AUSTRALIAN E-HEALTH RESEARCH CENTRE
www.csiro.au

Using SNOMED CT to do Deep Phenotyping of Cardiovascular Genetic Disorders

SNOMED International Pilot with Australian Genomics Cardiovascular Genetic Disorders Flagship

Alejandro Metke-Jimenez, Madonna Kemp, Donna Truran
June 30, 2020

SNOMED International
Ian Green

E-publish Manuscript Number: EP204331

Citation

Metke-Jimenez A, Kemp M, Truran D (2020) SNOMED International Pilot with Australian Genomics Cardiac Flagship: Using SNOMED CT to do Deep Phenotyping of Cardiovascular Genetic Disorders. CSIRO, Australia.

Copyright

Important disclaimer

Contents

Figures

Tables

Acknowledgments

This work was funded by SNOMED International and Australian Genomics. Australian Genomics is supported by the National Health and Medical Research Council (GNT1113531).

Executive summary

This document describes a pilot done with the Australian Genomics Renal Cardiac Flagship where the suitability of SNOMED CT for doing deep phenotyping in this clinical area was assessed. The project includes the design of REDCap forms to capture patient data, done in consultation with flagship clinicians, and the development of value sets to constrain the search space of autocomplete widgets used in the forms. The user interface uses a pattern that allows entering free text descriptions whenever a SNOMED CT concept is not found. These free text descriptions were used to find gaps in SNOMED CT based on data for 276 patients. The report concludes that SNOMED CT is suitable to do deep phenotyping in this clinical area and also provides recommendations on how to improve the issues that were found. The value sets that were developed are also made available and can be used as the basis for one or more cardiovascular genomics reference sets.

1 Introduction

Australian Genomics is a national initiative building evidence and developing a framework to support the implementation of genomics medicine in the Australian healthcare system. The initiative encompasses four programs of work: establishing a national diagnostic and research network in genomics; developing a national approach to data federation and analysis; informing health policy, conducting health economic analyses, applying implementation science methods and addressing ethical implications; and evaluating the needs of the genomic workforce.

The Australian e-Health Research Centre is part of Program 2 and leads the Phenotype Capture subgroup, which is working towards the standardisation of patient phenotype data currently captured by clinical projects. The team interacts with several flagships (the organisations doing clinical research) and is currently developing information models and value sets for the flagships' clinical areas.

This document describes a pilot with the Australian Genomics Cardiovascular Flagship that looks to incorporate the use of SNOMED CT to capture detailed patient phenotype information. The pilot has two main goals: to assess the suitability of SNOMED CT to do deep phenotyping in this specific clinical domain, and to provide recommendations on how to improve SNOMED CT if any gaps are found.

2 Background

The Cardiovascular Genetic Disorders Flagship is enabling genomic testing for patients with genetic heart conditions such as inherited cardiomyopathy, primary arrhythmia and congenital heart disease. The Flagship is recruiting up to 600 participants from across the country to undergo whole genome sequencing.

Some of the goals of the Flagship include:

- Collect evidence to develop "best practices" in cardiac genetic testing. This will help inform policy and educate the clinical workforce.
- Evaluate genetic testing for congenital heart disease.
- Determine the effectiveness of whole genome sequencing in the clinical diagnoses and treatment.

To be able to achieve these goals, it is necessary to collect high quality clinical data alongside the genomics data generated as part of the whole genome sequencing. The clinical data is currently being collected in the REDCap platform which, even though flexible and easy to use, hinders the standardisation of this information.

The Phenotype Capture group, led by the Australian eHealth Research Centre, has developed a tool called RedMatch, that allows transforming REDCap forms into Fast Healthcare Interoperable Resource (FHIR) resources, a new HL7 standard that is gaining momentum worldwide. Because of the flexibility offered by REDCap, it is important that the forms are designed in a way that facilitates their transformation to FHIR. The team worked with generic counsellors and clinicians in the flagship to incorporate best practices in the design of the forms used to capture the patients' clinical data. This included using the FHIR Ontology External Module[1], a REDCap plugin developed by the AEHRC, that allows using a FHIR terminology server to implement autocomplete-style search for coded data. The AHERC has also developed Ontoserver (Metke-Jimenez, Steel, Hansen, & Lawley, 2018), a world-leading FHIR terminology server that supports SNOMED CT natively and implements an algorithm designed specifically to enable quick searching of medical concepts.

A fundamental part of the creation of these FHIR resources is selecting the terminology used to populate the coded fields. In this project we have chosen to use SNOMED CT because it is the most comprehensive clinical terminology currently available, but in this particular space there are other terminologies that are also commonly used, such as the Human Phenotype Ontology (HPO). One of the goals of this project is to assess the suitability of SNOMED CT to capture detail clinical information in this clinical area (also referred to as "deep phenotyping" (Robinson, 2012)).

[1] https://github.com/aehrc/redcap_fhir_ontology_provider

3 Methods

3.1 REDCap form design

The REDCap form used to capture patient data was designed from the ground up, because the pilot started at the same time as the project. This enabled us to use our experience with previous flagships to influence the design of the form and facilitate its transformation to FHIR. Also, we were able to use the FHIR Ontology External Module to capture coded data.

The process included the following stages:

- An initial prototype was built based on a minimum clinical data set form draft that was developed by the flagships clinicians based on an previous project.
- The prototype was presented at the first Face to Face Cardiovascular Genetic Disorders Flagship meeting on February 7th, 2019.
- A series of workshops were conducted with the help of a senior genetic counsellor, a senior terminologist, flagship clinicians and a REDCap administrator, to implement changes to the initial draft.
- A final version of the form was produced and agreed upon by all the flagship stakeholders.

3.1.1 Overall design considerations

The main goal when designing the form was to collect high quality, detailed patient phenotype information (i.e., “deep phenotyping”). However, there are several limitations that had to be considered:

- The people capturing the clinical data, mostly genetic counsellors, have a limited amount of time to carry out the data entry.
- REDCap will work as an independent electronic data capture solution (i.e., it won’t be integrated to any existing EHRs or LIMS).
- There are limited options to change the REDCap user interface and known limitations, such as the need to define the maximum number of elements of a list in advance.

The key obstacle when designing the form was the obvious tension between the desire to capture detailed clinical data and the limited amount of time and resources available to collect this data. The fact that REDCap is not integrated with any existing systems means that no pre-population of data is possible, which only increases the time required to do data entry.

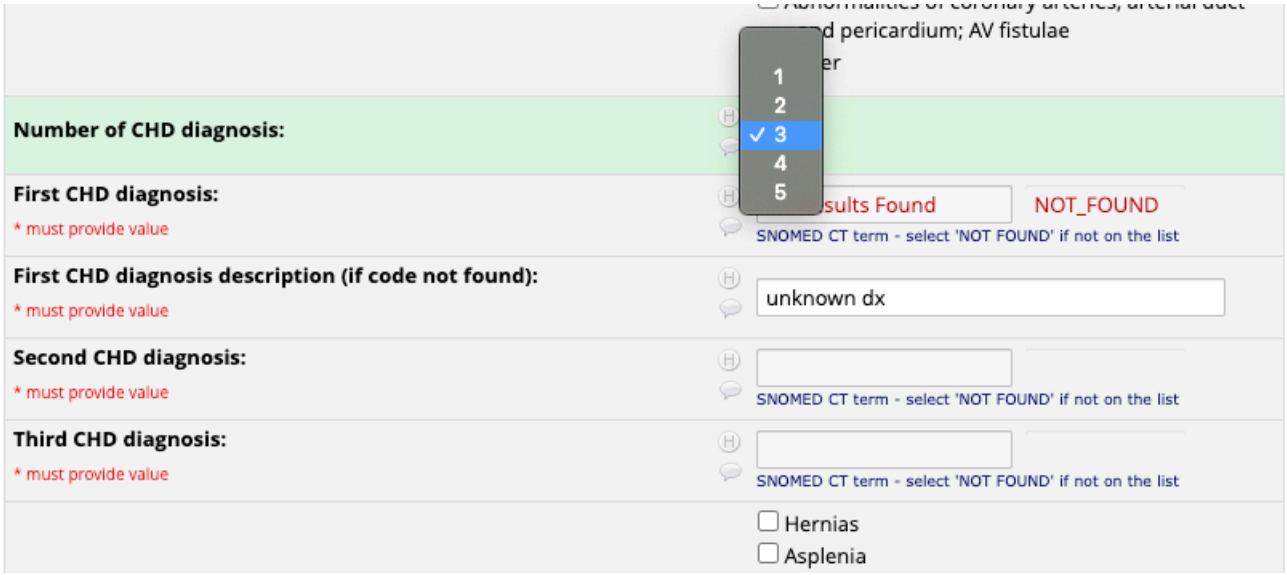

Figure 1 Pattern used to develop the Australian Genomics Cardiovascular Genetic Disorders Flagship REDCap form where the maximum number of repeating elements is defined in advance and branching logic is used to hide unused elements.

In order to achieve a balance between the quality and depth of the data being collected and the time required to do data entry, the following overall design decisions were made:

- The maximum number of elements in lists were determined by consulting the flagship clinicians in the workshops.
- Each element or group of elements in the list is only displayed based on user input. For example, Figure 1 shows a section of the form that allows capturing a maximum of five congenital heart disease (CHD) diagnoses and how branching logic allows displaying only three, based on the selection made by the user.
- High-level categories were designed as checkboxes, to facilitate data entry. When certain elements are selected, more detail can be added using autocomplete fields. Figure 2 illustrates how this pattern works. In this case, when the user selects cancer as a diagnosis, additional fields are displayed that allow entering several specific cancer types, using autocomplete style search. The autocomplete fields are bound to a SNOMED CT value set.

