Deliverable D5 (D5.1): Semantic models of key clinical conditions and outcome measures

Date

30th November 2016

Status

Final submission

Distribution

Public

Authors

Harshana Liyanage, Stacy Shinneman, Filipa Ferreira, Simon de Lusignan

Contents

DEFINITIONS

- **MIROI** – the MOCHA Research Opportunity Instrument is a survey instrument used to collect essential metadata (data about data) from health databases.
- **Semantic model** – provides detail about the meaning of the network of concepts and their relationships described in our ontologies, enabling their precise definition (denotation).
- **Ontology** – an explicit specification of a conceptualisation.
- **OWL** – Ontology Web Language is a semantic web language designed to represent rich and complex knowledge about things, groups of things, and relations between things.
- **Protégé** – a software tool used for authoring ontologies.
- **Measure** - a defined and specific process for assessment (should contain aspects such as What, Who, When, How, Where).
- **Indicator** - the result of relating a measure to policy target, a standard, or to its position within a distribution; a measure set against a benchmark scientifically grounded.
- **Health indicator** - measure of a health problem.
- **System indicator** - measure of the most important aspects of primary care functioning divided over the seven domains identified by Kringos, et al. (2013)[1]. The information that could be extracted when having data on these system indicators could then function as a measure of primary care functioning.
- **Outcome** - quality of life, health status or characteristic (including determinants of health) of a patient or population that results from a treatment or program.

Chapter 1 - Introduction

1.1 Background and Rationale

In the first year of the MOCHA project we identified case-based databases with the potential to provide data to allow the appraisal of child primary health services. This was achieved by using the MIROI (MOCHA International Research Opportunity) survey instrument which collected essential metadata describing the responding databases. (Appendix C lists the questions featured in the MIROI.) The list of databases to approach for inclusion was compiled from expert knowledge, literature and other projects, and from the Country Agents which the MOCHA project has retained in each of its 30 study countries.[1]

This deliverable focuses on creating semantic models that enable us to assess the feasibility of databases to participate in specific studies conducted within MOCHA.

Semantic interoperability and semantic models

Within medical informatics a key concept is one of semantic interoperability. This is where data with a precise meaning in one computerised medical record (CMR) system is transferred to another CMR while precisely preserving its meaning. Whilst we don't plan to make data sources semantically interoperable, we do want definitions planned to appraise child health systems to be subject to semantic, more precise definitions. This is to enable us to take account of how differing access or processes in different health systems might impact on whether or how a condition might be recorded. Developing ontologies to model semantics, which require precise definitions of concepts and their relationships, is a key part of this process.

For the topic specific profiling and subsequent data requests to be meaningful, we need to have more precise definitions than simply the name of the condition, indicators and outcome measures. We need instead to develop greater semantic precision – developing the requirements of individual subject leads using our three-step ontological process.[2] Whilst the comparisons required in MOCHA may require many variables, the core elements of any study are cases and outcome measures. Depending on what is being measured, the population denominator can be of vital importance. Hence its inclusion in this deliverable.

This deliverable takes us forward a step further towards actually making comparisons. The databases we have identified through MIROI contain different types of data in a variety of structures and different granularities. We look to identify how different data sources might be harmonised to support international collaborations. To do this, our methodology explored whether three key elements of data are available:

[1] MOCHA community on EMIF Web Catalogue: https://emif-catalogue.eu/c/mocha1

1. **Population denominator.** The size of the population from which the cases are drawn. This is important because it is not possible to calculate incidence or prevalence without it. The age-sex structure of a population also allows standardisation of rates, either against a standard population (e.g. European standard population) or against another study population group. If a specific study requires additional population characteristics (such as ethnicity), we will consider incorporating them as well.
2. **Case definitions.** Case definitions were identified and generalised to develop ontologies capturing key concepts and their relationships. A range of case definitions were to be considered in order to identify cases from different datasets. There were, for example, case definitions based on diagnostic labels, on test results, or based on therapy prescribed. We built upon the systematic reviews conducted in WP1 as the foundation for our case definitions. For each case definition we looked for precise semantic meaning of concepts associated to clinical conditions to maximise the chance that data from different sources are semantically interoperable – from our perspective, reusable between studies. This approach can utilise technical approaches designed to achieve semantic interoperability between health care computer systems.[3]
3. **Outcome measures.** As in the case definitions, there will not just be a single outcome measure. There will need to be a range of outcome measures, to allow for the different scope of the databases and study requirements. The outcome measures are anticipated to cover the six quality domains identified by the Institute of Medicine, in the United States.[4] These areas are:
    a. **Safe:** Avoiding harm to patients from the care that is intended to help them.
    b. **Timely:** Reducing waits and sometimes harmful delays for both those who receive and those who give care.
    c. **Effective:** Providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit (avoiding underuse and misuse, respectively).
    d. **Efficient:** Avoiding waste, including waste of equipment, supplies, ideas, and energy.
    e. **Equitable:** Providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status.
    f. **Patient-centred:** Providing care that is respectful of and responsive to individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions.

