT

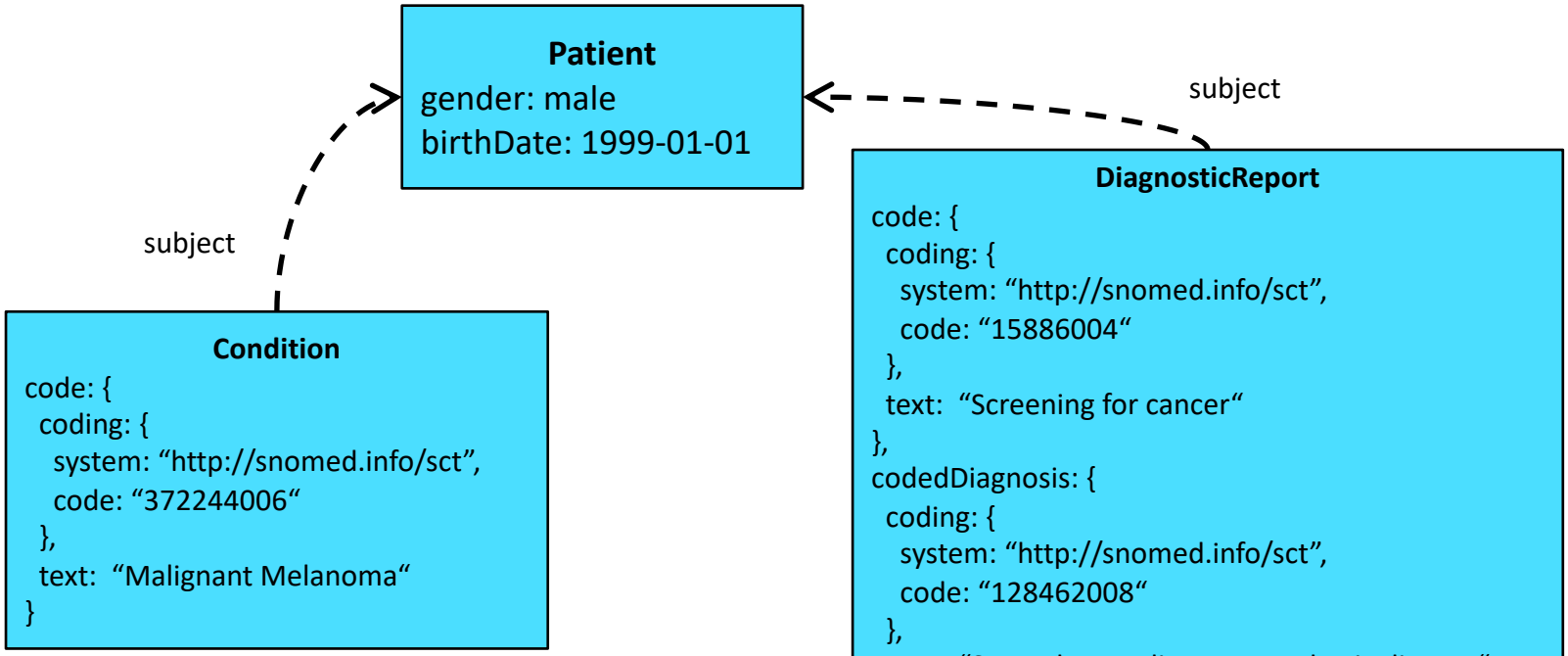
Initial encounter, medical history captured, consent obtained, genetic test ordered

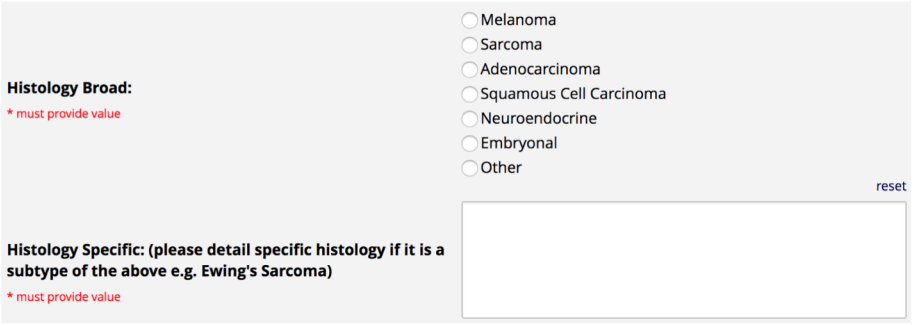
Second encounter, test results interpreted, treatment potentially adjusted

