NHS England 100,000 Genomes Project
SNOMED CT Rare Diseases Datasets Project

National and International Review - Report

Version: 0.5
Date: 23 Dec 2016

Contacts:

All questions and issues relating to the content detailed within this document should be forwarded to the following contact details:

Name: Monica Jones
Company: Animo Consultancy Ltd
Telephone: +44 1274 653462 / +44 7876 403693
Email: mcmjones@animoconsultancy.co.uk
monica.jones@nhs.net [secure data]
Web: www.animoconsultancy.co.uk

Document History:

Version	Date	Status	Author
0.1	1 Nov 2016	Initial Version	Monica Jones
0.2	9 Nov 2016	Amended version following publications research	Monica Jones
0.3	21 Nov 2016	Update following IHTSDO meeting	Monica Jones
0.4	9 Dec 2016	Update following North Thames GMC / GOSH meeting	Monica Jones
0.5	23 Dec 2016	Further updates (from follow up)	Monica Jones

This is a controlled document. On receipt of a new version please destroy all previous versions (unless a specified earlier version is in use for a particular stage of the Project).

Circulation List:

Name	Organisation
Kathy Farndon	NHS England
Angie Quinn	NHS England

Contents

National and International Review

Genomics Rare Diseases and Clinical Terminologies

Background

SNOMED CT is the world's most comprehensive terminology for electronic health information. SNOMED CT contributes to the improvement of patient care by underpinning the development of Electronic Health Records (EHR) that record clinical information in ways that enable meaning-based retrieval. This provides effective access to information required for decision support and consistent reporting and analysis.

Patients benefit from the use of SNOMED CT because it improves the recording of EHR information and facilitates better communication, leading to improvements in the quality of care. SNOMED CT is owned, maintained and distributed by the International Health Terminology Standard Development Organisation (IHTSDO); a not-for-profit association which is governed by its national members. Each country has a National Release Centre (NRC). In the UK, the NRC is based in Leeds (under NHS Digital's administration) and is called the UK Terminology Centre (UKTC).

The Genomics SNOMED CT Project covers England, but with close links to other UK devolved nations. As defined in the Project Initiation Document (PID) there have been discussions around the links between National and International SNOMED CT development and from a specialty perspective it was felt that the following would be appropriate avenues to seek possible linkage and shared learning.

Objective

Two of the agreed project deliverables are:

- The National and International search and review to identify whether any country has developed rare disease subsets. Support in terms of providing insight into previous related project work terminology development
- Investigation of the existing work being undertaken between Orphanet and IHTSDO to develop SNOMED CT Rare diseases subsets and whether this work could be utilised for use in the 100,000 Genome Project

This was completed by performing an environmental scan for activities related to terminology systems and the genomics rare diseases domain. The survey includes bibliographic searches and internet resources, as well as finding from face to face meetings and teleconferences. This report presents the findings of this survey and literature review.

Methods

Using the search terms (HPO) Human Phenotype Ontology map to SNOMED CT? and Development of Rare Diseases subsets - work on terminology? it was possible to identify projects which are engaged in terminology development addressing encoding of genomics - rare diseases, clinical terminology and SNOMED CT. The bibliography was created using PubMed, EuroPubMed and Orphanet. Contact was also made with IHTSDO (Clinical Engagement) and [North Thames] Genomics Medicine Centre[1].

Findings

Queries were made of known organisations involved in the genomics - rare diseases domain and medical terminology and classification developers. In addition, the results from academic journals are incorporated. In a review of those organisations engaged in terminology development and publication, there is coverage of some aspects of the genomic and rare disease domain. This report is presented in three parts: an overview, a table summarising activities by organisation, a bibliography and appendix for review.

The IHTSDO Management Board proposal was presented at the beginning of the year (a copy of this report was made available). It was agreed that IHTSDO needs a strategy for genomics and are putting together an expert group. The approach was agreed to be good and has assigned funding towards it. They are recruiting a genomics clinician (request for help is out currently – RFP across Europe).

Issues include:

- Currently there is no single source of the truth for genomics
- IHTSDO do not want to add in the whole of HPO into SNOMED CT
- There needs to be an agreement with HPO to define a value set somewhere else (RF2 allows these to link to attributes from SNOMED CT)
- SNOMED CT gives the content in the Electronic Health Record (EHR)
- Partnering opportunities with other genomics organisations and funders such as NIH

IHTSDO has an agreement with INSERM (based in Paris) to incorporate the Orphanet database into SNOMED CT. It does not have a text description for all rare diseases but these can be developed. It requires authoring of 2500-3000 terms per release. Intend to add at diagnosis level with modelling of concepts. INSERM have their priority list and it will be two and a half years before it is complete. IHTSDO intend to create an equivalence table.