We anticipate that not all case and outcome measure definitions will be available from all data sets. However, where comparisons are made between child health systems, the case definitions and outcome measures must have gone through a rigorous process to try to ensure comparability and availability at the required level of granularity. The process involves assessing the semantic models described in this deliverable with data available at the data source. The outcome measures will be embedded within the semantic model presented in this deliverable. The study teams will be able to identify specific outcome measures that will be used to assess databases during the feasibility assessment process.

This deliverable explains the process for developing semantic models (i.e. ontologies), seeing how those concepts are represented in clinical coding systems, which will lead to the development of topic specific questionnaires to explore which databases can provide data to make the proposed comparisons.

Finally, a fourth element – the intervention or exposure - can often be a vital part of any epidemiological or observational investigation. Whilst not formally part of this deliverable, we will look to include exposure, and potentially other comparison requirements, within the ontologies identified while working with MOCHA colleagues. Where feasible, this will be included within topic specific questionnaires we send to data sources regarding their data requirements.

1.2 Methods

1.2.1 Process overview

The purpose of this process is to explore the extent to which the data sources we have identified using the MOCHA International Research Opportunities Instrument (MIROI) can be utilised to make comparisons between models or to otherwise appraise different models of child health. The process will have two phases:

Phase 1: Semantic models that assist to explore if MOCHA data sources can meet research needs

This initial phase entail the development of semantic models that support exploring the data sources potential to enable the planned appraisal of different models of child health. This involves converting the concepts (often attributes of a disease or condition) into an ontology which is then tested to see if it can be mapped to clinical codes. We will then explore whether the required codes for these variables are likely to be available in our data sources. This phase will utilise the data collected so far using MIROI and follow-up feasibility assessment with selected databases. The outcome will be that all, some or none of the variables required are likely to be available.

In Phase 1 the steps in the method associated include:

(1) **Population denominator requirements** and the potential for such data to be available from participant data sources.

(2) **Development of semantic models as ontologies and semantic definitions for the key variables of interest** - the ontological process will help provide semantic precision about each variable.[5] This process identifies the key concepts and their relationships that define cases and outcome measures of interest, and potentially other variables.

The ontologies will be formally recorded in OWL (Ontology Web Language) using Protégé software. The associated concepts (diagnosis, symptoms, test results, therapies or treatments, health-related behaviours and process of care) are detailed in a hierarchical manner within each of the four ontologies.

Phase 2: Feasibility assessment of databases to explore the potential for data sharing
Development of semantic models in this deliverable will enable study specific enquiries to be raised with the other participating data sources. Development of semantic models will be evaluated against selected databases by taking the ontologies for variables of interest and exploring how these might be mapped to the coding and classification systems[6] used in the databases. Most databases will have clinically-coded data, and although we prefer key variables to be defined ontologically (i.e. conceptually), it is worth considering that any definition likely to be used in practice must be capable of being mapped to the data available, and this will mainly be coded.[7] The ontological mapping annotation of the data source will be an iterative process where mapping would take place against all potential databases short-listed from the MOCHA web catalogue which contains responses for the MIROI survey. A decision will be made as to whether the MOCHA data sources have the potential to answer all or any of the intended questions.

1.2.2 Phase 1: Developing semantic models

- **Population denominator requirements**

Whilst not part of the process of development of a semantic model, each variable will be checked to ascertain whether it requires a population denominator – and with what degree of granularity. The degree of granularity might be crude population, for a stated age-band, or particular breakdown of data such as five-year age-sex bands.

Population denominator

The denominator includes all persons at-risk for the disease or condition, i.e. disease-free or condition-free individuals in the population at the start of the time period.[8]

The population denominator requirements for each of the key variables may be required to give an incidence or prevalence rate, or for standardisation against a reference or other population.

- **Development of semantic models and definitions for the key variables of interest**

This step, which is the focus of this deliverable, involves creating ontologies for the key study variables – "a case" and an "outcome." Precise definitions will be created and then disaggregated into constituent concepts. For example: a case might be defined by (1) Diagnosis of the condition; (2) A test result associated with having the condition; (3) Treatment used to manage that condition; or (4) Business process records – like disease notification or claiming a fee for conducting a procedure. The ontology may include a generalised set of concepts that will encapsulate all identified types of case definitions.

In this task, we will use clinical concepts identified above to develop ontologies for the three conditions of interest. The ontologies will be developed using the OWL (Web Ontology Language) standards using Protégé as the Ontology building environment.

