

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)



- 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

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**

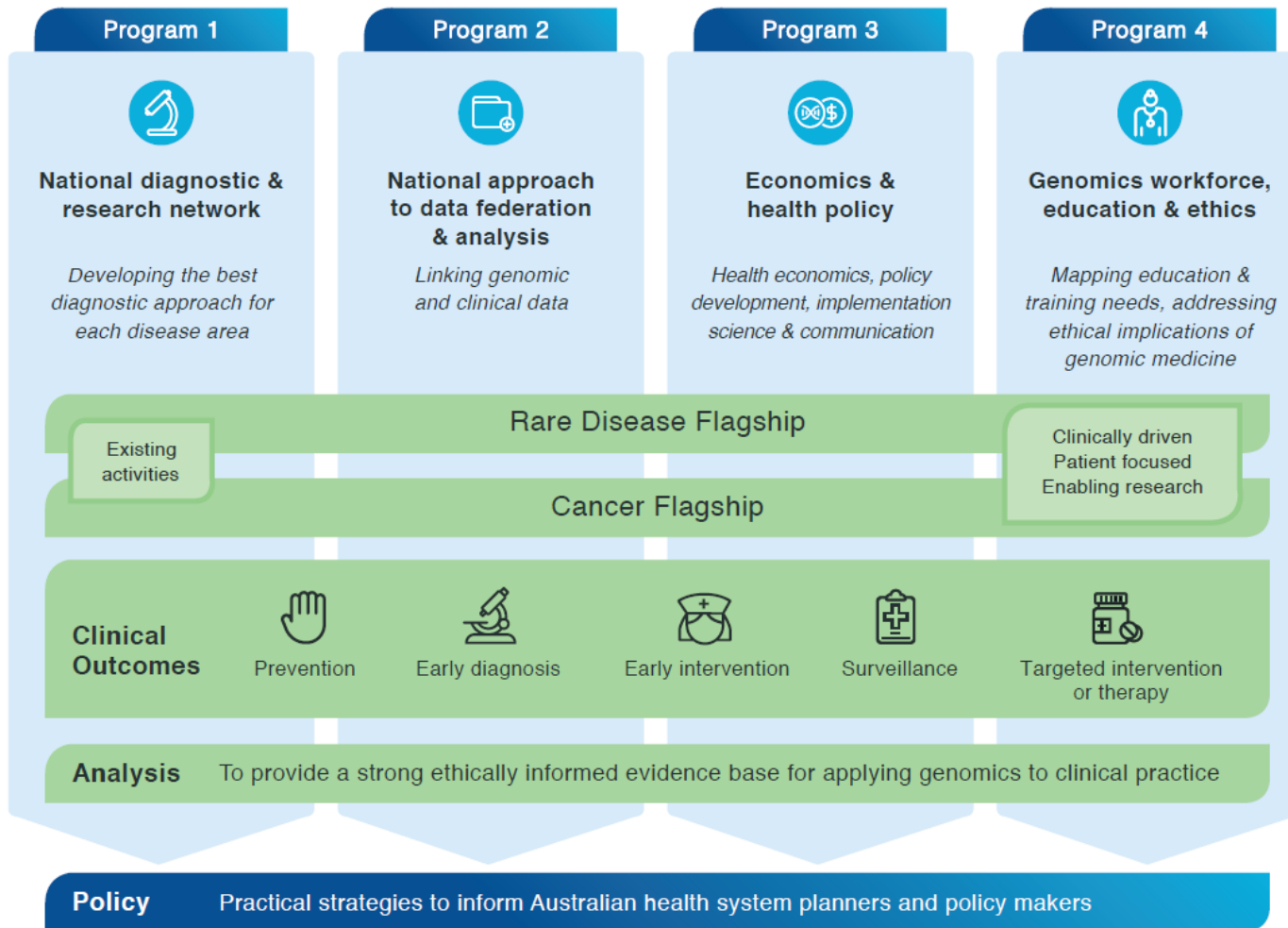


- Royal Darwin Hospital
- The University of Queensland**
Lady Cilento Children's Hospital
Institute for Molecular Bioscience
- QIMR Berghofer Medical Research Institute**
Wesley Hospital
- Royal Brisbane and Women's Hospital/Genetic Health Queensland**
Princess Alexandra Hospital
Diamantina Institute
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
- Royal Hobart Hospital
- Melbourne Genomics Health Alliance**
Monash Health
Victorian Comprehensive Cancer Centre
Floreys Institute of Neuroscience and Mental Health