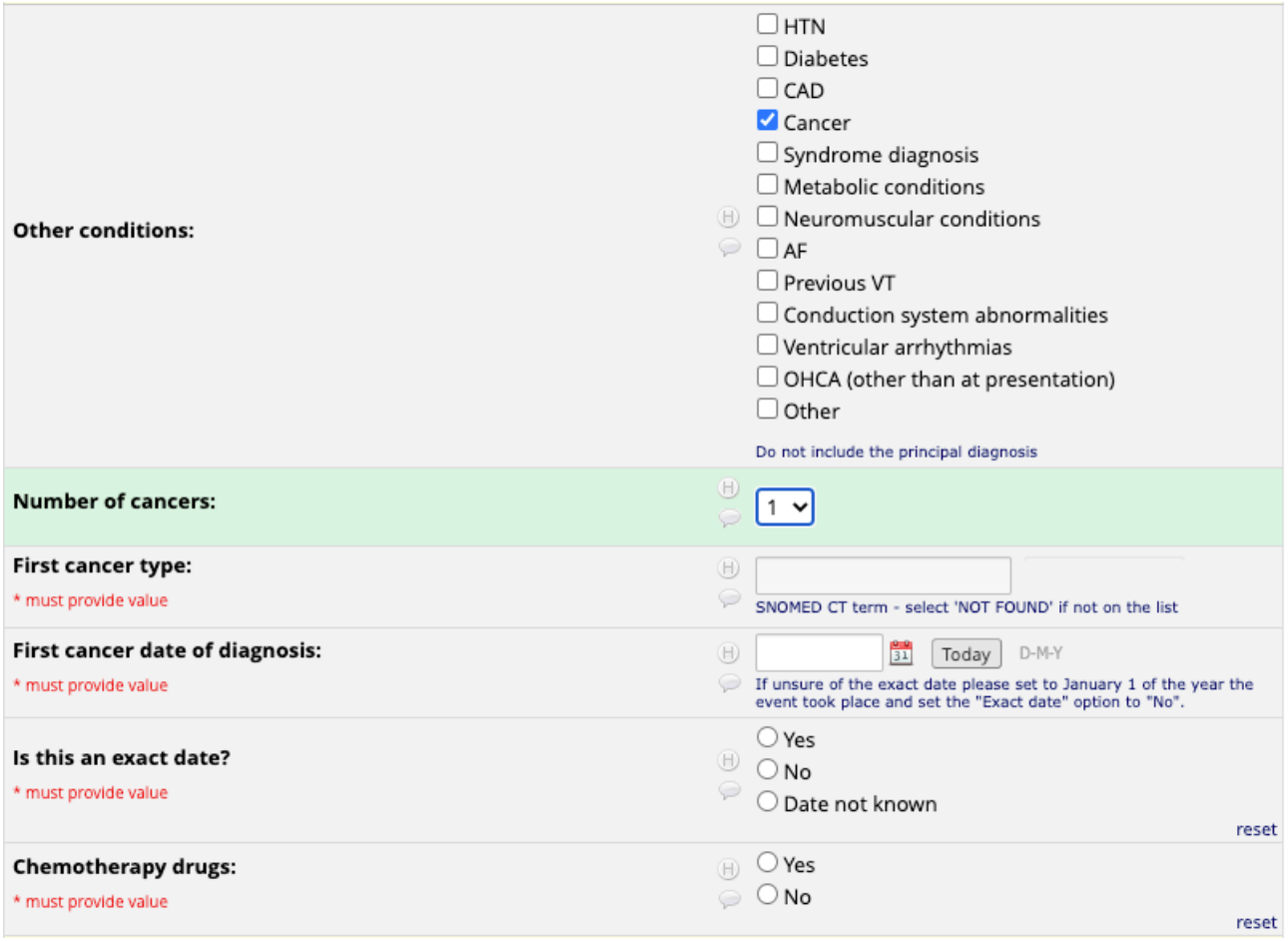

Figure 2 Pattern used to capture high-level information using checkboxes and allowing the user to enter more detailed information using autocomplete fields.

3.1.2 Specific design considerations for capturing coded data

In this project the FHIR Ontology External Module was used to create autocomplete style fields to capture coded data. Each field can be bound to a value set that constrains the search space to a subset of SNOMED CT that is applicable to that field. This improves the quality of the data because the user has fewer and more relevant choices to select.

Even though capturing coded data is preferable, it is always possible that a concept is not available in the terminology or that a user is unable to find it for some reason. In this case, the system should allow entering free text, but at the same time it should encourage the users to search for a concept first. Therefore, this pattern provides a single autocomplete field and returns a 'NOT_FOUND' element if the search produces no results. Only if the 'NOT_FOUND' element is explicitly selected, a free text field is displayed. This pattern is shown in Figure 1, where a free text box is displayed when the user cannot find a concept for the first CHD diagnosis.

The implementation of this pattern allows checking for potentially missing concepts or synonyms in the terminology or suboptimal terminology server behaviour.

3.2 Value set definitions for coded fields

Once the fields in the REDCap form were agreed upon by the flagship clinicians, the team created SNOMED CT value sets for most of the autocomplete fields, except for ancestry, given the preference for the Human Ancestry Ontology (Morales, et al., 2018). The contents of the value sets were discussed with the flagship clinicians as part of the workshops and both the literature (Priori, 2013) and other code systems used in this clinical area were consulted (Franklin, Jacobs, Tchervenkov, & Béland, 2002). The following sections describe the nine initial value sets that were defined.

3.2.1 Coronary Artery Disease Conditions

This value set is used to capture types of coronary artery disease (CAD). It is defined as the descendants of the SNOMED CT concept 251015000 | Coronary artery finding (finding)|. An example of one of the fields that uses this value set is shown in Figure 3.

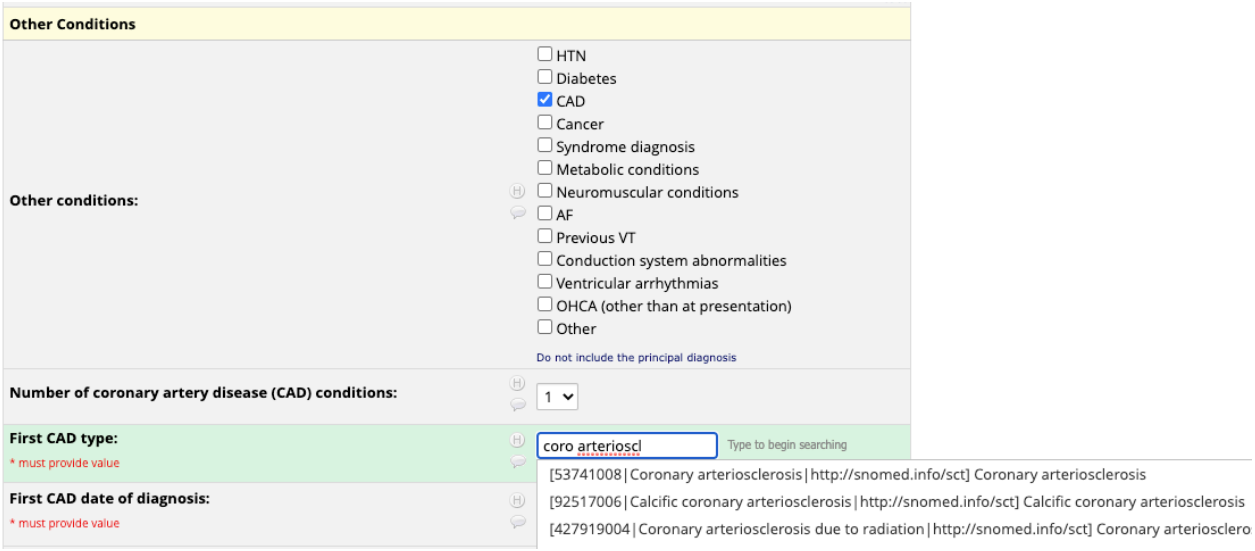

Figure 3 Autocomplete field bound to the Coronary Artery Disease Conditions value set.

3.2.2 Cancer Types

This value set is used to capture specific cancer types once the user has indicated that the patient has cancer. It is defined as the descendants of the SNOMED CT concept 363346000 | Malignant neoplastic disease (disorder) |. An example of one of the fields that uses this value set is shown in Figure 4.

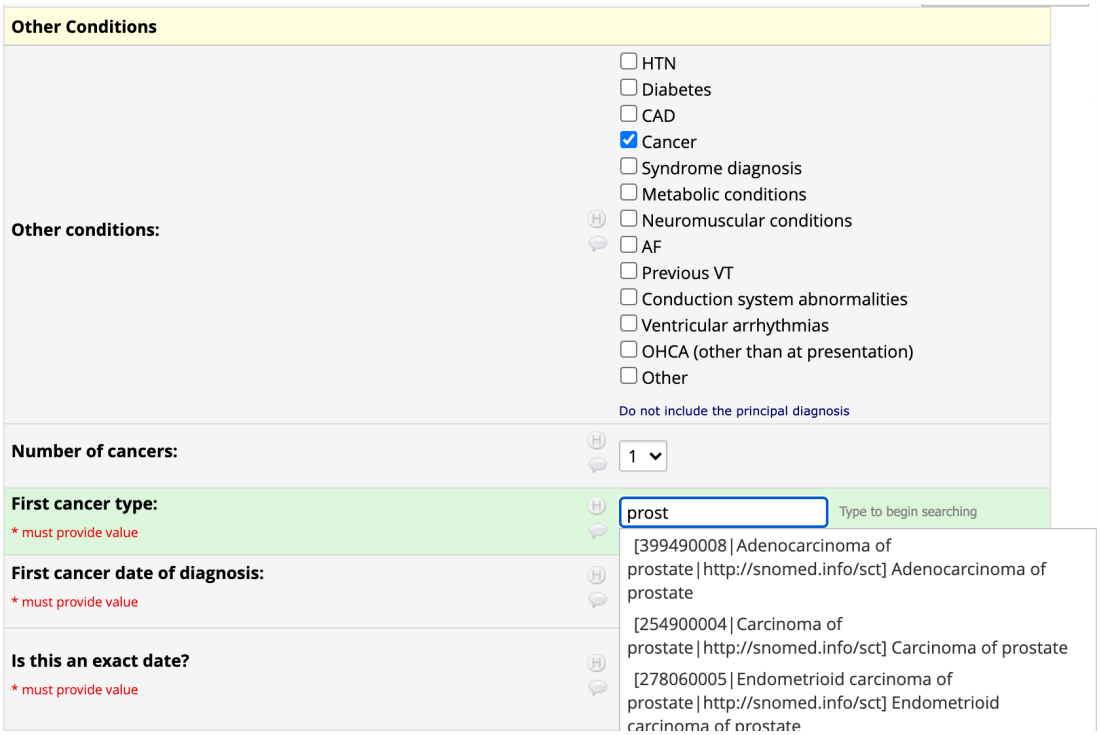

Figure 4 Autocomplete field bound to the Cancer Types value set.

3.2.3 Conduction Abnormalities

This value set is used to capture specific types of conduction abnormalities. It is defined as the descendants of the SNOMED CT concept 44808001 | Conduction disorder of the heart (disorder)|. An example of one of the fields that uses this value set is shown in Figure 5.