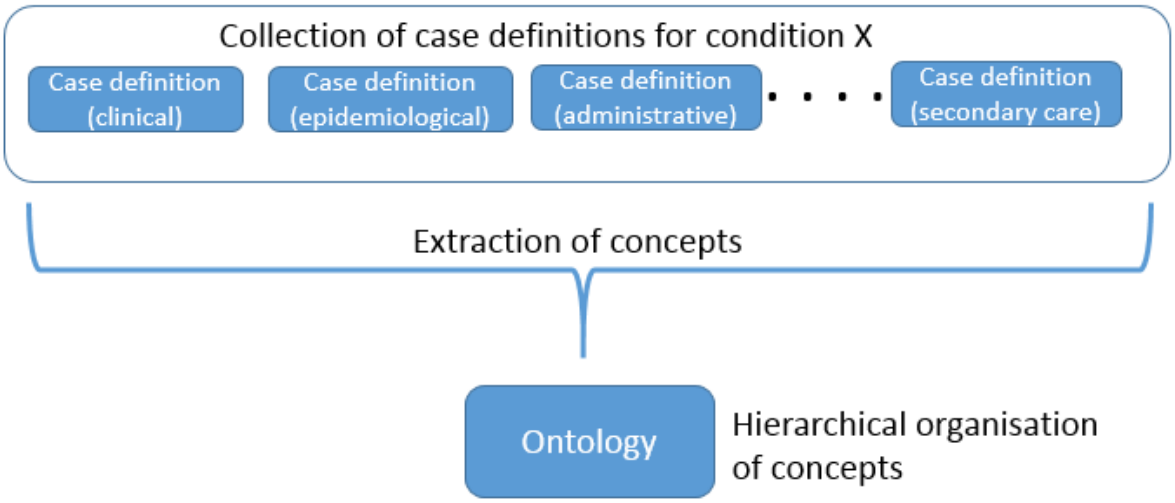

Figure 1 – Process of developing ontologies

The ontologies developed in this step will be compared subsequently with the likely available data from data sources that have completed the MIROI survey. The ontologies will also contain concepts related to outcome measures that will possibly exist in a database. The assessment of the ontology will be carried for all databases that can potentially support studies as indicated through the MIROI responses.

It is likely that most matching at this stage will be between clinical concept and how data are recorded in the database. Generally research databases contain coded clinical data. Where data are coded, the concepts will be mapped to a data dictionary for the given coding system. Where the match is poor, other data sources can be looked for. Additionally, if there is poor matching because the level of granularity is wrong, the ontology can be revisited. Researchers can make as many iterations as they wish.

As our MIROI databases for supporting studies are heterogeneous and have data at different granularities, it is possible that different databases and revision of variable definitions will take place.

Once the final ontologies and associated code lists are created, it will then be possible to proceed to Phase 2 (beyond the scope of this deliverable) where study specific versions of MIROI are developed and sent to eligible data sources.

1.2.3 Phase 2: Feasibility assessment of databases to explore the potential for data sharing

The process of assessing and engaging databases is illustrated in the flow diagram below. In this deliverable we will focus on the components that will support the feasibility assessment of databases. The distributed data analysis is not covered in this deliverable.