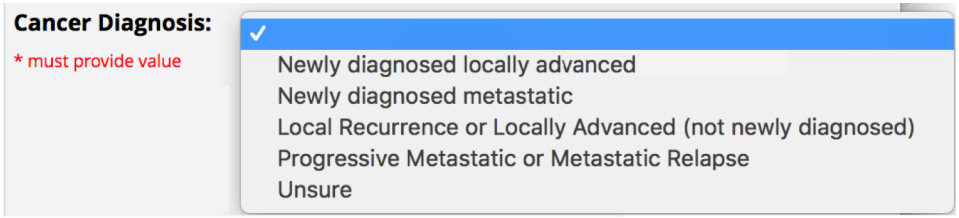
Third encounter, follow up, genetic test value assesed



Examples of Mapping Issues







Free text might not always map to SCT

Some detail is lost in the mapping

Kidgen Flagship Pilot

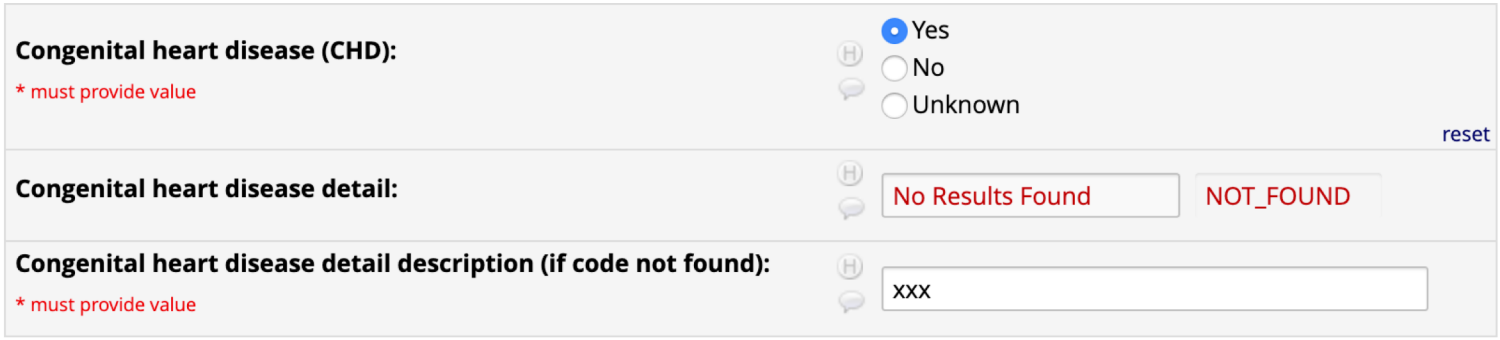
- Renal Genetics Flagship
  - Integrates genomic medicine into care for those with inherited kidney diseases, with the aim to better diagnose, manage and treat these diseases
- The project has two main goals:
  - Improving the quality of the data that has been collected
  - Exploring analytics use cases using SNOMED CT and FHIR

Phase 1 - Data Capture

Referral diagnosis (revised)	unclear, Biopsy suggestive of inherited GBM abnormality vs clinical dx of SRNS Expand
Clinical diagnosis (revised)	Alport Expand
Final genetic diagnosis (revised) (Kush to input final diagnosis based on WES report):	autosomal recessive Alport syndrome Expand