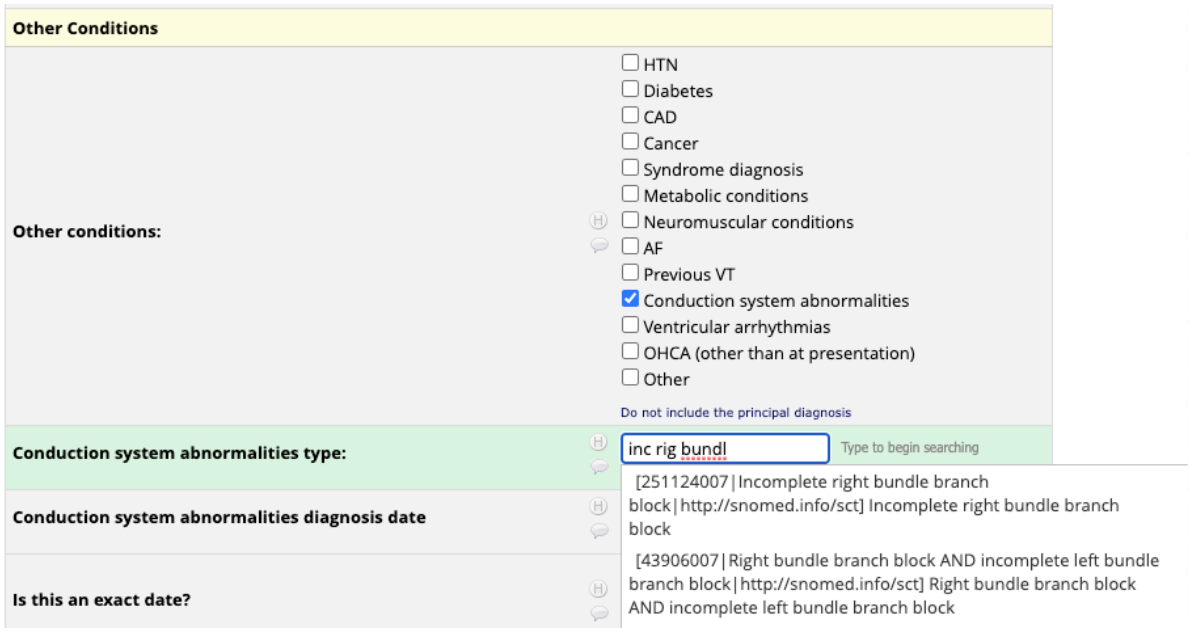

Figure 5 Autocomplete field bound to the Conduction Abnormalities value set.

3.2.4 Ventricular Arrhythmias

This value set is used to capture specific types of ventricular arrhythmias. It is defined as the descendants of the SNOMED CT concept 44103008| Ventricular arrhythmia (disorder)|. An example of one of the fields that uses this value set is shown in Figure 6.

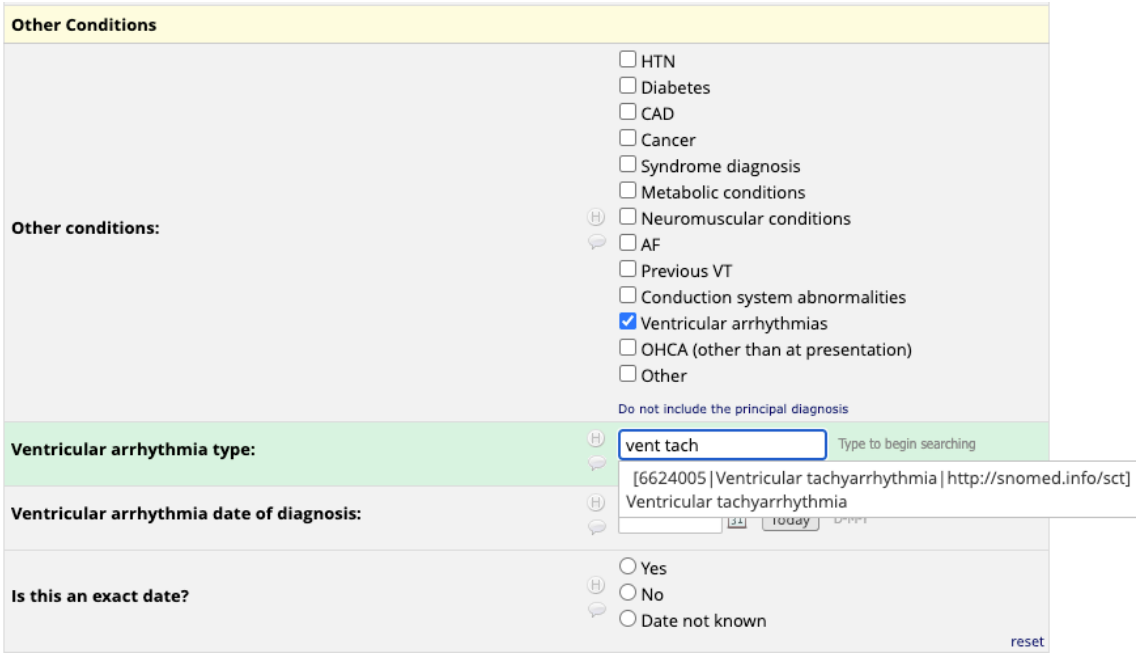

Figure 6 Autocomplete field bound to the Ventricular Arrhythmias value set.

3.2.5 Other Conditions

This value set is used to capture additional conditions that don't fall under the predefined high-level categories. It is defined as the descendants of the SNOMED CT concept 64572001 | Disease (disorder) |, excluding the content in the CAD Conditions, Cancer Types and Ventricular

Arrhythmias value sets, in addition to the concepts that represent diabetes and hypertensions. An example of one of the fields that uses this value set is shown in Figure 7.

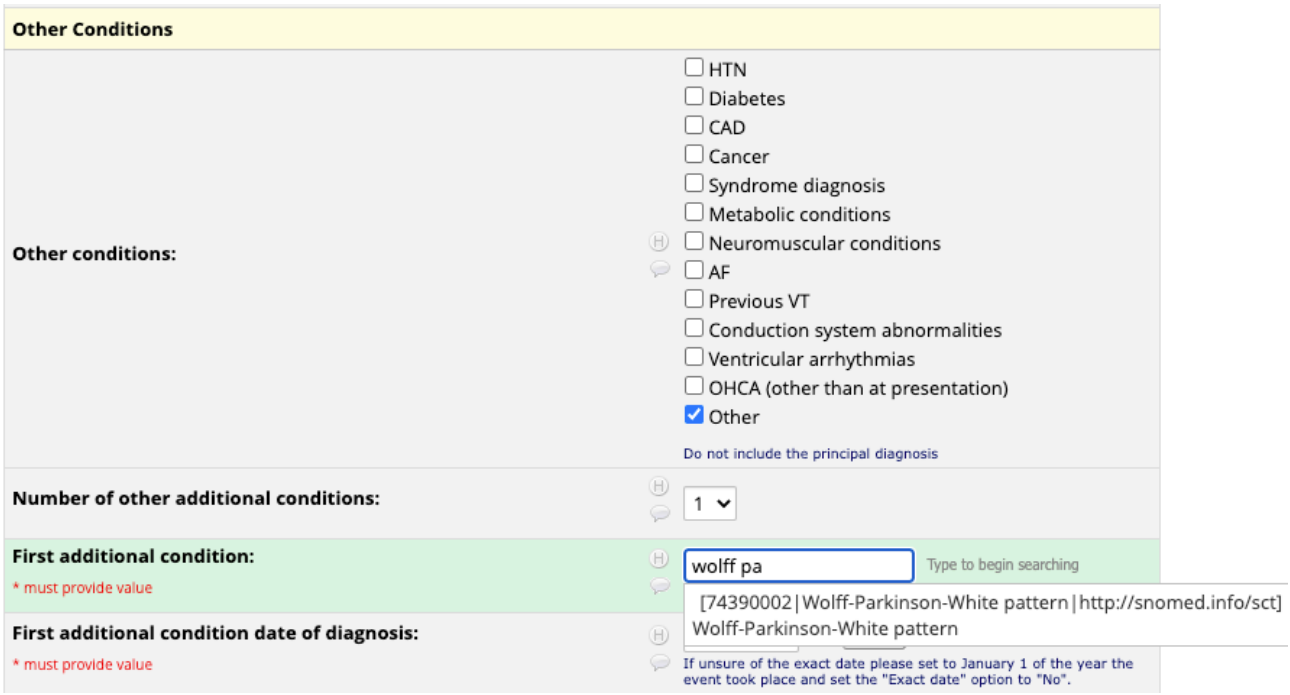

Figure 7 Autocomplete field bound to the Other Conditions value set.

3.2.6 Echocardiogram Valve Disease

This value set is used to capture a patient's specific type of valve disease detected in an echocardiogram. It is defined as the descendants of the SNOMED CT concept 368009 | Heart valve disorder (disorder) |. An example of one of the fields that uses this value set is shown in Figure 8.

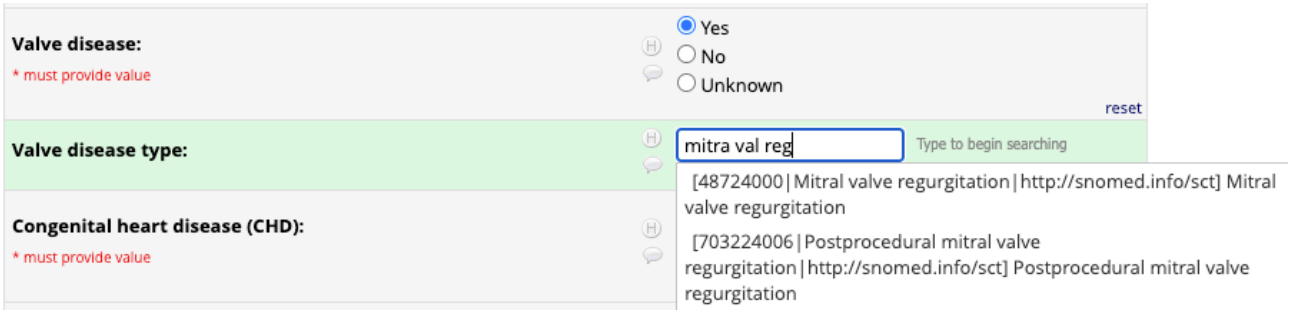

Figure 8 Autocomplete field bound to the Echocardiogram Valve Disease value set.

3.2.7 Congenital Heart Disease Diagnosis

This value set is used to capture details of a congenital heart disease diagnosis. It is defined as the union of the descendants of the following SNOMED CT concepts:

- 253272009 | Congenital abnormality of cardiac connection (disorder) |
- 86299006 | Tetralogy of Fallot (disorder) |
- 449442001 | Congenital abnormality of great cardiac vein (disorder) |

- 253363004 | Abnormality of atrial septum (disorder) |
- 253428008 | Abnormality of common atrioventricular valve in atrioventricular septal defect (disorder) |
- 253549006 | Ventricular septal abnormality (disorder) |
- 28574005 | Congenital anomaly of coronary artery (disorder) |
- 400159008 | Congenital vascular malformation (disorder) |

An example of one of the fields that uses this value set is shown in Figure 9.