Australian Genomics

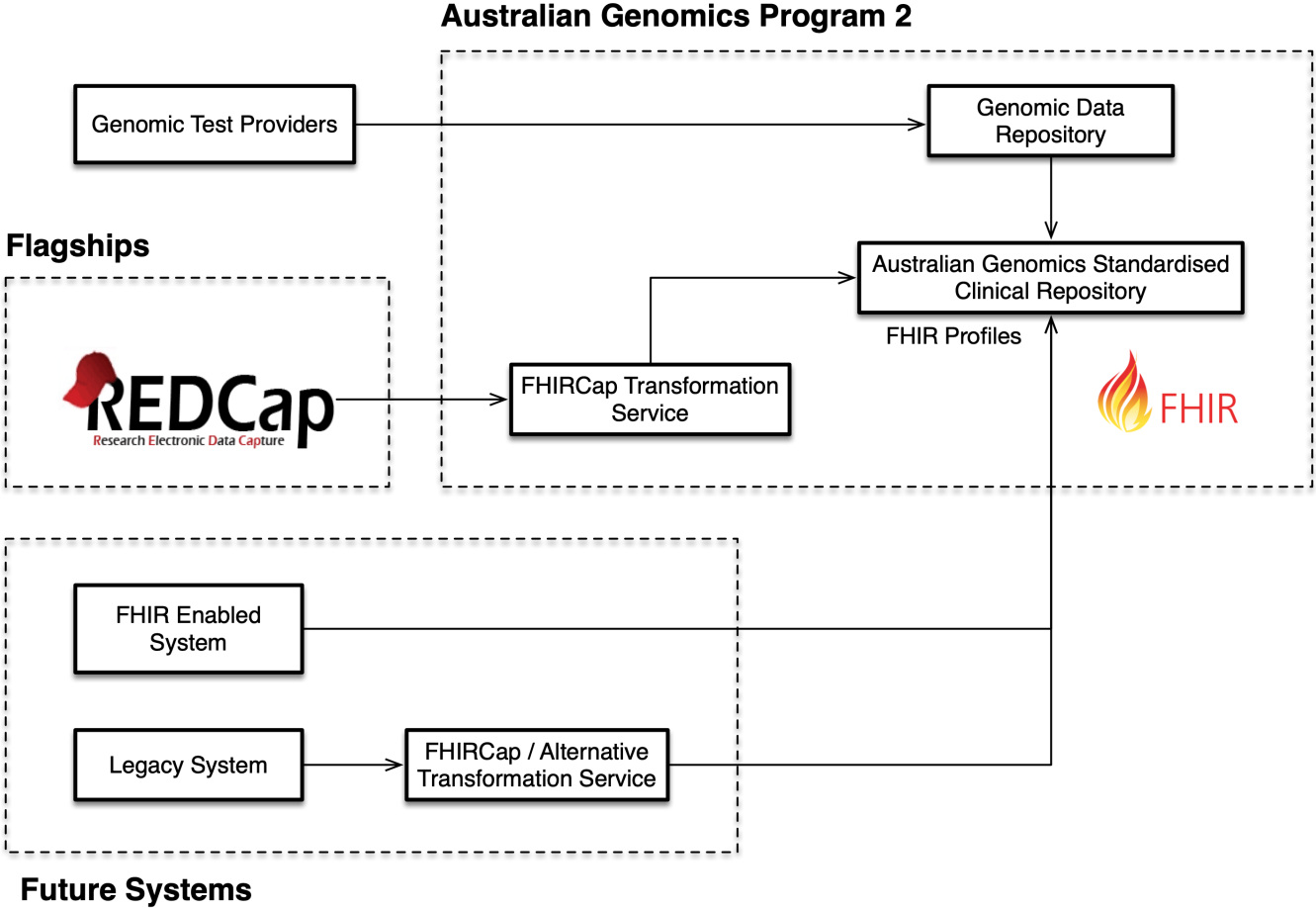
PROGRAMS, FLAGSHIPS AND PROJECTS



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

The screenshot displays the REDCap interface for project **AGHA_20180827**. The user is logged in as **site_admin**. The sidebar on the left shows navigation options: **My Projects** or **Control Center**, **Project Home** or **Project Setup**, and **REDCap Messenger**. The project status is **Development**. Under **Data Collection**, there are links for **Record Status Dashboard**, **Add / Edit Records**, and **Study Number**. The **Data Collection Instruments** list includes: **Study Identifiers** (highlighted with a blue box), **Participant Tracking**, **Kidgen Participant Tracking**, **GI Participant Tracking**, **Ethics Control**, **Demographics**, **Personal Contact Info**, **Family History**, **Family Identifiers**, **Genetic Counselling**, **NMD**, **Mito**, **MCD**, **EE**, **Kidgen**, **Kidgen Variant Interpretation And Result Return**, **Leuko**, and **GI** (highlighted with a red box).

The main form area shows the **Study Identifiers** form. The **Study Number** field is empty. The **Recruitment Site** dropdown is set to **VIC- RCH**. The **Flagship** dropdown is set to **Somatic Cancer/ Solid Tumors (IPR)**. The **Is patient still living?** radio buttons are set to **Alive**. The **Has the participant been withdrawn from the study?** dropdown is set to **No**. The **Form Status** section shows **Complete?** as **Incomplete**. The **Lock this record for this form?** checkbox is unchecked. The **Save & Exit Form** and **Save & ...** buttons are visible at the bottom right.