Phase 1 - Data Capture

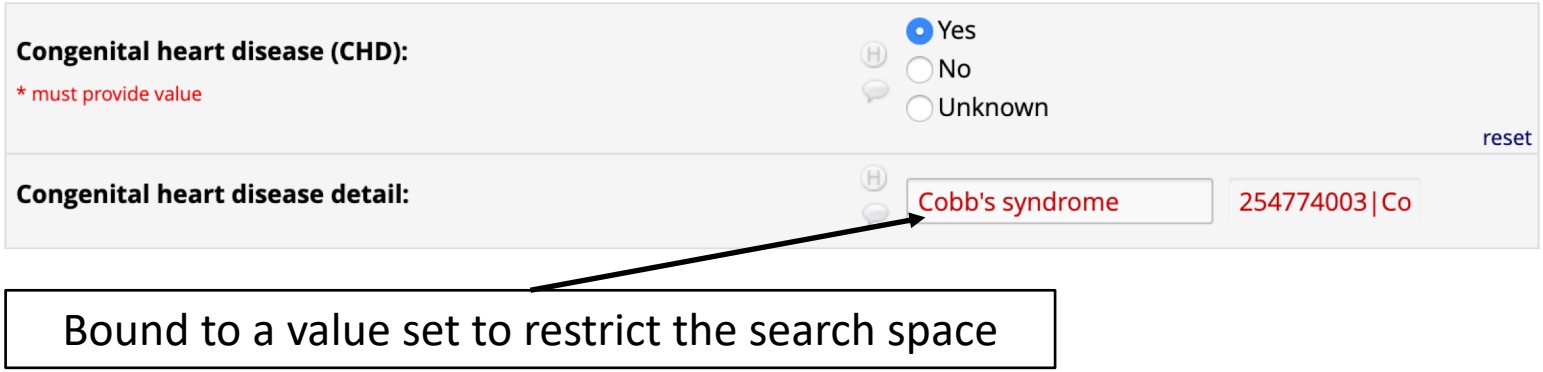
- Pattern to capture coded data



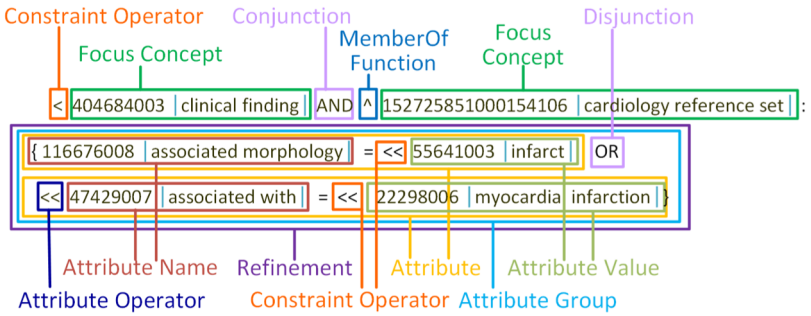
- It is still possible to enter free text if a concept is not in SNOMED CT
  - But, we want to encourage users to search for a concept
  - Free text should only be used if the concept does not exist or doesn't have good descriptions
  - The UI is designed to allow free text entry only if the user searches for a code and states that it couldn't be found