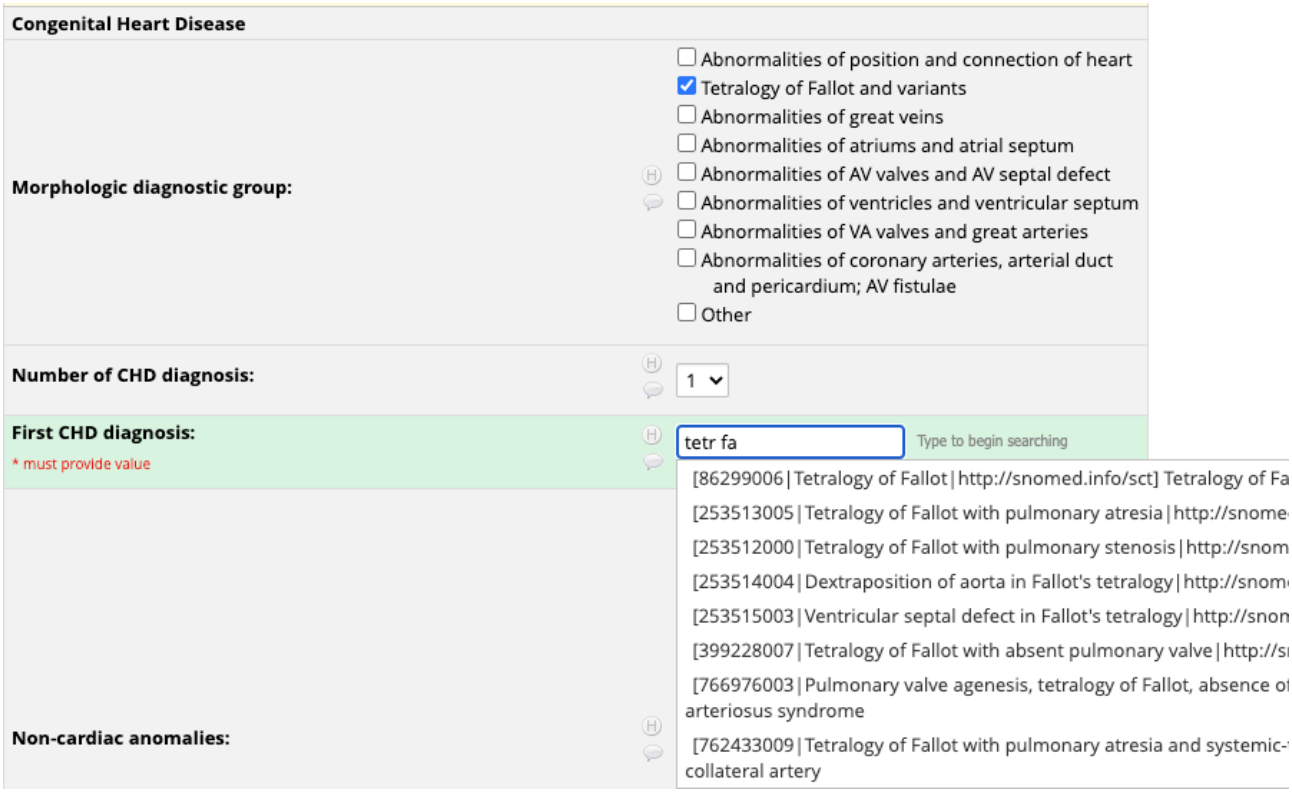

Figure 9 Autocomplete field bound to the Congenital Heart Disease Diagnosis value set.

3.2.8 Congenital Heart Disease Non-cardiac Anomalies

This value set is used to capture details of non-cardiac anomalies that are relevant for congenital heart disease. It is defined as the union of the following:

- Descendants and self of 39839004 | Diaphragmatic hernia (disorder) |
- Descendants and self of 40775003 | Intestinal hernia (disorder) |
- Descendants and self of 396232000 | Inguinal hernia (disorder) |
- Descendants of 707147002 | Asplenia (disorder) |
- Descendants of 89166001 | Congenital anomaly of liver (disorder) |
- Descendants of 44513007 | Congenital anomaly of the kidney (disorder) |
- Descendants of 31686000 | Congenital anomaly of lower alimentary tract (disorder) |

- Descendants of 69771008 | Congenital anomaly of oesophagus (disorder) |
- Descendants of 14532008 | Congenital anomaly of trachea (disorder) |
- Descendants of 74877002 | Congenital anomaly of spine (disorder) |
- Descendants of 73573004 | Congenital anomaly of musculoskeletal system (disorder) |
- Descendants of 127329003 | Congenital anomaly of visual system (disorder) |
- Descendants of 95827002 | Congenital hearing disorder (disorder) |
- Descendants of 87979003 | Cleft palate (disorder) |
- Descendants of 80281008 | Cleft lip (disorder) |
- Descendants of 276720006 | Dysmorphism (disorder) |
- Descendants of 57148006 | Congenital anomaly of brain (disorder) |
- Descendants of 700364009 | Neurodevelopmental disorder (disorder) |

An example of one of the fields that uses this value set is shown in Figure 10.

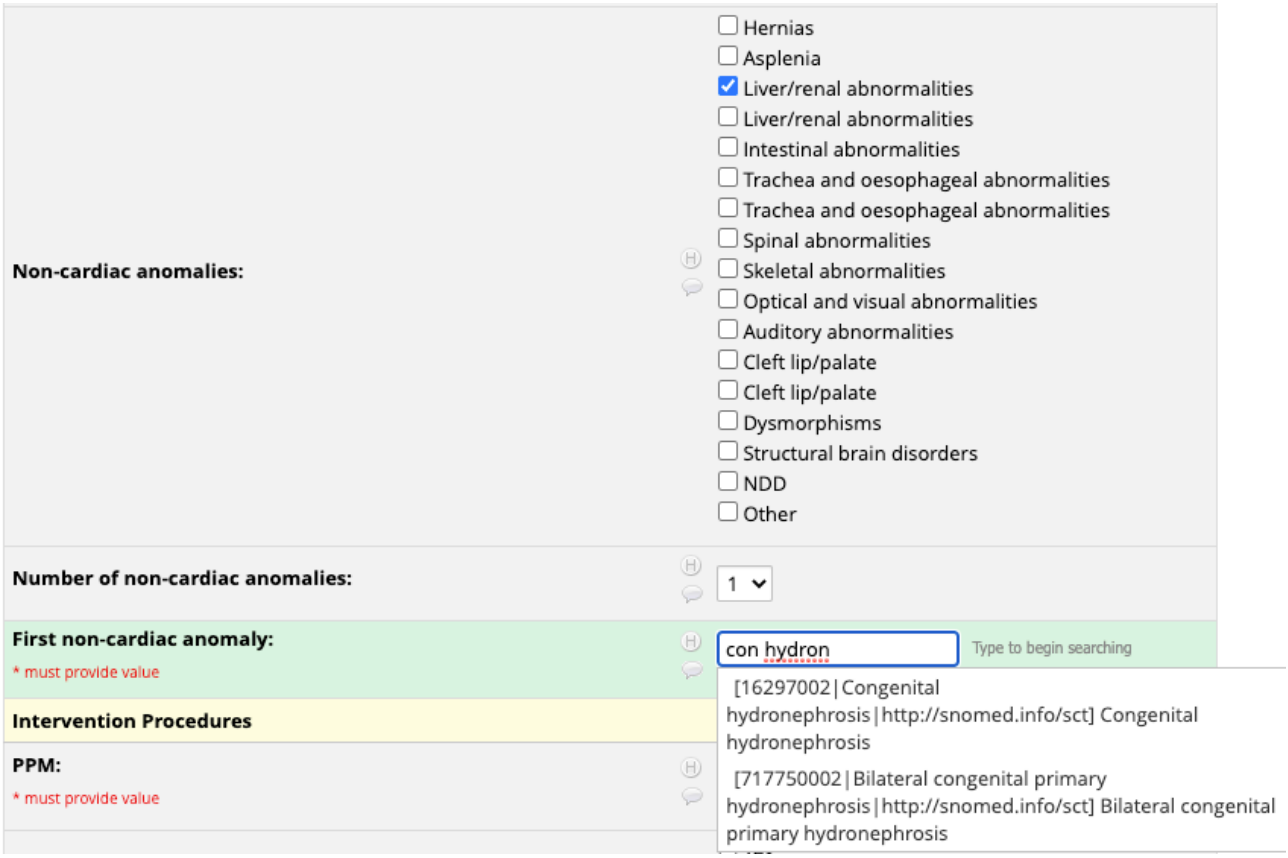

Figure 10 Autocomplete field bound to the Congenital Heart Disease Non-cardiac Anomalies value set.

3.2.9 Congenital Heart Disease Procedures

This value set is used to capture procedures that are relevant for congenital heart disease. It is defined as the descendants of the SNOMED CT concept 118672003 | Procedure on cardiovascular system |. An example of one of the fields that uses this value set is shown in Figure 11.

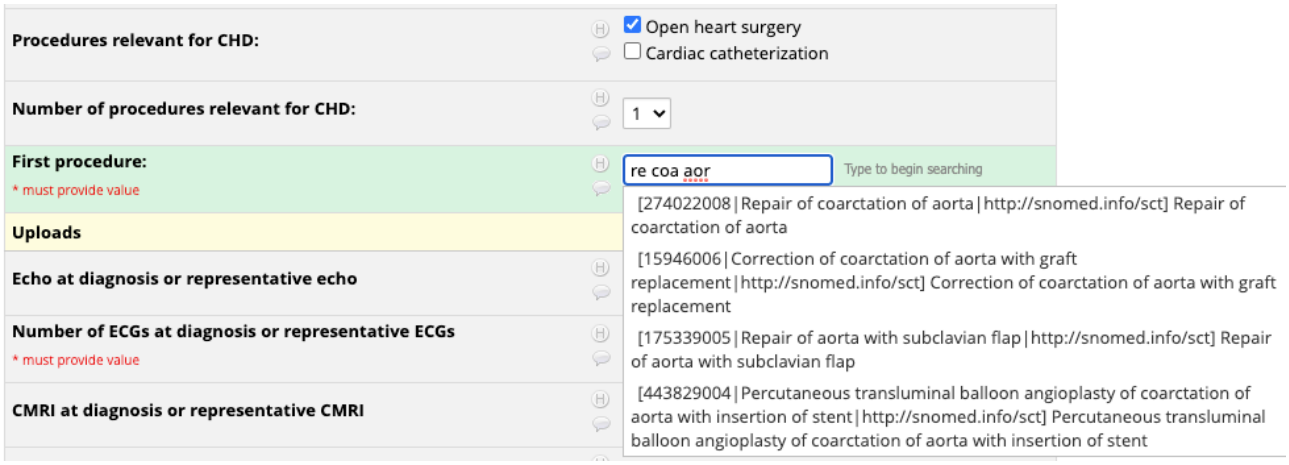

Figure 11 Autocomplete field bound to the Congenital Heart Disease Procedures value set.