REDCap in the Flagships

- Clinical data capture is not standardised

Other epilepsy features	
Convulsive status epilepticus	<input type="radio"/> Yes <input checked="" type="radio"/> No reset
Non-convulsive status epilepticus	<input type="radio"/> Yes <input checked="" type="radio"/> No reset
Febrile seizures	<input type="radio"/> Yes <input checked="" type="radio"/> No reset

Nerve Conduction Studies	
Nerve conduction studies	<input checked="" type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not assessed reset
Motor	<input type="checkbox"/> Demyelinating <input type="checkbox"/> Axonal <input type="checkbox"/> Mixed <input checked="" type="checkbox"/> Normal <input type="checkbox"/> Not assessed
Sensory	<input type="checkbox"/> Demyelinating <input checked="" type="checkbox"/> Axonal <input type="checkbox"/> Mixed <input type="checkbox"/> Normal <input type="checkbox"/> Not assessed

Primary Tumour Site:
<ul style="list-style-type: none">✓ CUPAdrenalBiliaryBladderBreastBone & Soft TissueCervicalColorectalEndometrialHead & NeckLiverLung

Principal phenotypic features	<div>Central hypotonia, apnoeas, feeding difficulties, tall forehead, bitemporal narrowing, scaphocephaly</div> Expand
-------------------------------	---

REDCap in the Flagships

- REDCap FHIR Terminology Plugin

Principal phenotypic features

Central hypotonia, apnoeas, feeding difficulties, tall forehead, bitemporal narrowing, scaphocephaly

Expand



Principal phenotypic features	Central hypotonia HP:0011398 HPO terms
Principal phenotypic features	Apnea HP:0002104 HPO terms
Principal phenotypic features	Feeding difficulties ... HP:0008872 HPO terms
Principal phenotypic features	Bitemporal hollowi... HP:0025386 HPO terms
Principal phenotypic features	Scaphocephaly HP:0030799 HPO terms

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:

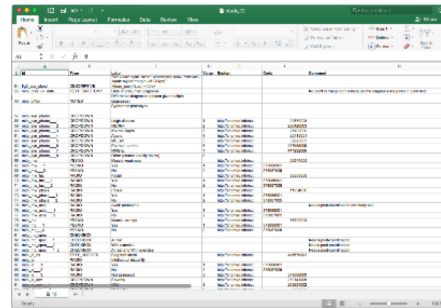
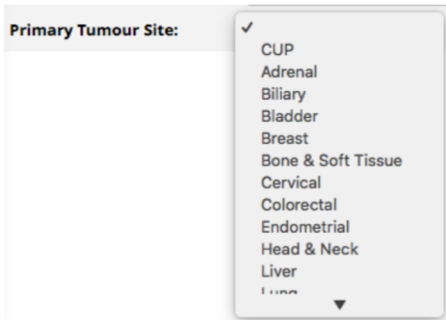
VALUE (pts) != null :

Condition<c1> ->

bodySite[0] = CONCEPT_SELECTED (ts),

code = CONCEPT_VALUE (
http://snomed.info/sct|108369006
);

Mappings:



A screenshot of a data table with columns for 'Code', 'Concept', 'Date', and 'Event'. The table contains multiple rows of data, likely representing mappings between SNOMED CT codes and concepts.

FHIRCap Transformation Example

Somatic Cancer Flagship REDCap Forms

Study Clinician Input 2

Do you think the patient's result was received in time for it to have a clinical impact?

Would the result have made a greater impact if available earlier in patient's treatment journey?

Did the recommendations in the report impact on the care of the patient?

Did any difficulties arise through the process of genomic testing for your patient [pers. first_name] [pers. surname]?

Do you feel the testing process (regardless of result) was of value to the patient?

Any further comments?

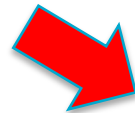
Data to be entered after result review (3 months after reporting)

Patient status

Result changed treatment used compared to previously listed planned treatment? Yes No

Patient ECOG (Eastern Central Oncology Group)

Please confirm if all data required after result review has been completed? Yes No