Phase 1 – Data Capture

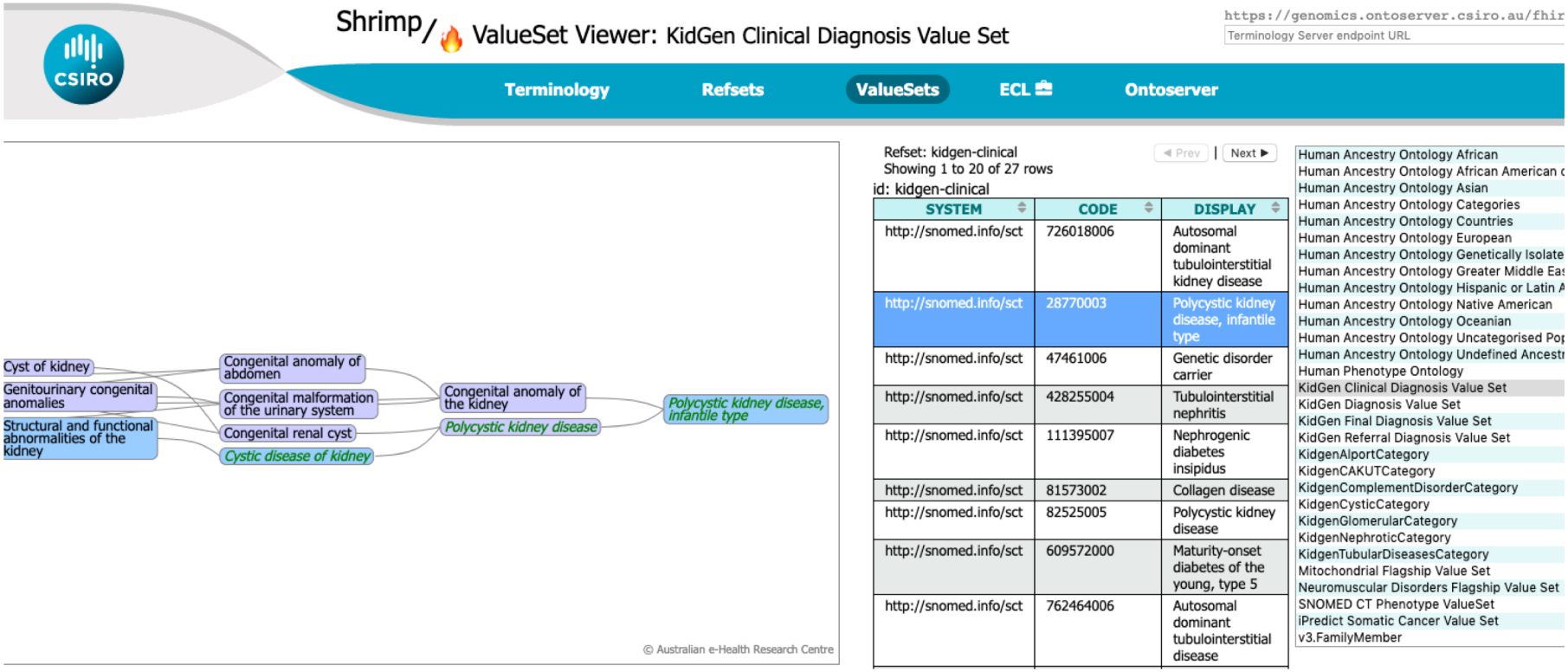
- Pattern to capture coded data



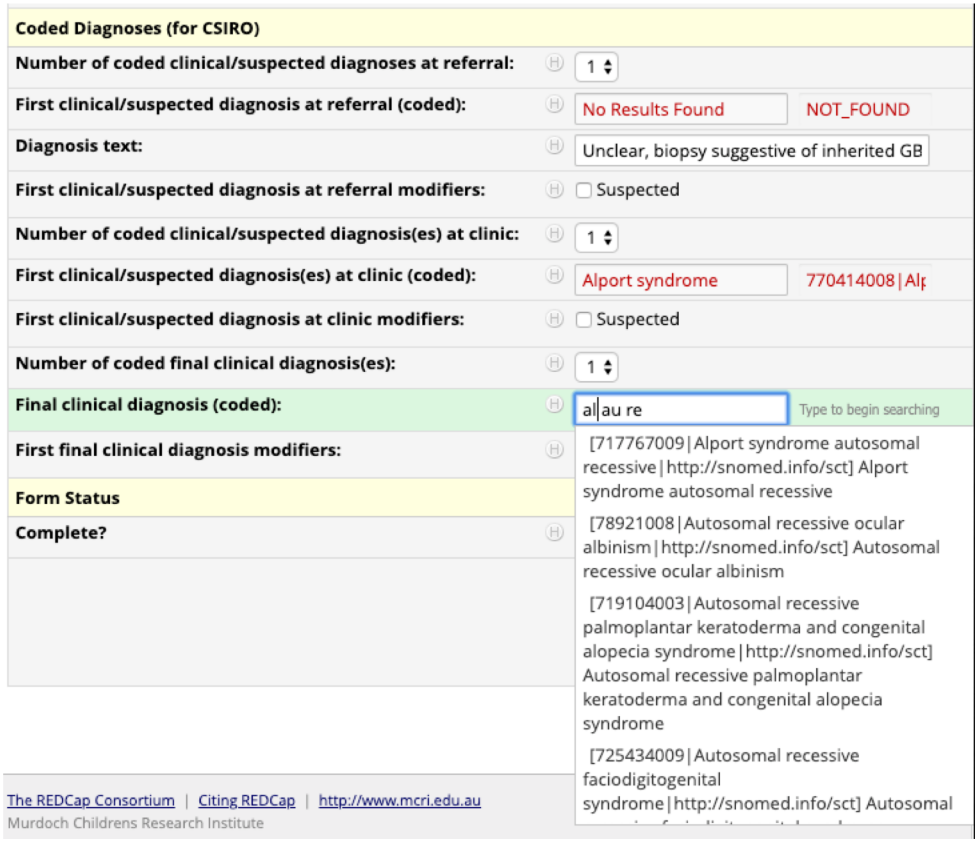
- Lots of different ways to create value sets
- SNOMED CT can use the Expression Constraint Language, which is really powerful