4 Results

The final version of the REDCap form was put into production on the 18th of April, 2019, and recruitment of patients started shortly after. A snapshot of the data was generated on June 12, 2020, and contained data for 276 patients. A review of the coded data in this set was conducted by an expert clinical terminologist and the flagship clinicians were consulted when necessary. The following sections show the results of this review.

4.1 SNOMED CT issues

The pattern used to capture coded data was designed to enable the review of the underlying terminology, in this case SNOMED CT, once the form was used to capture real patient data. The analysis revealed three types of problems related to SNOMED CT, which are discussed in the following sections.

4.1.1 Missing synonyms

One of the reasons a concept might not be found when using an autocomplete field is a missing synonym. In this setting, because of time constraints, it is unlikely that users will be willing to use different search strings to find a concept and would rather opt for selecting 'NOT_FOUND' and entering a free text description. This is one of the reasons why it is important to have a comprehensive set of synonyms for each concept. Table 1 shows the missing synonyms that were identified.

Table 1 Missing synonyms found for SNOMED CT concepts.

Text Entered by User	Concept Id	Concept FSN	Proposed Synonyms
cavus foot	36755004	Talipes cavus (disorder)	• Cavus foot
hypoplastic aortic arch	60787001	Congenital hypoplasia of aortic arch (disorder)	• Hypoplastic aortic arch
Rod Cone dystrophy	80328002	Progressive cone-rod dystrophy (disorder)	• Rod cone dystrophy
small muscular VSD	94706008	Muscular ventricular septum defect (disorder)	• Muscular VSD • Small muscular ventricular septum defect • Small muscular VSD
spontaneously closing VSD	123714004	Spontaneous closure of ventricular septal defect (finding)	• Spontaneously closing VSD
Bilateral hip dysplasia	205057007	Bilateral dysplastic hip (disorder)	• Bilateral hip dysplasia
tapered fingers	249768009	Tapering fingers (finding)	• Tapered fingers
short stature	418519005	Height below average (finding)	• Short stature
Poor growth/short stature	425115008	Abnormal height in relation to growth / age standard (finding)	• Poor growth / stature
Closure of ASD and VSD	699078008	Repair of atrioventricular septal defect with atrial and ventricular components (procedure)	• Repair of ASD and VSD
left vocal cord palsy	764724007	Paralysis of left vocal cord (disorder)	• Left vocal cord palsy
re-do sternotomy	3841000032100	Reopening of sternotomy (procedure)	• Redo sternotomy

4.1.2 Incorrect modelling

The autocomplete pattern implemented in REDCap also revealed modelling issues in SNOMED CT. The following sections describe these issues in detail.

Cavus foot

SNOMED CT contains three concepts for *pes cavus* as shown in Figures Figure 12, Figure 13 and Figure 14. One of the issues in these definitions is that the concepts are not named consistently. Each fully specified name uses a different variation of the main noun (e.g., *tapiles* cavus in Figure 12 and congenital *pes* cavus in Figure 13). The other noteworthy issue in the modelling is that *Congenital pes cavus* and *Acquired pes cavus* are not children of *Tapiles cavus*.

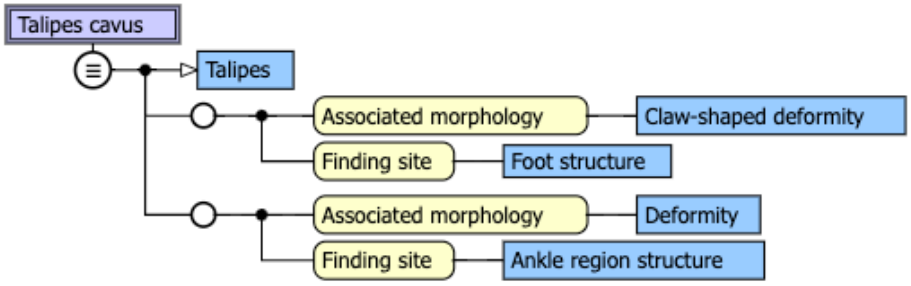

Figure 12 Talipes cavus concept definition in SNOMED CT.

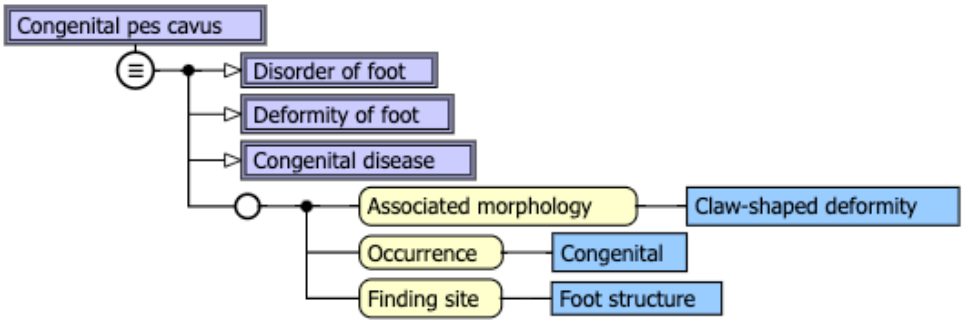

Figure 13 Congenital pes cavus concept definition in SNOMED CT.

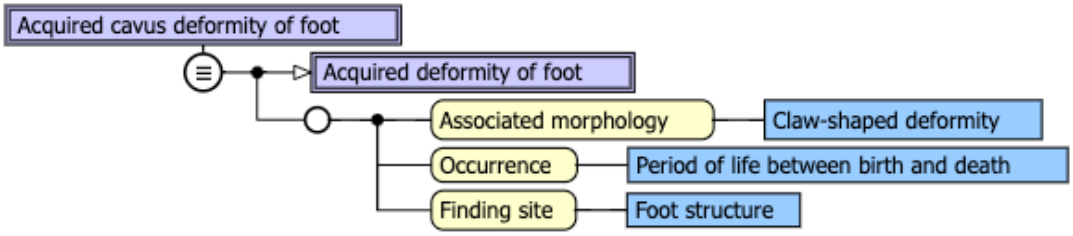

Figure 14 Acquired cavus deformity of foot concept definition in SNOMED CT.

Rod cone dystrophy

The SNOMED CT concept for *progressive cone-rod dystrophy*, shown in Figure 15, is not a child of *cone dystrophy* (Figure 16) nor *rod dystrophy* (Figure 17 Rod dystrophy definition in SNOMED CT. This is likely to be an issue with the definition of the finding site.

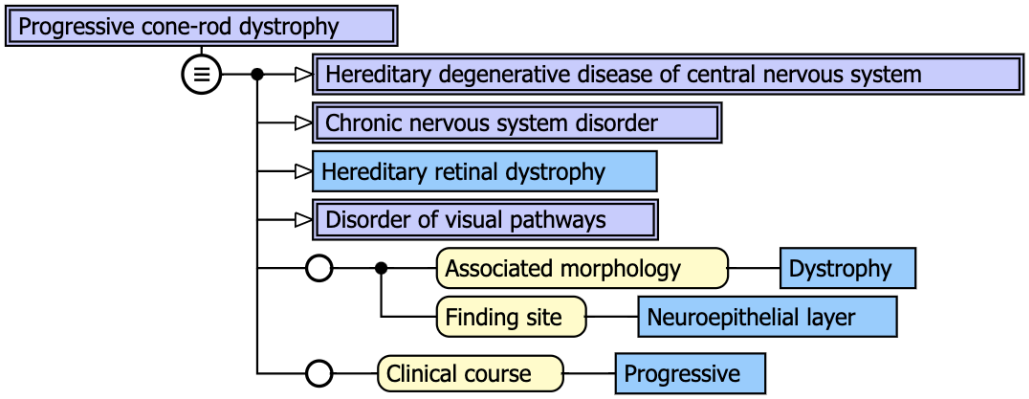

Figure 15 Progressive cone-rod dystrophy definition in SNOMED CT.

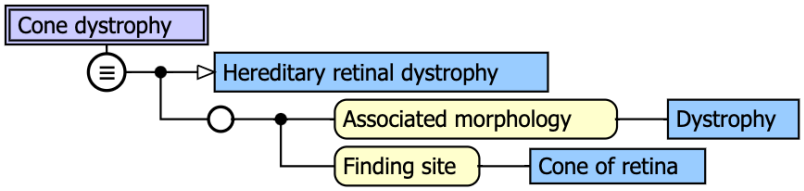

Figure 16 Cone dystrophy definition in SNOMED CT.

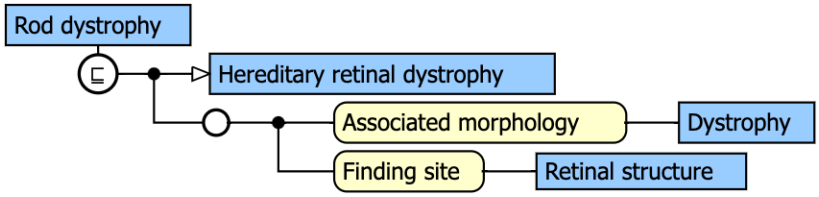

Figure 17 Rod dystrophy definition in SNOMED CT.

Mild hypoplasia of right ventricle

This concept is missing from SNOMED CT, as shown in Figure 18. The concept *Uhl's disease*, shown in Figure 19, has the synonym *Hypoplasia of right ventricle*, but is not part of this hierarchy. There are several things to note here. First, *Uhl's disease* is widely referred to in the literature as *Uhl's anomaly*, which is a synonym, but should probably be the preferred term.