This project has requested access to what has been put in already and process will be set up (INSERM online database is in the public domain). There will be in addition access to the codes that are being added to SNOMED CT. There are two ways of doing this, access to the daily build browser which will show you what is being added as they are being added. The other is to give access to the spreadsheets that are submitted by INSERM and then reviewed. The latter is preferred.

[1] The North Thames NHS Genomic Medicine Centre (NTGMC) is part of the Government's 100,000 Genomes Project which aims to sequence, or analyse, 100,000 genomes from participants with cancer, rare disorders and infectious diseases. There are currently 13 regional GMCs in England.

Biggest areas are childhood disorders and links have been made with GOSH. A meeting was held with key stakeholders on 5 December and agreements to closely align work moving forward was agreed. IHTSDO would like to put childhood rare diseases into SNOMED CT. This could be time boxed into SNOMED CT subsets (a constrained set) and is to be scoped.

Internationally there is work some work going on. Australian digital health (Dave Bunker in Queensland). Others include Spain, Netherlands and USA (University of Nebraska Hospital) – could be joint NIH funded. These are also being explored.

Conclusion

The National and International review concluded that there are a number of developments happening internationally but none that would duplicate the work of the NHS England subset development project. There are however some very useful citations that should be cross referenced and follow up contacts that should be included within the communications plan moving forward.

We will share project scope with the IHTSDO, UKTC and GEL as planned and work closely with these organisations. We should also link up with WHO / ISO and LOINC as required. In the review of the bibliography, researchers could be approached from the abstracts of their work, both are addressing terminology and classification development.

Terminology Development and Management Organisations

Organisation	Country	Project Description	Contact Details (if available)
Regenstrief Institute	International	LOINC is available international standard for tests, measurements, and observations	http://www.regenstrief.org/resources/loinc/ Daniel Vreeman - Director, Loinc And Health Data Standards dvreeman@loinc.org
World Health Organization	International	WHO Informatics and technology committee International Classification of Health Interventions	http://www.who.int/classifications/committees/itc/en http://www.who.int/classifications/ichi/en/
International Health Terminology Standards Development Organisation (IHTSDO)	International	SNOMED CT	www.ihtsdo.org Ian Green - Clinical Engagement and Education Services Business Manager igr@ihtsdo.org

Appendix A: Bibliographic Citations

Reference:	Review & Evaluate the following areas
PubMed (Online)	(HPO) Human Phenotype Ontology map to SNOMED CT?
	Development of Rare Diseases subsets - work on terminology?

Library	
https://www.ncbi.nlm.nih.gov/pubmed?term=Human%20Phenotype%20Ontology%20%26%20SNOMED%20CT	1
	Interoperability between phenotypes in research and healthcare terminologies--Investigating partial mappings between HPO and SNOMED CT. Dhombres F, Bodenreider O. J Biomed Semantics. 2016 Feb 9;7:3. doi: 10.1186/s13326-016-0047-3. PMID: 26865946 Free PMC Article Similar articles
	2
	SORTA: a system for ontology-based re-coding and technical annotation of biomedical phenotypedata. Pang C, Sollie A, Sijtsma A, Hendriksen D, Charbon B, de Haan M, de Boer T, Kelpin F, Jetten J, van der Velde JK, Smidt N, Sijmons R, Hillege H, Swertz MA. Database (Oxford). 2015 Sep 18;2015. pii: bav089. doi: 10.1093/database/bav089. PMID: 26385205 Free PMC Article Similar articles
	3
	Extending the coverage of phenotypes in SNOMED CT through post-coordination. Dhombres F, Winnenburg R, Case JT, Bodenreider O. Stud Health Technol Inform. 2015;216:795-9. PMID: 26262161 Similar articles

https://www.ncbi.nlm.nih.gov/bioproject?term=CT%5BRare%20Diseases%5D&cmd=DetailsSearch	4
Search: Bioproject CT (Rare Diseases)	Women's Health Initiative (WHI) Organism: Homo sapiens Taxonomy: Homo sapiens (human) Project data type: Phenotype or Genotype

	Scope: Multiisolate FRED HUTCHINSON CANCER RESEARCH CENTER Accession: PRJNA312995 ID: 312995