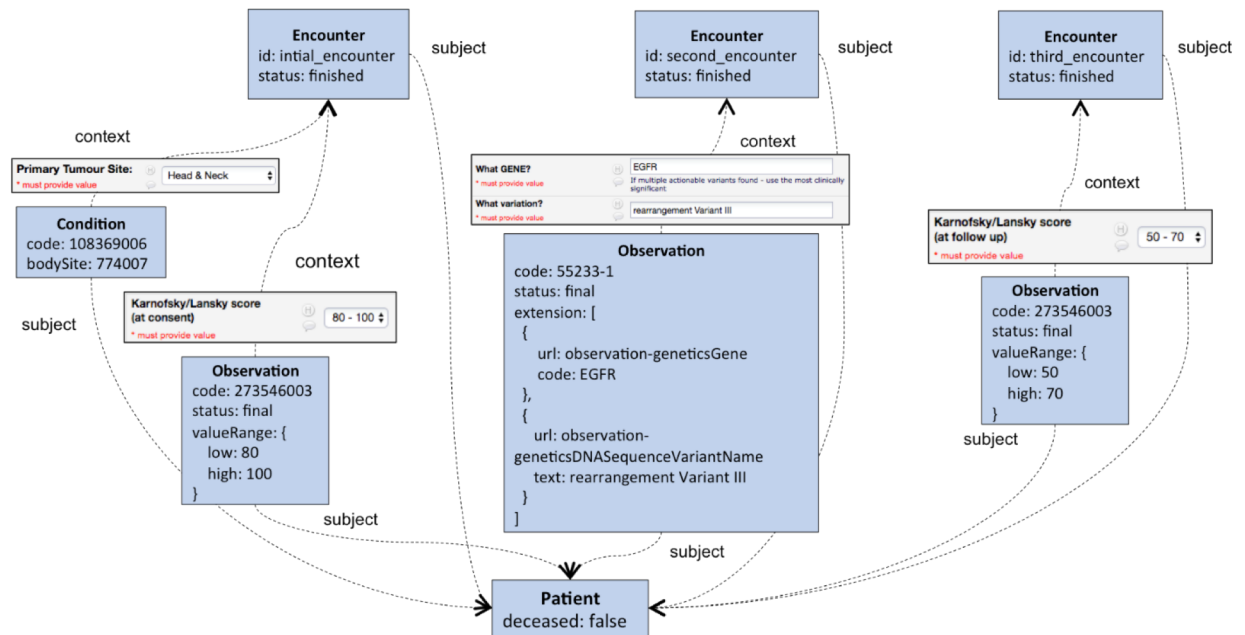


FHIR Representation using SNOMED CT

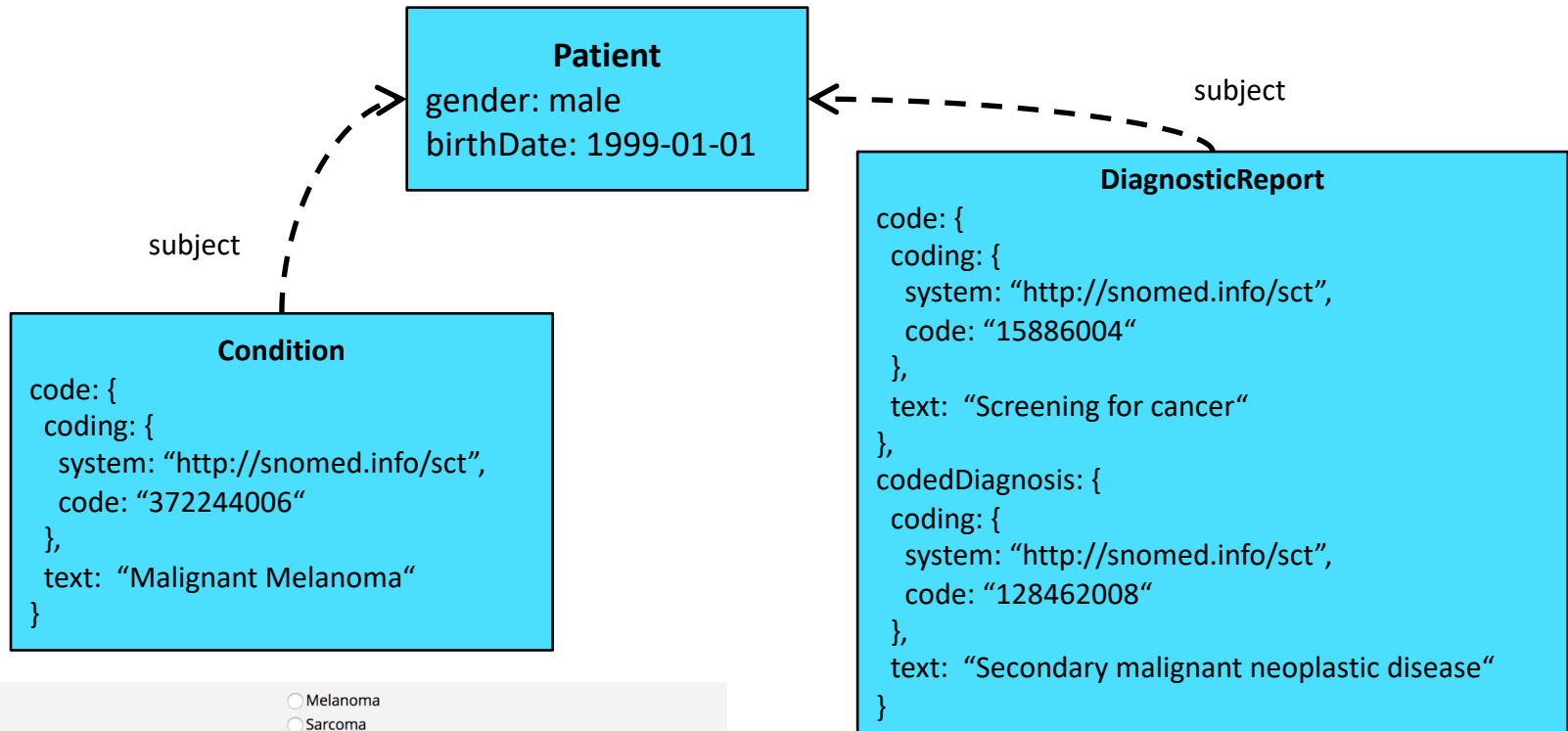
Initial encounter, medical history captured, consent obtained, genetic test ordered

Second encounter, test results interpreted, treatment potentially adjusted

Third encounter, follow up, genetic test value assessed



Examples of Mapping Issues



Histology Broad:
* must provide value

- Melanoma
- Sarcoma
- Adenocarcinoma
- Squamous Cell Carcinoma
- Neuroendocrine
- Embryonal
- Other

reset

Histology Specific: (please detail specific histology if it is a subtype of the above e.g. Ewing's Sarcoma)
* must provide value

Cancer Diagnosis:
* must provide value

Newly diagnosed locally advanced

Newly diagnosed metastatic

Local Recurrence or Locally Advanced (not newly diagnosed)