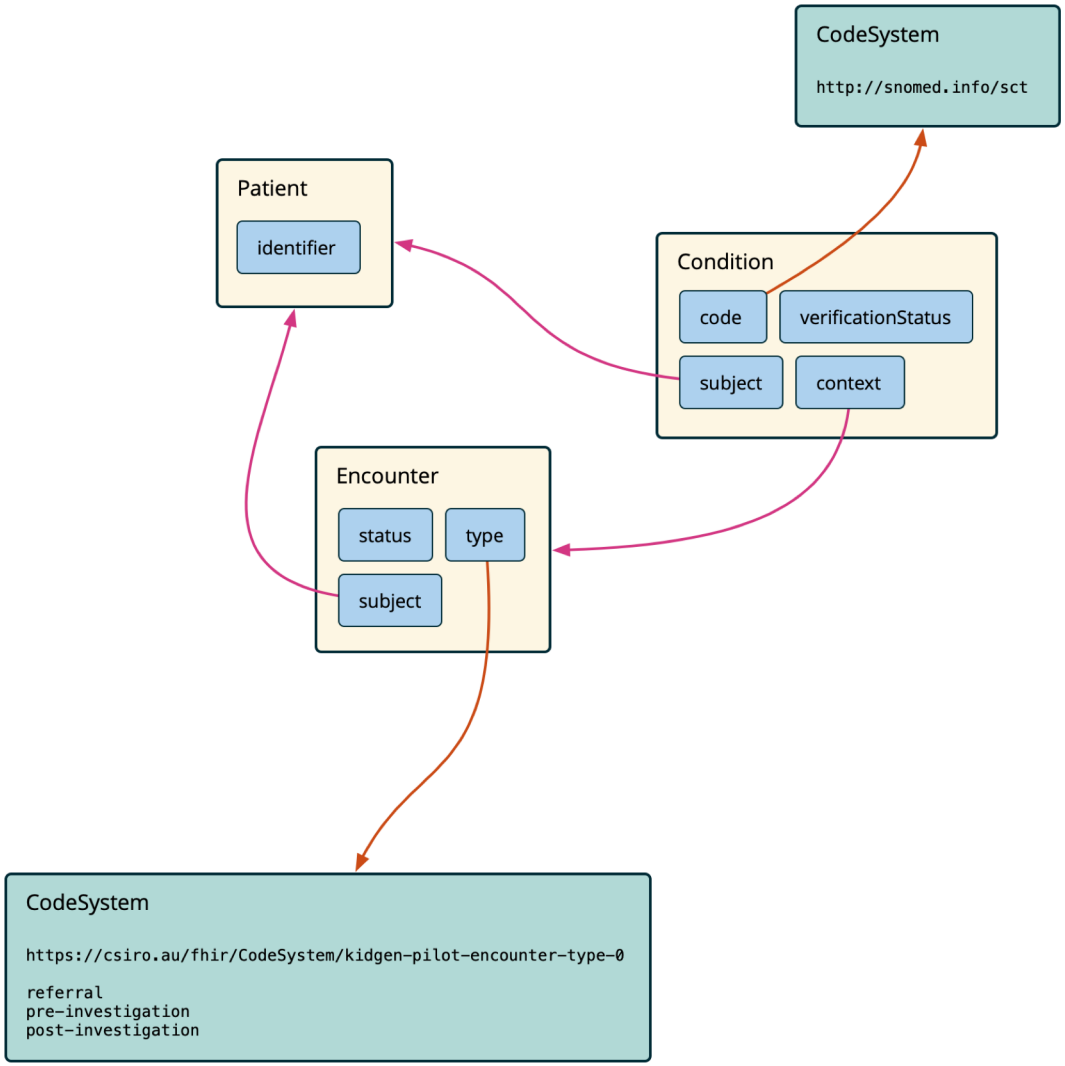
Phase 1 - Data Capture



Phase 1 - Data Capture



Phase 1 – Transformation to FHIR



Phase 1 – Modelling Recommendations

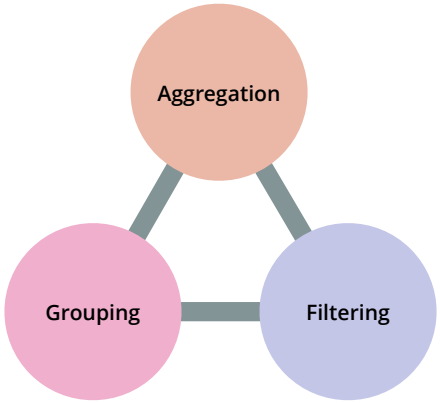
- 16 modelling issues were reported
- Issues included:
  - Incorrect modelling, e.g., concepts in the wrong hierarchy
  - Incomplete modelling, e.g., concepts missing relationships
  - Missing concepts
  - Missing synonyms

Phase 2 - Analytics

- A patient attending a Kidgen renal clinic gets diagnosed multiple times:
  - At referral: preliminary diagnosis made by the referrer.
  - Before genomic testing is performed: multi-disciplinary team reviews preliminary diagnosis with all information available.
  - After genomic testing: multi-disciplinary team reviews diagnosis taking into account the results of genomic testing.
- Interested in understanding how diagnoses evolve during this journey