Search: PubMed journal on datasets on rare diseases	5
https://www.ncbi.nlm.nih.gov/pubmed?term=journal+on+datasets+on+rare+diseases&cmd=DetailsSearch	Sul JH, Cade BE, Cho MH, Qiao D, Silverman EK, Redline S, Sunyaev S. Am J Hum Genet. 2016 Oct 6;99(4):846-859. doi: 10.1016/j.ajhg.2016.08.015. PMID: 27666371 Similar articles
	6 Epidemiology of severe anaphylaxis: can we use population-based data to understand anaphylaxis? Turner PJ, Campbell DE. Curr Opin Allergy Clin Immunol. 2016 Oct;16(5):441-50. doi: 10.1097/ACI.0000000000000305. PMID: 27490124 Similar articles
	7 Molecular digital pathology: progress and potential of exchanging molecular data. Roy S, Pfeifer JD, LaFramboise WA, Pantanowitz L. Expert Rev Mol Diagn. 2016 Sep;16(9):941-7. doi: 10.1080/14737159.2016.1206472. PMID: 27471996 Similar articles
	8 Design, set-up and utility of the UK facioscapulohumeral muscular dystrophy patient registry. Evangelista T, Wood L, Fernandez-Torron R, Williams M, Smith D, Lunt P, Hudson J, Norwood F, Orrell R, Willis T, Hilton-Jones D, Rafferty K, Guglieri M, Lochmüller H. J Neurol. 2016 Jul;263(7):1401-8. doi: 10.1007/s00415-016-8132-1. PMID: 27159994 Similar articles

9

RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data.

Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ.
Bioinformatics. 2016 May 1;32(9):1423-6. doi: 10.1093/bioinformatics/btw079.
PMID: 27153000

Similar articles

10

A Systematic Evaluation of Feature Selection and Classification Algorithms Using Simulated and Real miRNA Sequencing Data.

Yang S, Guo L, Shao F, Zhao Y, Chen F.
Comput Math Methods Med. 2015;2015:178572. doi: 10.1155/2015/178572.
PMID: 26508990

Free PMC Article

Similar articles

11

Strategies for Imputing and Analyzing Rare Variants in Association Studies.

Hoffmann TJ, Witte JS.
Trends Genet. 2015 Oct;31(10):556-63. doi: 10.1016/j.tig.2015.07.006. Review.
PMID: 26450338

Free PMC Article

Similar articles

12

The NIH BD2K center for big data in translational genomics.

Paten B, Diekhans M, Druker BJ, Friend S, Guinney J, Gassner N, Guttman M, Kent WJ, Mantey P, Margolin AA, Massie M, Novak AM, Nothaft F, Pachter L, Patterson D, Smuga-Otto M, Stuart JM, Van't Veer L, Wold B, Haussler D.
J Am Med Inform Assoc. 2015 Nov;22(6):1143-7. doi: 10.1093/jamia/ocv047.
PMID: 26174866

Free PMC Article

Similar articles

13

Identifying personal microbiomes using metagenomic codes.

Franzosa EA, Huang K, Meadow JF, Gevers D, Lemon KP, Bohannan BJ, Huttenhower C.
Proc Natl Acad Sci U S A. 2015 Jun 2;112(22):E2930-8. doi: 10.1073/pnas.1423854112.
PMID: 25964341

Similar articles

14

A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research.

Choquet R, Maaroufi M, de Carrara A, Messiaen C, Luigi E, Landais P.

J Am Med Inform Assoc. 2015 Jan;22(1):76-85. doi: 10.1136/amiajnl-2014-002794.

PMID: 25038198

Free PMC Article

Similar articles

15

The Rare Diseases Clinical Research Network's organization and approach to observational research and health outcomes research.

Krischer JP, Gopal-Srivastava R, Groft SC, Eckstein DJ; Rare Diseases Clinical Research Network.

J Gen Intern Med. 2014 Aug;29 Suppl 3:S739-44. doi: 10.1007/s11606-014-2894-x.

PMID: 25029976

Free PMC Article

Similar articles

16

Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users.

Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S.

Hum Mutat. 2012 May;33(5):803-8. doi: 10.1002/humu.22078.

PMID: 22422702

Similar articles

Reference:	Review & Evaluate the following areas
Europubmed (Online)	(HPO) Human Phenotype Ontology map to SNOMED CT?
	Development of Rare Diseases subsets - work on terminology?