Progressive Metastatic or Metastatic Relapse

Unsure

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

Congenital heart disease (CHD): <i>* must provide value</i>	<input checked="" type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Unknown	reset
Congenital heart disease detail:	<input type="text" value="No Results Found"/> <input type="text" value="NOT_FOUND"/>	
Congenital heart disease detail description (if code not found): <i>* must provide value</i>	<input type="text" value="xxx"/>	

- 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

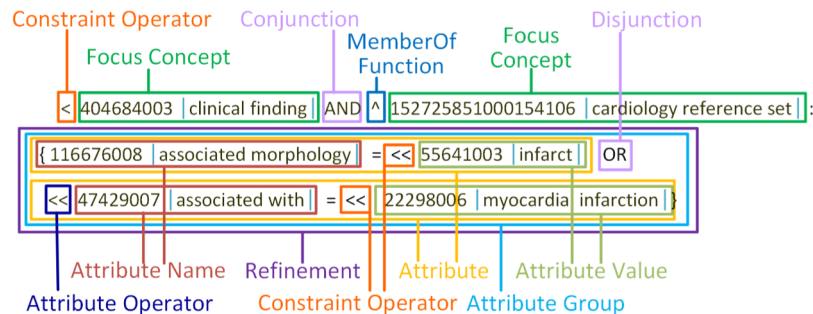
- Pattern to capture coded data

Congenital heart disease (CHD): Yes No Unknown reset
** must provide value*

Congenital heart disease detail:

Bound to a value set to restrict the search space

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



Phase 1 - Data Capture



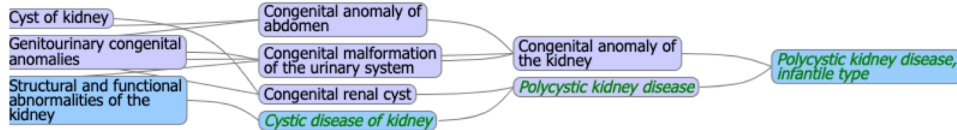
Terminology

Refsets

ValueSets

ECL 🏠

Ontoserver



© Australian e-Health Research Centre

Refset: kidgen-clinical
Showing 1 to 20 of 27 rows
id: kidgen-clinical

◀ Prev | Next ▶

SYSTEM	CODE	DISPLAY
http://snomed.info/sct	726018006	Autosomal dominant tubulointerstitial kidney disease
http://snomed.info/sct	28770003	Polycystic kidney disease, infantile type
http://snomed.info/sct	47461006	Genetic disorder carrier
http://snomed.info/sct	428255004	Tubulointerstitial nephritis
http://snomed.info/sct	111395007	Nephrogenic diabetes insipidus
http://snomed.info/sct	81573002	Collagen disease
http://snomed.info/sct	82525005	Polycystic kidney disease
http://snomed.info/sct	609572000	Maturity-onset diabetes of the young, type 5
http://snomed.info/sct	762464006	Autosomal dominant tubulointerstitial disease

- Human Ancestry Ontology African
- Human Ancestry Ontology African American c
- Human Ancestry Ontology Asian
- Human Ancestry Ontology Categories
- Human Ancestry Ontology Countries
- Human Ancestry Ontology European
- Human Ancestry Ontology Genetically Isolate
- Human Ancestry Ontology Greater Middle East
- Human Ancestry Ontology Hispanic or Latin A
- Human Ancestry Ontology Native American
- Human Ancestry Ontology Oceanian
- Human Ancestry Ontology Uncategorised Pop
- Human Ancestry Ontology Undefined Ancestr
- Human Phenotype Ontology
- KidGen Clinical Diagnosis Value Set
- KidGen Diagnosis Value Set
- KidGen Final Diagnosis Value Set
- KidGen Referral Diagnosis Value Set
- KidgenAlportCategory
- KidgenCAKUTCcategory
- KidgenComplementDisorderCategory
- KidgenCysticCategory
- KidgenGlomerularCategory
- KidgenNephroticCategory
- KidgenTubularDiseasesCategory
- Mitochondrial Flagship Value Set
- Neuromuscular Disorders Flagship Value Set
- SNOMED CT Phenotype ValueSet
- iPredict Somatic Cancer Value Set
- v3.FamilyMember