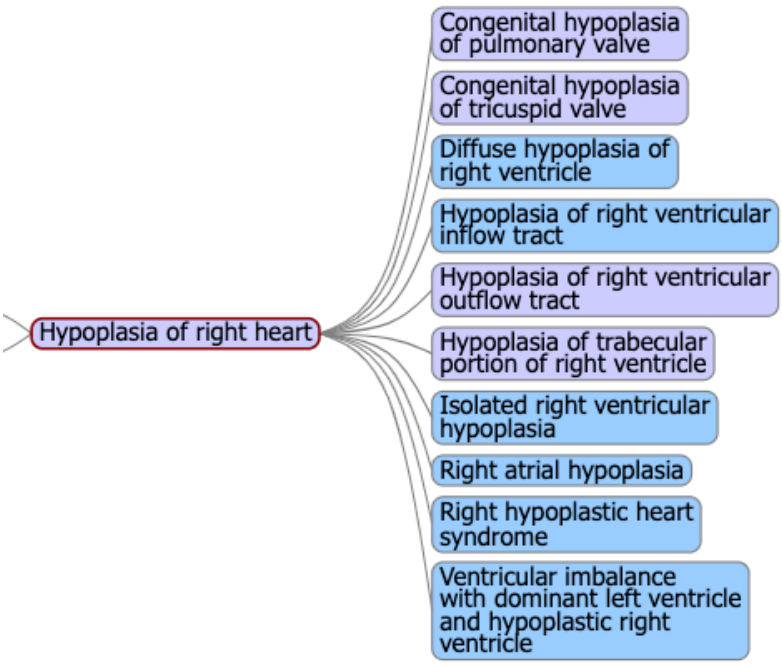

Figure 18 Children of Hypoplasia of right heart concept in SNOMED CT.

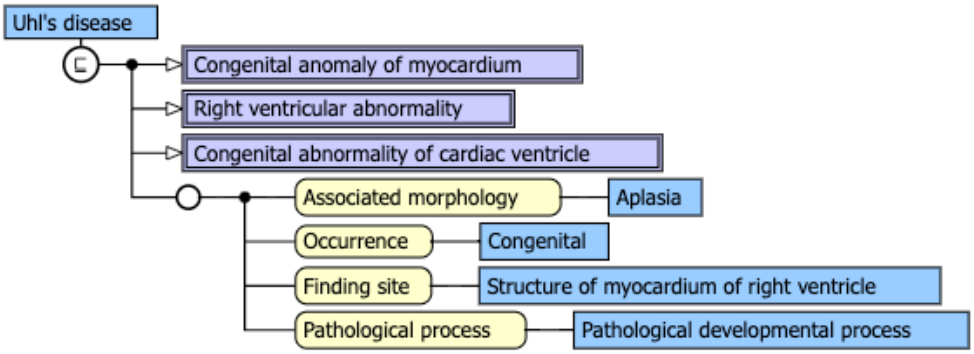

Figure 19 Ulh's disease definition in SNOMED CT.

Also, the associated morphology for *Uhl's disease* is *Aplasia*. *Aplasia* and *Dysplasia* are siblings, as shown in Figure 20. Given that these concepts are similar but *Hypoplasia* is less severe, this suggests that *Aplasia* might be a subtype of *Hypoplasia*. If this were the case then *Uhl's disease* could be made a child of either *Hypoplasia of right heart*, or a new, intermediate concept *Hypoplasia of right ventricle*.

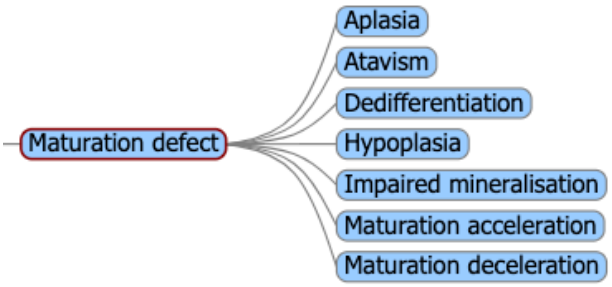

Figure 20 Maturation defect hierarchy in SNOMED CT.

Right auditory neuropathy

The SNOMED CT concept for *Auditory neuropathy spectrum disorder* is shown in Figure 21. There is no side differentiation, which is required in this case, which suggests additional concepts might be required. Also, this condition was originally termed *auditory neuropathy* and was then renamed to *Auditory Neuropathy / Auditory Dys-synchrony (AN/AD)*, before consensus was reached to rename it *Auditory neuropathy spectrum disorder*. Therefore, adding those names as synonyms will likely facilitate locating this concept when clinicians are familiar with the old names.

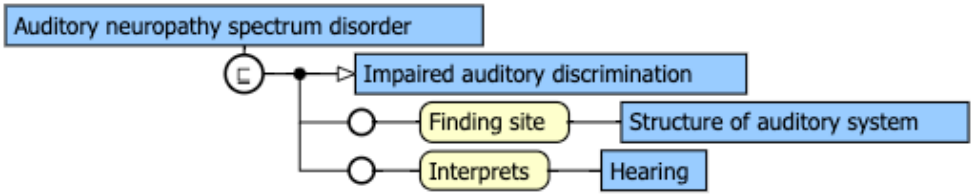

Figure 21 Auditory neuropathy spectrum disorder definition in SNOMED CT.

4.1.3 Missing concepts

The pattern used to capture coded data also allowed identification of missing concepts. There were three main areas where missing concepts were identified. Congenital Heart Disease Diagnosis, Congenital Heart Disease Procedures and Congenital Heart Disease Non-cardiac Anomalies value sets.

Table 2, Table 3 and Table 4 show the concepts that were not found in SNOMED CT in the Congenital Heart Disease Diagnosis, Congenital Heart Disease Procedures and Congenital Heart Disease Non-cardiac Anomalies value sets.

Table 2 CHD diagnosis concepts not found in SNOMED CT.

Text	Comment
Mild hypoplasia of right ventricle	The concept 39589002 | Hypoplasia of right heart (disorder) | refers to a more generalised site. The “mild” qualifier is also missing but this could be added in the information model.
possible progressive heart block	There is no corresponding concept. This would fit in the situation hierarchy.

Table 3 CHD procedure concepts not found in SNOMED CT.

Text	Comment
dacron patch closure of VSD	The SNOMED CT concept 80733000 | Closure of ventricular septal defect with patch (procedure) | matches “closure of VSD with patch” but the fact that it is a “dacron” patch is missing.
pericardial gusset repair of pulmonary artery	Part of this concept can be mapped to 175254001 | Repair of pulmonary artery using prosthesis (procedure) |, but there is no concept for “gusset repair”.
Stent angioplasty of recoarctation and RPA stenosis	The closest match is 233215007 | Placement of stent in pulmonary artery (procedure) |. There is no matching concept for the procedure as described.
balloon dilation of supra-valvular pulmonary valve stenosis	The concept 46897002 |Percutaneous balloon valvuloplasty of pulmonary valve (procedure) | refers to a more generalised site.
patent ductus arteriosus stenting by cath	The closest concept would be 233232006 | Maintenance of ductus arteriosus patency (procedure) |, but it does not specify “stenting by catheter”.
Yasui procedure and closure of atrial ventricular septal defect	No matching concept.
repair of partial AVSD	The closest concept would be 174836000 | Repair of defect of the atrioventricular septum (procedure) | but this does not refer to a partial AVSD.
Repair of partial atrioventricular canal defect	The closest concept would be 42582005 | Repair of atrioventricular canal defect (procedure) | but this is missing "partial", although there is a disorder concept for “partial atrioventricular canal defect”.
insertion of pulmonary homograft after previous repair of TOF	The closest concept would be 174944001 | Homograft pulmonary valve replacement (procedure) | but this is missing “previous TOF repair”.

Table 4 CHD non-cardiac anomalies concepts not found in SNOMED CT.

Text	Comment
multiple vertebral segmentation anomolies	The closest match is 205455005 | Defect of vertebral segmentation (disorder) |, but refers to a single anomaly.
Bifid right fifth rib	The closest match is 66102006 | Congenital fusion of ribs (disorder) |, but this concept does not define the specific rib as the site of fusion.
Hypoplastic alae nasi	No matching concept.
round blunted nasal tip	No matching concept, but SNOMED CT does have other concepts for 'appearance of nose' e.g.: “square nasal tip”.
unusual philtrum	No matching concept.
hypoplastic 5th fingers bilaterally	No matching concept.
5th toe on left foot hypoplastic with complete absence of nail	No matching concept.

4.2 Form / information model issues

4.2.1 Value set definition issues

Some of the cases where users where unable to find concepts using the autocomplete plugin were due to improperly defined value sets.

Congenital heart disease diagnosis

The value set used to capture congenital heart disease diagnoses, described in Section 3.2.7, was built by taking the union of all the children of the SNOMED CT concepts that correspond to the categories shown in Figure 22.

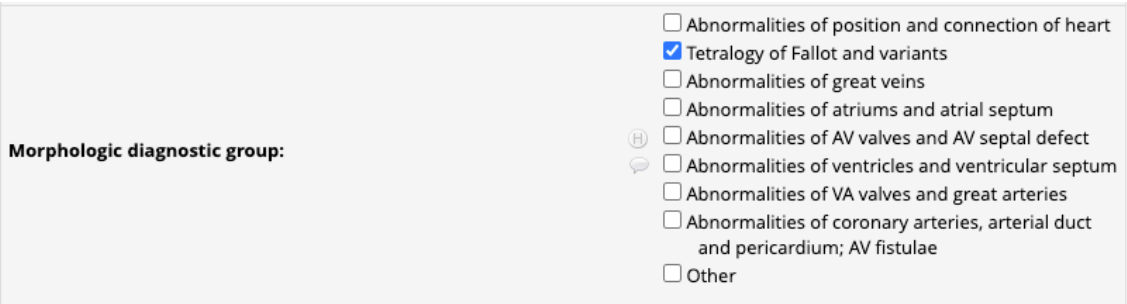

Figure 22 Categories of congenital heart disease diagnoses.

However, this union does not contain some of the concepts that were being searched for by users. These are listed in Table 5.

Table 5 SNOMED CT concepts missing from the original CHD diagnoses value set.

Concept Id	Concept FSN
62067003	Hypoplastic left heart syndrome (disorder)
72352009	Bicuspid aortic valve (disorder)
75372006	Congenital anomaly of mitral valve (disorder)
56786000	Pulmonic valve stenosis (disorder)
63042009	Congenital atresia of tricuspid valve (disorder)
123714004	Spontaneous closure of ventricular septal defect (finding)
7305005	Coarctation of aorta (disorder)

There are several reasons why this happened. First, the categories defined by the clinicians do not necessarily map one to one to the corresponding SNOMED CT concepts and this creates gaps. Also, some concepts that can be considered a congenital heart disease diagnosis are not necessarily always congenital and therefore will not be in the value set that was originally defined. One example is *pulmonary valve stenosis*, a condition that is commonly caused by congenital heart disease, but can also be caused by a malignant tumour.

Because of these issues the value set was updated to include the children of 106063007| Cardiovascular finding (finding) |, which is a concept higher up in the hierarchy. We believe that even though the search space is now larger, the quality of the data will not decrease and the new value set definition will cover all of the concepts that are relevant for this field.

Congenital heart disease non-cardiac anomalies

This value set, described in Section 3.2.8, is used to capture congenital heart disease non-cardiac anomalies and was built based on the categories shown in Figure 23.