Library	
http://europepmc.org/search?query=Human%20Phenotype%20Ontology%20%26%20SNOMED%20CT	Interoperability between phenotypes in research and healthcare terminologies--Investigating partial mappings between HPO and SNOMED CT. (PMID:26865946 PMCID:PMC4748471) Free full text article Dhombres F, Bodenreider O, J Biomed Semantics [2016, 7:3] Cited: 0 times
	NEO: Systematic Non-Lattice Embedding of Ontologies for Comparing the Subsumption Relationship in SNOMED CT and in FMA Using MapReduce. (PMID:26306275 PMCID:PMC4525277) Free full text article Zhu W et al , AMIA Jt Summits Transl Sci Proc [2015, 2015:216-220] Cited: 1 time
	Extending the coverage of phenotypes in SNOMED CT through post-coordination. (PMID:26262161) Dhombres F , Winnenburg R , Case JT , Bodenreider O Stud Health Technol Inform [2015, 216:795-799] Cited: 0 times
	SORTA: a system for ontology-based re-coding and technical annotation of biomedical phenotype data. (PMID:26385205 PMCID:PMC4574036) Free full text article Pang C et al , Database (Oxford) [2015, 2015] Cited: 1 time
	A comparative analysis of the density of the SNOMED CT conceptual content for semantic harmonization.

	(PMID:25890688 PMCID:PMC4457611) Free full text article He Z et al , Artif Intell Med [2015, 64(1):29-40] Cited: 1 time
	Categorization of allergic disorders in the new World Health Organization International Classification of Diseases. (PMID:25905010 PMCID:PMC4405839) Free full text article Tanno LK et al , Clin Transl Allergy [2014, 4:42] Cited: 6 times
	Coverage of rare disease names in standard terminologies and implications for patients, providers, and research. (PMID:25954361 PMCID:PMC4419993) Free full text article Fung KW et al , AMIA Annu Symp Proc [2014, 2014:564-572] Cited: 0 times
	Next generation phenotyping using the unified medical language system. (PMID:25601137 PMCID:PMC4288084) Free full text article Adamusiak T et al JMIR Med Inform [2014, 2(1):e5] Cited: 2 times
	Literature review of SNOMED CT use. (PMID:23828173 PMCID:PMC3957381) Free full text article Lee D et al , J Am Med Inform Assoc [2014, 21(e1):e11-9] Cited: 5 times
	Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: the eMERGE Network experience. (PMID:21597104 PMCID:PMC3128396) Free full text article Pathak J et al,

	J Am Med Inform Assoc [2011, 18(4):376-386] Cited: 36 times
	Leveraging terminological resources for mapping between rare disease information sources. (PMID:23920611 PMCID:PMC4296515) Free full text article Rance B et al , Stud Health Technol Inform [2013, 192:529-533] Cited: 2 times
	Terminological mapping for high throughput comparative biology of phenotypes. (PMID:14992504 PMCID:PMC2891021) Free full text article Lussier YA and Li J, Pac Symp Biocomput [2004:202-213] Cited: 9 times

Reference:	Review & Evaluate the following areas
Orphanet (Online)	(HPO) Human Phenotype Ontology map to SNOMED CT?
	Development of Rare Diseases subsets - work on terminology?

Library	
https://ojrd.biomedcentral.com/articles?query=Human+phenotype+ontology+%26+SNOMED+CT&searchType=journalSearch&tab=keyword	Rare diseases in ICD11: making rare diseases visible in health information systems through appropriate coding Because of their individual rarity, genetic diseases and other types of rare diseases are under-represented in healthcare coding systems; this contributes to a lack of ascertainment and recognition of their impact. **Ségolène Aymé, Bertrand Bellet and Ana Rath** *Orphanet Journal of Rare Diseases* 2015 **10**:35 Published on: 26 March 2015
Reference to WHO not SNOMED CT	
This is a subset but not SNOMED CT	A comprehensive database of Duchenne and Becker muscular dystrophy patients (0–18 years old) in East China Currently, there is no cure for Duchenne and Becker muscular dystrophies (DMD/BMD). However, clinical trials with new therapeutic strategies are being conducted or considered. A comprehensive database is criti... **Xihua Li, Lei Zhao, Shuizhen Zhou, Chaoping Hu, Yiyun Shi, Wei Shi, Hui Li, Fang Liu, Bingbing Wu and Yi Wang** *Orphanet Journal of Rare Diseases* 2015 **10**:5 Published on: 23 January 2015