  - Important to calculate diagnostic yield
  - Also helps evaluate the multi-disciplinary clinic model

Interactive analytics, delivered via REST API

- Users want to access real-time analytics interactively, via web applications and mobile devices
- Interacting with data warehouses directly via SQL is difficult to secure and imposes a high-level of complexity on clients
- Users need an integrated solution for querying clinical records and the terminology within them
- Solution: provide a FHIR-based REST API that can satisfy aggregate query requests, and also drill-down to individual records where the user is authorised



Number of patients

grouped by

Patient has a recorded diagnosis that is a type of lung infection

where

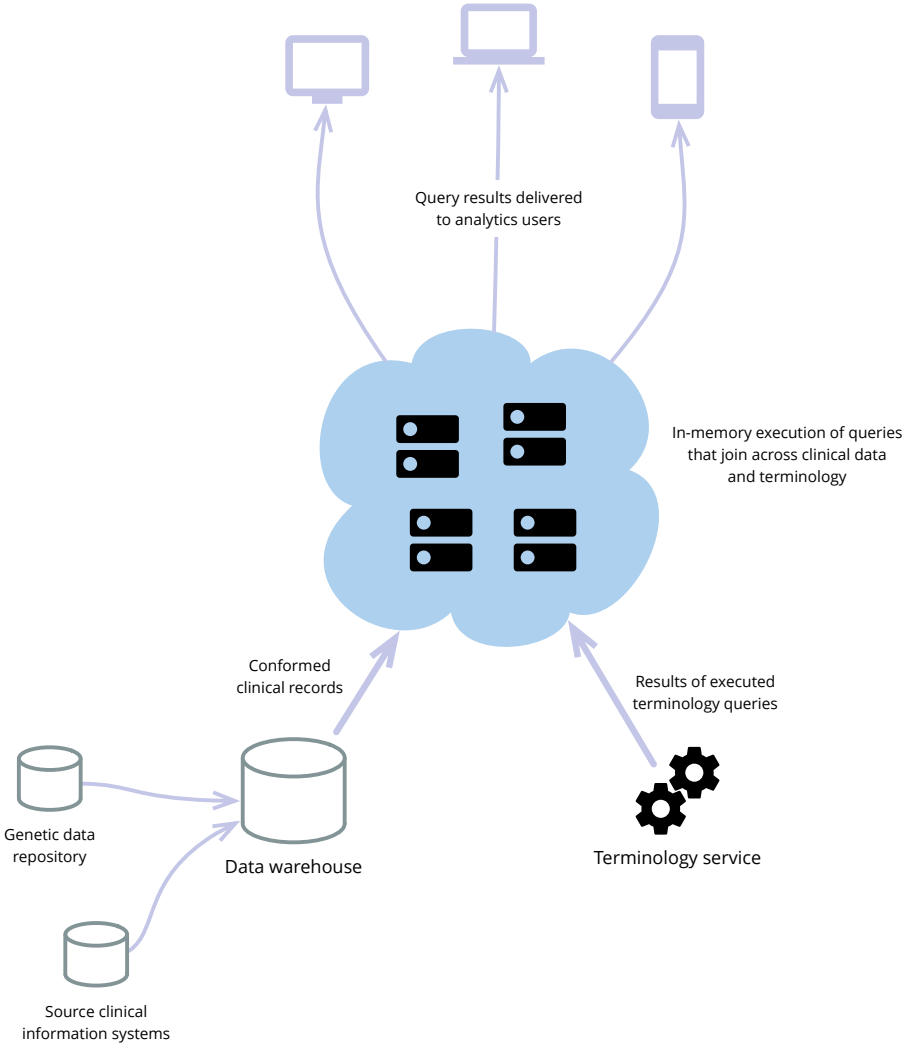
Patient has recorded diagnosis that is a type of autoimmune arthritis