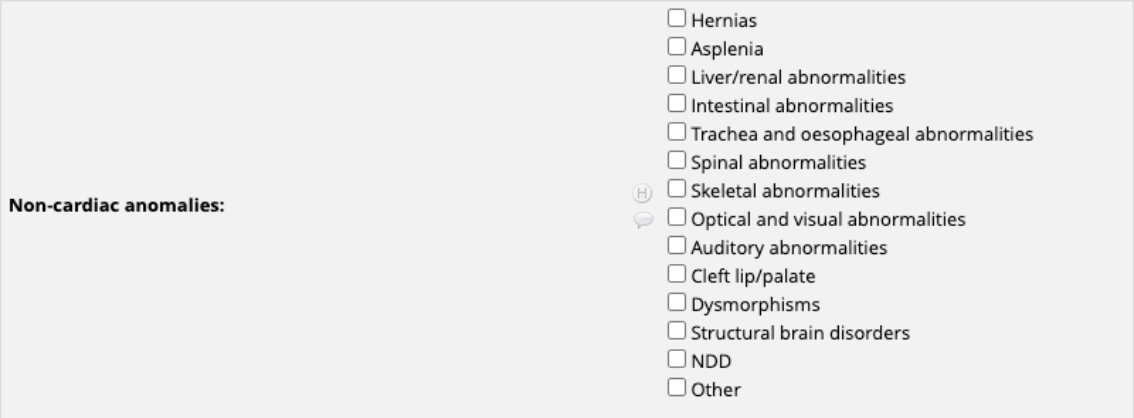

Figure 23 Categories of congenital heart disease non-cardiac anomalies.

However, again in this case, some concepts that clinicians searched for were not found in the value set, despite being in SNOMED CT. These are listed in Table 6.

Table 6 SNOMED CT concepts missing from the original CHD Non-cardiac Anomalies value set.

Concept Id	Concept FSN
29260007	Immunoglobulin A deficiency (disorder)
40650009	Obstruction of colon (disorder)
92937004	Congenital abnormal shape of pinna (disorder)
95515009	Low set ears (disorder)
161129001	Learning difficulties (finding)
205091006	Congenital pes cavus (disorder)
205455005	Defect of vertebral segmentation (disorder)
224958001	Global developmental delay (disorder)
248200007	Dysmorphic facies (finding)
249768009	Tapering fingers (finding)
298382003	Scoliosis deformity of spine (disorder)
413839001	Chronic lung disease (disorder)
418519005	Height below average (finding)
425601005	Unilateral neural hearing loss (situation)
764724007	Paralysis of left vocal cord (disorder)

There are several reasons for these gaps. First, the “other” category is problematic and was purposely not mapped when developing the value set because doing so would defeat the purpose of constraining the search space. Many of the missing concepts belong to this category. To solve this issue, the value set has been updated to include the descendants of the following concepts:

- 414029004 | Disorder of immune function (disorder) |
- 373930000 | Cognitive function finding (finding) |
- 417893002 | Deformity (finding) |
- 5294002 | Developmental disorder (disorder) |
- 50043002 | Disorder of respiratory system (disorder) |

Also, the original value set mapped the *dysmorphisms* category defined by the clinicians to the descendants of the SNOMED CT concept 276720006 | Dysmorphism |, but this concept has no descendants and its modelling should be reviewed. To solve this issue, the value set now includes the descendants of 276654001| Congenital malformation (disorder) | instead.

Finally, most categories were mapped to the descendants of the corresponding *congenital* concepts in SNOMED CT, but sometimes the users searched for non-congenital concepts (e.g., 205455005 | Defect of vertebral segmentation|). Also, as mentioned previously, in some cases disorders are typically congenital but can also have other causes. Therefore, the value set definition was updated and the descendants of congenital concepts were replaced with the descendants of one of the parent, non-congenital concepts, as shown in Table 7.

Table 7 Concepts used to replace the congenital versions used in the CHD non-cardiac anomalies value set.

Congenital concept	Non-congenital parent
89166001 | Congenital anomaly of liver (disorder) |	235856003 | Disorder of liver (disorder) |
44513007 | Congenital anomaly of the kidney (disorder) |	235856003 | Disorder of liver (disorder) |
31686000 | Congenital anomaly of lower alimentary tract (disorder) |	53619000 | Disorder of digestive system (disorder) |
69771008 | Congenital anomaly of esophagus (disorder) |	53619000 | Disorder of digestive system (disorder) |
14532008 | Congenital anomaly of trachea (disorder) |	50043002 | Disorder of respiratory system (disorder) |
74877002 | Congenital anomaly of spine (disorder) |	410730009 | Disorder of spinal region (disorder) |
73573004 | Congenital anomaly of musculoskeletal system (disorder) |	928000 | Disorder of musculoskeletal system (disorder) |
127329003 | Congenital anomaly of visual system (disorder) |	128127008 | Visual system disorder (disorder) |
95827002 | Congenital hearing disorder (disorder) |	362966006 | Disorder of auditory system (disorder) |
57148006 | Congenital anomaly of brain (disorder) |	81308009 | Disorder of brain (disorder) |

Congenital heart disease procedures

This value set is used to capture procedures related to heart disease. The only concept that could not be found in the value set was 3841000032100 | Reopening of sternotomy surgical site (procedure) |. The original definition of the value set, described in Section 3.2.9, includes only the descendants of 118672003 | Procedure on cardiovascular system |, but a reopening of a

sternotomy is clearly a relevant concept for heart disease. Therefore, the value set was updated to also include the descendants of 238327005 | Chest wall procedure (procedure) |.

4.2.2 Form issues

Many clinical concepts have qualifiers but not every combination can be represented as a precoordinated concept in a terminology. For example, a user searched for the concept “severe pulmonary stenosis”, and no corresponding concept was found in SNOMED CT. However, the concept for “pulmonary stenosis” is available (56786000 | Pulmonic valve stenosis |). In this case, the user interface should allow capturing the “severe” qualifier.

A similar issue was identified with suspected conditions. The value sets that were developed did not include concepts from the “Situation with explicit concept” hierarchy and therefore the user interface should allow a user to indicate that a concept is suspected.

The underlying FHIR representation, based on a FHIR implementation guide that is being developed for Australian Genomics, will support both of these attributes. However, because of the user interface limitations imposed by REDCap and the tension between depth and straightforward data entry, the ability to add these in the user interface was not implemented in this pilot.

4.3 Data entry issues

Some of the issues identified during the analysis were due to improper use of the autocomplete form elements. In some cases, after failing to find a concept, users would write complex descriptions that resemble stories rather than concepts. This is expected given that clinical notes tend to have this format. For example, the entry “GIT issues including colonic obstruction requiring resection” contains a significant amount of information. The elements that make up this “story” can be coded with the current REDCap form and guidance should be given to users around how to correctly code this kind of entry. However, the form doesn’t support capturing relationships between concepts, for example, indicating that the *resection* was due to *colonic obstruction*. This is due to the limitations imposed by the REDCap user interface.

This limitation was also the likely cause of another issue that was identified where a user added the free text entry “small size”, after failing to find a suitable concept, presumably to qualify the concept 85502002 | Bilateral inguinal hernia(disorder) |. Again in this case, guidance is needed to inform the users that it is currently not possible to qualify concepts.

5 Discussion

Three types of issues directly related to SNOMED CT were identified in the pilot: missing synonyms, missing concepts and incorrect modelling. Specific recommendations to improve SNOMED CT were given in the previous section. Issues were also identified in the value sets that were defined to bind some of the autocomplete fields in the REDCap form and these have been fixed as part of the project. The final versions of these value sets are available in Appendix A

The rest of the issues that were found are not related directly to SNOMED CT but are still important to consider when developing a system that uses a clinical terminology to capture coded data. One of the most interesting problems that was identified has to do with the difficulties of composing coded concepts when a precoordinated concept does not exist. Many clinicians are used to expressing complex concepts, more akin to stories, using natural language. Providing a user interface that allows doing this just as easily but instead using a computable, interoperable model, such FHIR with SNOMED CT, is very challenging and out of scope for this project, especially given the requirement of using REDCap.

Another interesting challenge has to do with the tension between the level of detail of the patient information being captured and the burden on the clinicians entering the data. This was handled through consultation with key members of the flagship and also with the use of the FHIR terminology autocomplete plugin and Ontoserver, which allows to quickly search for a concept using prefix-based search. This type of search has been shown to work well when dealing with large clinical terminologies (Sevenster, van Ommering, & Qian, 2012).

The pilot revealed that one major advantage of SNOMED CT over other terminologies, such as HPO, is its comprehensive coverage. HPO, for example, does not include any content around procedures and even though these are not technically part of a patient's phenotype, they are very important in this clinical area, and therefore capturing them in coded form is an advantage when considering doing data analysis in the future.

Finally, although not directly part of the pilot, it is worth mentioning that the decision to use SNOMED CT raised concerns among clinicians from other flagships that are currently using HPO. The main reason for this was that certain software tools, such as the Alissa curation platform[2], only work with HPO concepts. A SNOMED CT to HPO map that was under development as a separate project in Australian Genomics was used to overcome these issues. However, interoperability with other terminologies used in some niche areas is an important issue to consider. Even within Australian Genomics, it will be the case that some flagships will contain codes from other terminologies and this should be accounted for when doing analytics. A concept map is one strategy that can be useful in some scenarios, but it is important to keep these maps up to date.

[2] https://web.archive.org/web/20180630165951/https://www.agilent.com/en/products/software-informatics/clinical-informatics-(alissa-platform)/alissa-clinical-informatics-overview

6 Conclusion

The two main goals of this genomics pilot were assessing the suitability of SNOMED CT to do deep phenotyping of patients in the Cardiovascular Genetic Disorder Flagship and providing recommendations to improve SNOMED CT. The scope of the project also included the design of the REDCap forms used to collect patient data and the definition of the value sets bound to different autocomplete form elements.

The pilot has shown that SNOMED CT is suitable to do deep phenotyping in this clinical area. Only a small number of missing concepts and modelling issues were identified, mostly in the congenital heart disease space. However, there is room for improvement and recommendations have been made to include missing synonyms and concepts, and to review some of the current modelling.

Additional considerations around the use of SNOMED CT in electronic data capture systems were also discussed. These include existing challenges around coded data capture, such as the difficulties in creating an easy to use interface that allows capturing complex “stories” and the need to interoperate with other niche terminologies such as HPO. These insights can be used to inform future work.

7 Bibliography

Franklin, R., Jacobs, J., Tchervenkov, C., & Béland, M. (2002). European Paediatric Cardiac Code short list crossmapped to STS/EACTS short list with ICD-9 & ICD-10 crossmapping. *Cardiology in the Young, 12*, 23--49.