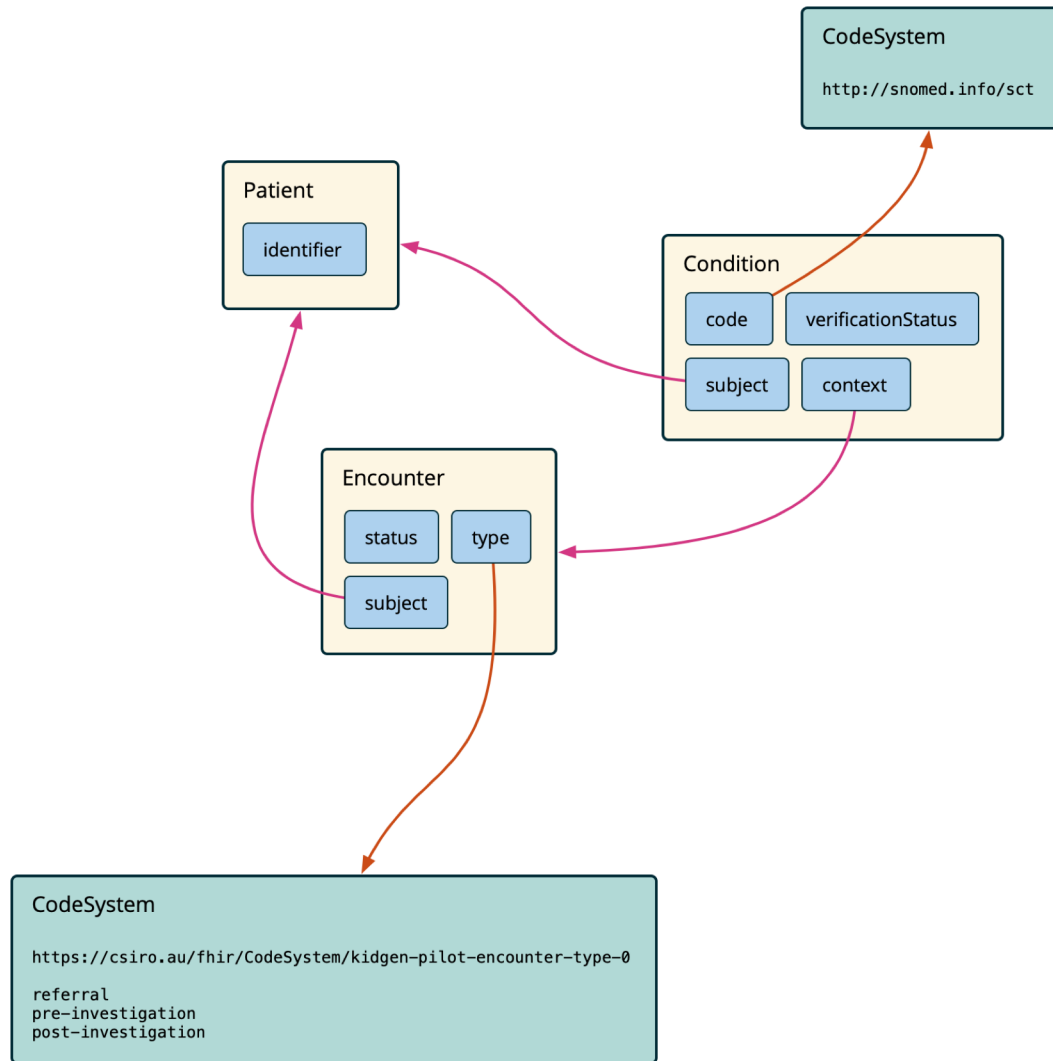


Phase 1 - Data Capture

Coded Diagnoses (for CSIRO)	
Number of coded clinical/suspected diagnoses at referral:	<input type="text" value="1"/>
First clinical/suspected diagnosis at referral (coded):	<input type="text" value="No Results Found"/> NOT_FOUND
Diagnosis text:	<input type="text" value="Unclear, biopsy suggestive of inherited GB"/>
First clinical/suspected diagnosis at referral modifiers:	<input type="checkbox"/> Suspected
Number of coded clinical/suspected diagnosis(es) at clinic:	<input type="text" value="1"/>
First clinical/suspected diagnosis(es) at clinic (coded):	<input type="text" value="Alport syndrome"/> 770414008 Al
First clinical/suspected diagnosis at clinic modifiers:	<input type="checkbox"/> Suspected
Number of coded final clinical diagnosis(es):	<input type="text" value="1"/>
Final clinical diagnosis (coded):	<input type="text" value="a au re"/> Type to begin searching
First final clinical diagnosis modifiers:	<input type="text" value=""/> [717767009] Alport syndrome autosomal recessive http://snomed.info/sct] Alport syndrome autosomal recessive
Form Status	
Complete?	<input type="checkbox"/> [78921008] Autosomal recessive ocular albinism http://snomed.info/sct] Autosomal recessive ocular albinism
	[719104003] Autosomal recessive palmoplantar keratoderma and congenital alopecia syndrome http://snomed.info/sct] Autosomal recessive palmoplantar keratoderma and congenital alopecia syndrome
	[725434009] Autosomal recessive faciodigitogenital syndrome http://snomed.info/sct] Autosomal

[The REDCap Consortium](#) | [Citing REDCap](#) | <http://www.mcri.edu.au>
Murdoch Childrens Research Institute

Phase 1 – Transformation to FHIR



Phase 1 – Modelling Recommendations

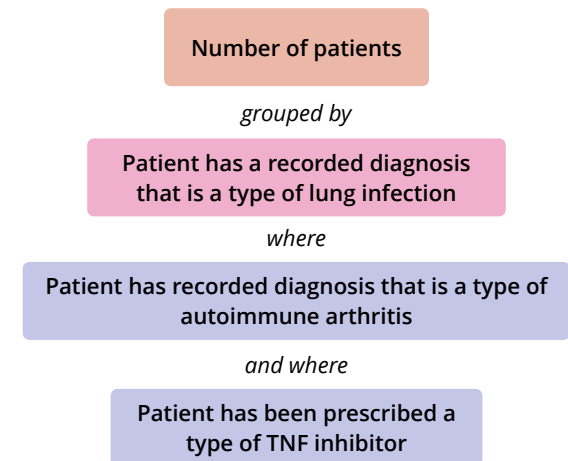
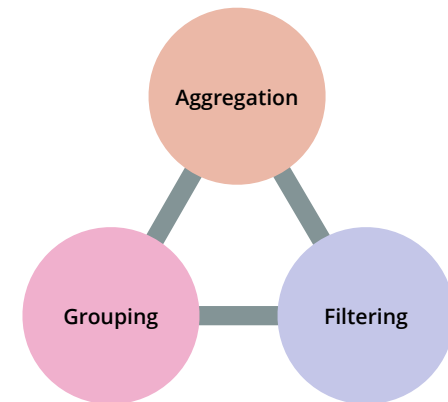
- 16 modelling issues were reported
- Issues included:
 - Incorrect modelling, e.g., concepts in the wrong hierarchy
 - Incomplete modelling, e.g., concepts missing relationships
 - Missing concepts
 - Missing synonyms