and where

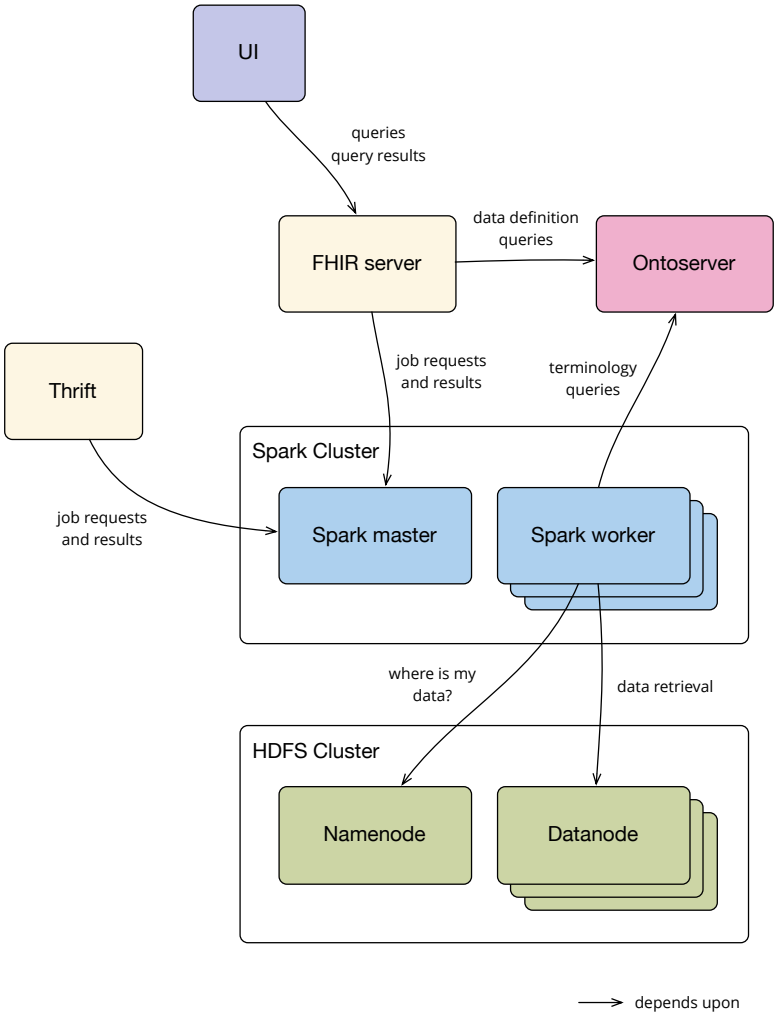
Patient has been prescribed a type of TNF inhibitor

Run-time resolution of terminology queries

- Traditional data warehousing technology relies heavily on pre-computation of value sets used in aggregation and filtering
- This does not work when dealing with large and complex terminologies such as SNOMED CT, as the range of possible expressions within queries is unbounded
- Solution: delegate execution of terminology expressions to a dedicated service