Metke-Jimenez, A., Steel, J., Hansen, D., & Lawley, M. (2018). Ontoserver: a syndicated terminology server. *Journal of Biomedical Semantics*.

Morales, J., Welter, D., Bowler, E. H., Cerezo, M., Harris, L. W., McMahon, A. C., . . . Parkinson, H. (2018). A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. *Genome Biology*.

Priori, S. G. (2013). HRS/EHRA/APHRS expert consensus statement on the diagnosis and management of patients with inherited primary arrhythmia syndromes. *Heart Rythm, 10*, 1932-1963.

Robinson, P. N. (2012). Deep phenotyping for precision medicine. *Human mutation*, 777-780.

Sevenster, M., van Ommering, R., & Qian, Y. (2012). Algorithmic and user study of an autocompletion algorithm on a large medical vocabulary. *Journal of Biomedical Informatics*, 107-119.

Appendix A Value Set Definitions

A.1 Coronary artery disease conditions


```
{
    "resourceType": "ValueSet",
    "id": "sct-cardiac-cc-cad",
    "text": {
      "status": "generated",
      "div": "<div xmlns=\"http://www.w3.org/1999/xhtml\"><p>Cardiac CAD Conditions
value set.</p></div>"
    },
    "url": "http://www.australiangenomics.org.au/cardiac/cc/cad",
    "version": "0.1.0",
    "name": "Cardiac Flagship CAD Conditions Value Set",
    "status": "draft",
    "compose": {
        "include": [
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "251015000"
                    }
                ]
            }
        ]
    }
}
```


A.2 Cancer types


```
{
    "resourceType": "ValueSet",
    "id": "sct-cardiac-cc-cancer",
    "text": {
      "status": "generated",
      "div": "<div xmlns=\"http://www.w3.org/1999/xhtml\"><p>Cardiac comorbidities
cancer types value set.</p></div>"
    },
    "url": "http://www.australiangenomics.org.au/cardiac/cc/cancer",
    "version": "0.1.0",
    "name": "Cardiac Flagship Comorbidities Cancer Types Value Set",
    "status": "draft",
    "compose": {
        "include": [
```

```
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "363346000"
                    }
                ]
            }
        ]
    }
}
```

A.3 Conduction abnormalities

```
{
    "resourceType": "ValueSet",
    "id": "sct-cardiac-cc-cond",
    "text": {
      "status": "generated",
      "div": "<div xmlns=\"http://www.w3.org/1999/xhtml\"><p>Cardiac comorbidities
conduction abnormalities value set.</p></div>"
    },
    "url": "http://www.australiangenomics.org.au/cardiac/cc/cond",
    "version": "0.1.0",
    "name": "Cardiac Flagship Comorbidities Conduction Abnormalities Value Set",
    "status": "draft",
    "compose": {
        "include": [
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "44808001"
                    }
                ]
            }
        ]
    }
}
```

A.4 Ventricular arrhythmias

```
{
    "resourceType": "ValueSet",
    "id": "sct-cardiac-cc-vent-arr",
    "text": {
      "status": "generated",
```

```
      "div": "<div xmlns=\"http://www.w3.org/1999/xhtml\"><p>Cardiac comorbidities
ventricular arrhythmias value set.</p></div>"
    },
    "url": "http://www.australiangenomics.org.au/cardiac/cc/vent/arr",
    "version": "0.1.0",
    "name": "Cardiac Flagship Comorbidities Ventricular Arrhythmias Value Set",
    "status": "draft",
    "compose": {
        "include": [
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "44103008"
                    }
                ]
            }
        ]
    }
}
```


A.5 Other conditions


```
{
    "resourceType": "ValueSet",
    "id": "sct-cardiac-cc",
    "text": {
      "status": "generated",
      "div": "<div xmlns=\"http://www.w3.org/1999/xhtml\"><p>Cardiac comorbidities value
set.</p></div>"
    },
    "url": "http://www.australiangenomics.org.au/cardiac/cc",
    "version": "0.1.0",
    "name": "Cardiac Flagship Comorbidities Value Set",
    "status": "draft",
    "compose": {
        "include": [
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "64572001"
                    }
                ]
            }
        ],
        "exclude": [
            {
                "system": "http://snomed.info/sct",
```

```
                "filter": [
                    {
                        "property": "concept",
                        "op": "is-a",
                        "value": "38341003"
                    }
                ]
            },
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "is-a",
                        "value": "73211009"
                    }
                ]
            },
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "is-a",
                        "value": "363346000"
                    }
                ]
            },
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "is-a",
                        "value": "53741008"
                    }
                ]
            }
        ]
    }
}
```

A.6 Echocardiogram valve disease

```
{
    "resourceType": "ValueSet",
    "id": "sct-cardiac-echo-vd",
    "text": {
      "status": "generated",
      "div": "<div xmlns=\"http://www.w3.org/1999/xhtml\"><p>Cardiac echo valve disease value set.</p></div>"
    },
    "url": "http://www.australiangenomics.org.au/cardiac/echo/vd",
```

```
    "version": "0.1.0",
    "name": "Cardiac Flagship Echo Valve Disease Value Set",
    "status": "draft",
    "compose": {
        "include": [
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "368009"
                    }
                ]
            }
        ]
    }
}
```


A.7 Congenital heart disease diagnosis


```
{
    "resourceType": "ValueSet",
    "id": "sct-cardiac-chd-dx",
    "text": {
      "status": "generated",
      "div": "<div xmlns=\"http://www.w3.org/1999/xhtml\"><p>Congenital heart disease
diagnosis value set.</p></div>"
    },
    "url": "http://www.australiangenomics.org.au/cardiac/congenital/dx",
    "version": "0.1.0",
    "name": "Congenital Heart Disease Diagnosis Value Set",
    "status": "draft",
    "compose": {
        "include": [
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "106063007 "
                    }
                ]
            }
        ]
    }
}
```

A.8 Congenital heart disease non-cardiac anomalies


```
{
    "resourceType": "ValueSet",
    "id": "sct-cardiac-chd-nc-anom",
    "text": {
      "status": "generated",
      "div": "<div xmlns=\"http://www.w3.org/1999/xhtml\"><p>Congenital heart disease
non-cardiac anomalies value set.</p></div>"
    },
    "url": "http://www.australiangenomics.org.au/cardiac/congenital/nc/anom",
    "version": "0.1.0",
    "name": "Congenital Heart Disease Non-cardiac Anomalies Value Set",
    "status": "draft",
    "compose": {
        "include": [
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "is-a",
                        "value": "39839004"
                    }
                ]
            },{
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "is-a",
                        "value": "40775003"
                    }
                ]
            },{
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "is-a",
                        "value": "396232000"
                    }
                ]
            },{
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "707147002"
                    }
                ]
            },
```

```
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "235856003"
                    }
                ]
            },
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "235856003"
                    }
                ]
            },
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "53619000"
                    }
                ]
            },
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "53619000"
                    }
                ]
            },
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "50043002"
                    }
                ]
            },
            {
                "system": "http://snomed.info/sct",
                "filter": [
```

```
                {
                    "property": "concept",
                    "op": "descendent-of",
                    "value": "410730009"
                }
            ]
        },
        {
            "system": "http://snomed.info/sct",
            "filter": [
                {
                    "property": "concept",
                    "op": "descendent-of",
                    "value": "928000"
                }
            ]
        },
        {
            "system": "http://snomed.info/sct",
            "filter": [
                {
                    "property": "concept",
                    "op": "descendent-of",
                    "value": "128127008"
                }
            ]
        },
        {
            "system": "http://snomed.info/sct",
            "filter": [
                {
                    "property": "concept",
                    "op": "descendent-of",
                    "value": "362966006"
                }
            ]
        },
        {
            "system": "http://snomed.info/sct",
            "filter": [
                {
                    "property": "concept",
                    "op": "descendent-of",
                    "value": "87979003"
                }
            ]
        },
        {
            "system": "http://snomed.info/sct",
            "filter": [
                {
                    "property": "concept",
                    "op": "descendent-of",
```

```
                    "value": "80281008"
                }
            ]
        },
        {
            "system": "http://snomed.info/sct",
            "filter": [
                {
                    "property": "concept",
                    "op": "descendent-of",
                    "value": "276654001"
                }
            ]
        },
        {
            "system": "http://snomed.info/sct",
            "filter": [
                {
                    "property": "concept",
                    "op": "descendent-of",
                    "value": "81308009"
                }
            ]
        },
        {
            "system": "http://snomed.info/sct",
            "filter": [
                {
                    "property": "concept",
                    "op": "descendent-of",
                    "value": "700364009"
                }
            ]
        },
        {
            "system": "http://snomed.info/sct",
            "filter": [
                {
                    "property": "concept",
                    "op": "descendent-of",
                    "value": "414029004"
                }
            ]
        },
        {
            "system": "http://snomed.info/sct",
            "filter": [
                {
                    "property": "concept",
                    "op": "descendent-of",
                    "value": "373930000"
                }
            ]
```

```
            },
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "417893002"
                    }
                ]
            },
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "5294002"
                    }
                ]
            },
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "50043002"
                    }
                ]
            }
        ]
    }
}
```

A.9 Congenital heart disease procedures

```
{
    "resourceType": "ValueSet",
    "id": "sct-cardiac-chd-px",
    "text": {
      "status": "generated",
      "div": "<div xmlns=\"http://www.w3.org/1999/xhtml\"><p>Congenital heart disease
procedures value set.</p></div>"
    },
    "url": "http://www.australiangenomics.org.au/cardiac/congenital/px",
    "version": "0.1.0",
    "name": "Congenital Heart Disease Procedures Value Set",
    "status": "draft",
    "compose": {
        "include": [
            {
                "system": "http://snomed.info/sct",
```

```
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "118672003"
                    }
                ]
            },
            {
                "system": "http://snomed.info/sct",
                "filter": [
                    {
                        "property": "concept",
                        "op": "descendent-of",
                        "value": "238327005"
                    }
                ]
            }
        ]
    }
}
```

CONTACT US

t 1300 363 400
+61 3 9545 2176
e csiroenquiries@csiro.au
w www.csiro.au

AT CSIRO, WE DO THE EXTRAORDINARY EVERY DAY

We innovate for tomorrow and help improve today – for our customers, all Australians and the world.

Our innovations contribute billions of dollars to the Australian economy every year. As the largest patent holder in the nation, our vast wealth of intellectual property has led to more than 150 spin-off companies.

With more than 5,000 experts and a burning desire to get things done, we are Australia's catalyst for innovation.

CSIRO. WE IMAGINE. WE COLLABORATE. WE INNOVATE.

FOR FURTHER INFORMATION

The Australian eHealth Research Centre
Alejandro Metke-Jimenez
t +61 7 3253 3645
e alejandro.metke@csiro.au
w aehrc.com