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

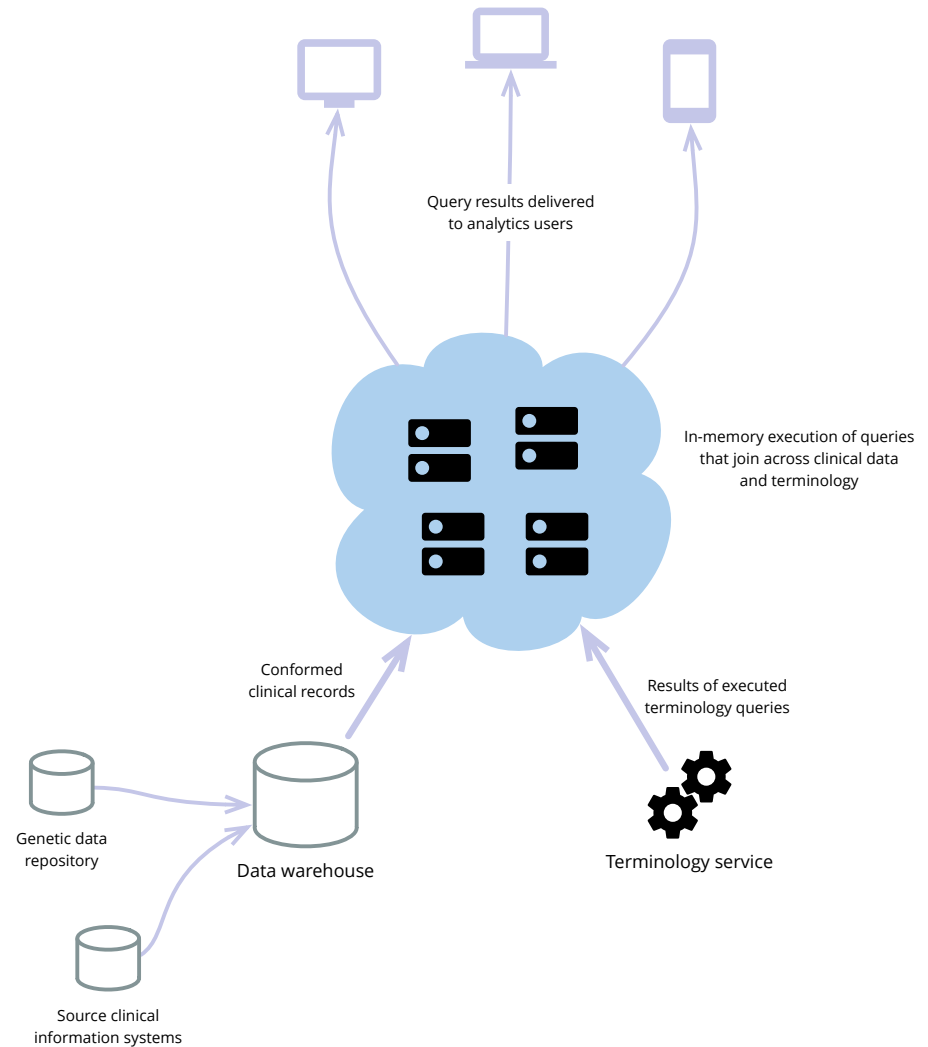
Interactive analytics, delivered via REST API