Current solution



Notebooks

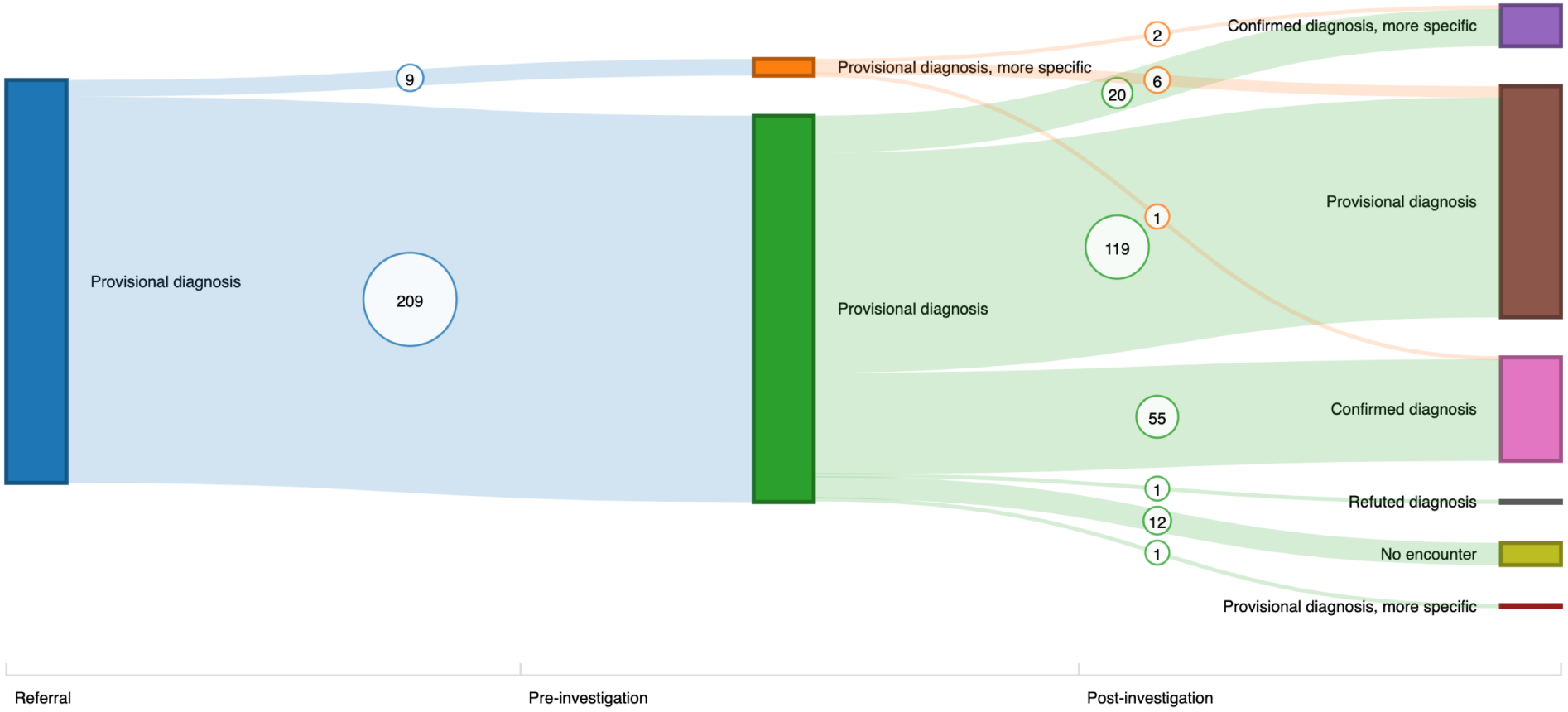
- We have created a number of notebooks demonstrating analytic techniques for analysing SNOMED CT coded data
- Examples demonstrate the use of the Spark Python API to query the data warehouse and Ontoserver, and combine the results
- Future experimentation with Zeppelin and R-based notebooks is planned

https://jupyter.org/

https://zeppelin.apache.org/

Phase 2 - Analytics

Number of genomic test patients by encounter type, diagnosis verification status and diagnosis specificity relative to previous diagnoses



Thank you

Alejandro Metke
The Australian e-Health Research Centre

e: alejandro.metke@csiro.au