- 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

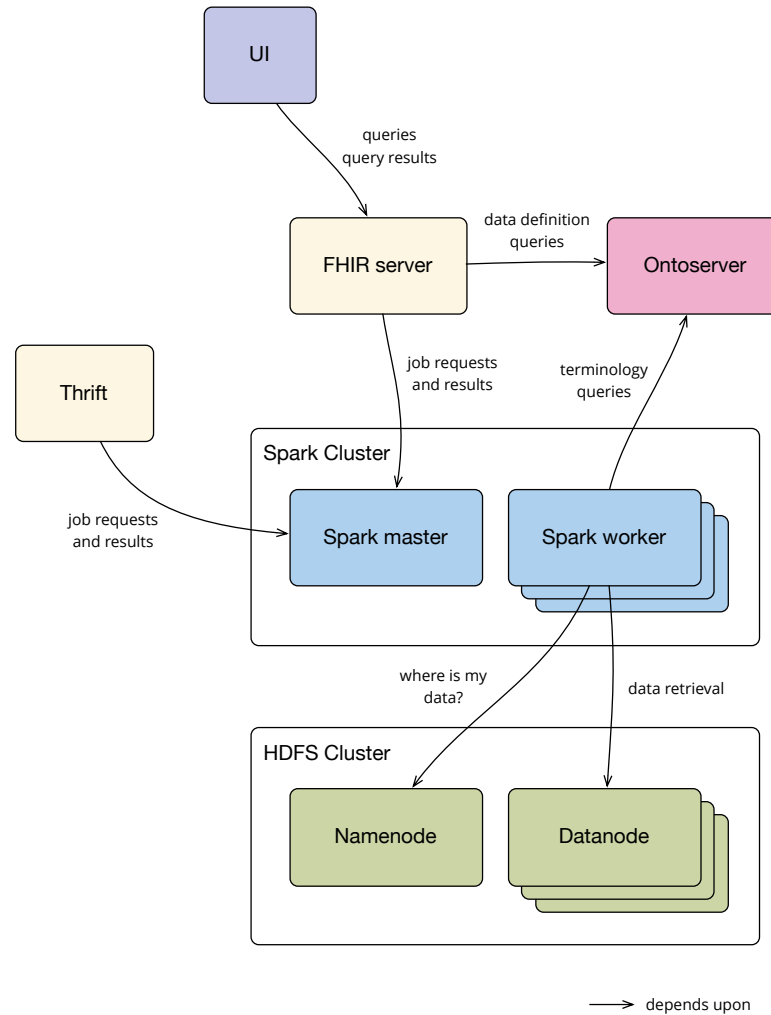


Run-time resolution of terminology queries

- 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

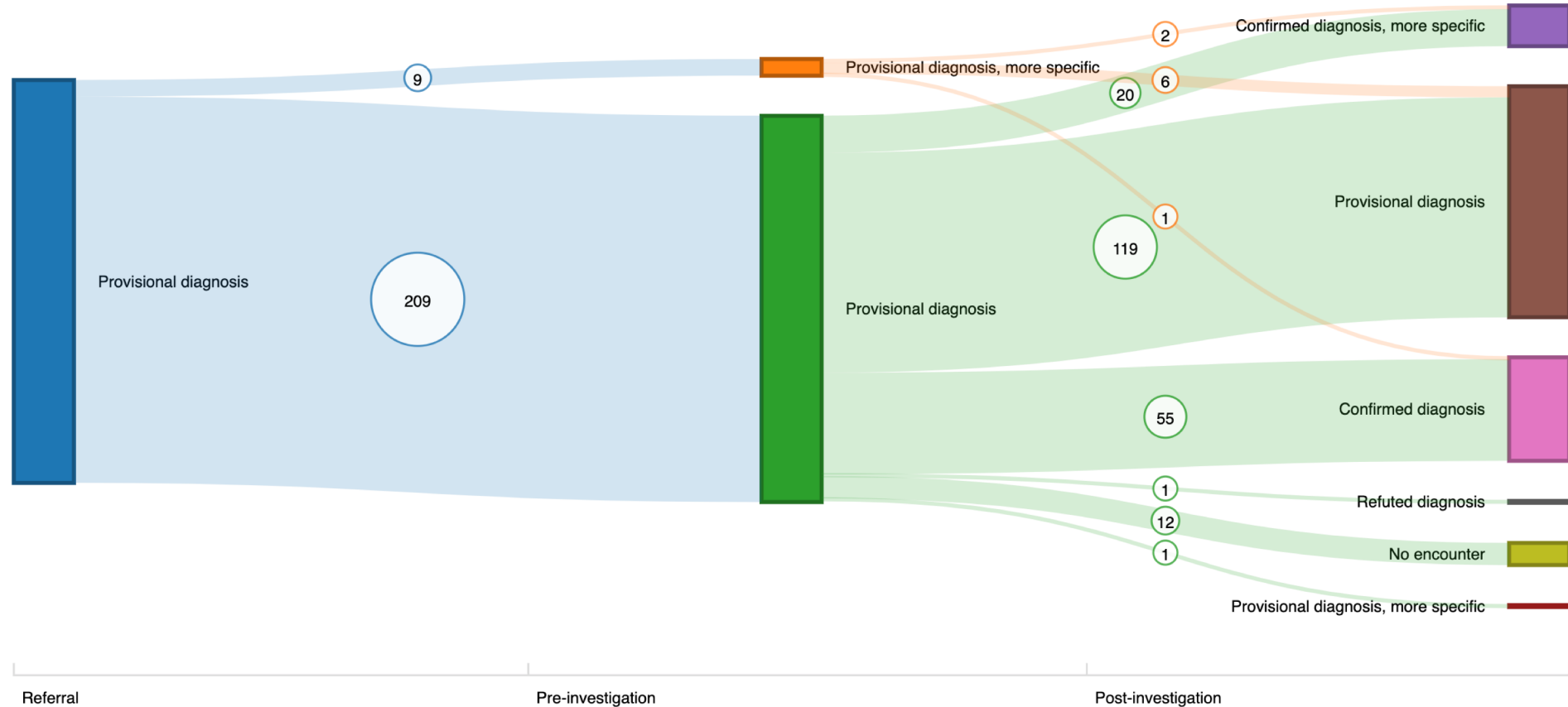
<https://jupyter.org/>

<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

THE AUSTRALIAN E-HEALTH CENTRE
www.csiro.au