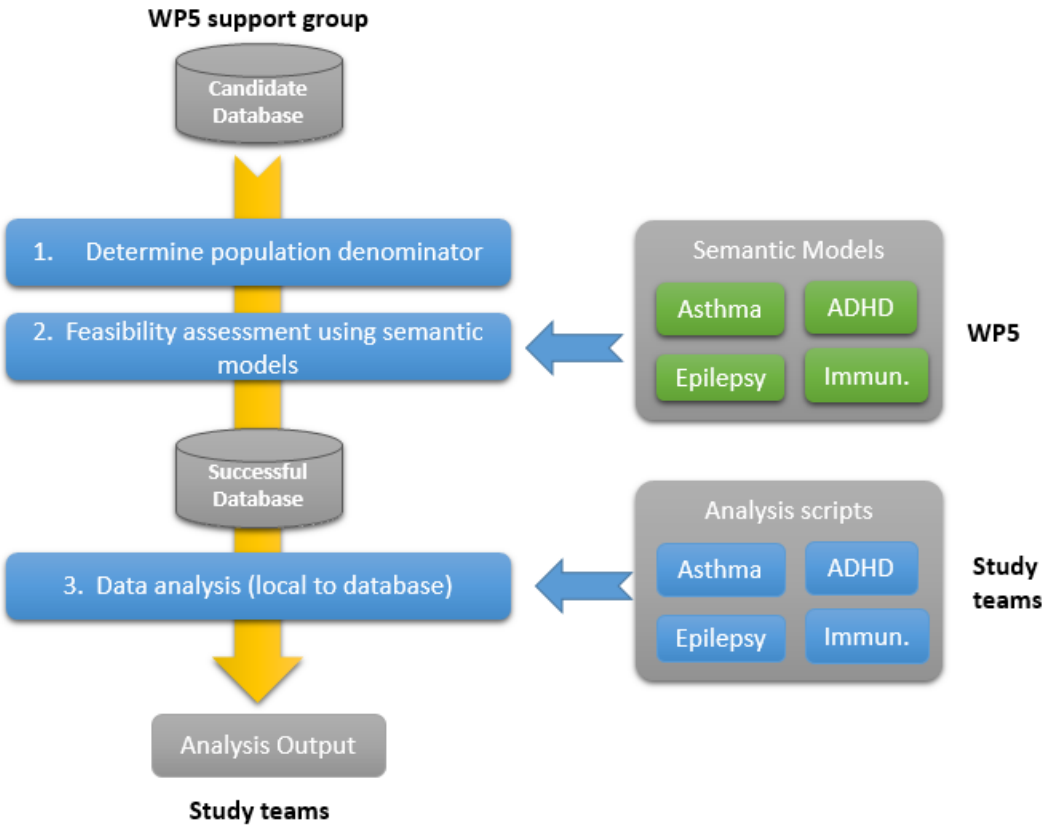

Figure 2 - Outline process flow

The breakdown of the task in this process can be specified as follows:

1) Identify candidate database (using the initial general MIROI survey)
2) Assess feasibility to provide datasets to support studies related to one or more index conditions using semantic models and related study specific surveys
   a) Identify population denominator
   b) Identify which of the case definitions each data source can identify
   c) Identify which of the outcome definitions each database can report for each condition
   d) Develop study specific instruments for distribution to selected databases
   e) Report the child health model in place
3) Conduct distributed analysis

It is important to note that a database could be considered as a successful candidate if it is determined to be feasible to support investigation of one or more index conditions.

1.3 About semantic models presented in this deliverable

The subsequent chapters of this deliverable will focus on specific clinical conditions of interest within the MOCHA project. Each chapter will contain a collection of case definitions used for building the ontology for the clinical condition. Additionally, the chapters introduce a disease ontology and a disease outcome measure ontology.

The 8 ontologies developed for this deliverable and their concept (class) count is given in the table below.

Ontology	Concept (class) count
1. Asthma	286
2. Asthma outcome measures	95
3. Epilepsy	129
4. Epilepsy outcome measures	63
5. ADHD	154
6. ADHD outcome measures	98
7. Immunisation	7845
8. Immunisation outcome measures	31

The ontologies provided can be viewed by using the desktop version of the Protégé software (Appendix A) or as online ontologies accessible through the Web Protégé application (Appendix B). A brief description of OWL is given in Appendix D.

Chapter 2 – Semantic model for asthma

In this chapter, we present the various case definitions that were considered for developing the ontology for asthma. A diverse range of definitions have been used to ensure that wide range of asthma related concepts are included in the ontology. The source from which the definition was obtained is also indicated along with each definition. The resulting ontology is hosted online as a web resource and the link to the ontology is provided at the end of this chapter.

2.1 Asthma case definitions

a) MOCHA Asthma case definition

Reference:

MOCHA - Work Package 1: Identification of models of children's primary care: Systematic Review and Meta-analysis of the Literature – Part 2

Asthma is a heterogeneous (many variations) disease, usually characterized by chronic airway inflammation. It is defined by the history of respiratory symptoms such as wheeze, shortness of breath, chest tightness and cough that vary over time and in intensity, together with variable expiratory airflow limitation.

b) US National Institutes of Health case definition

Reference:

National Asthma Education and Prevention Program's Expert Panel Report 3 (EPR-3): Guidelines for the Diagnosis and Management of Asthma-Summary Report 2007. The Journal of allergy and clinical immunology. 2007;120(5 Suppl):S94-13 – pg 6

http://www.nhlbi.nih.gov/health-pro/guidelines/current/asthma-guidelines/full-report

Asthma is a chronic inflammatory disorder of the airways. In susceptible individuals, this inflammation causes recurrent episodes of coughing (particularly at night or early in the morning), wheezing, breathlessness, and chest tightness. These episodes are usually associated with widespread but variable airflow obstruction that is often reversible either spontaneously or with treatment.

Asthma is a complex disorder characterized by variable and recurring symptoms, airflow obstruction, bronchial hyperresponsiveness, and an underlying inflammation. The interaction of these features determines the clinical manifestations and severity of asthma and the response to treatment. Airway inflammation (and therefore airway limitation) is caused by: bronchoconstriction, airway hyperresponsiveness and airway oedema.

Episodic symptoms of airflow obstruction or airway hyperresponsiveness are presented. Airflow obstruction is at least partially reversible, measured by spirometry. Reversibility is determined by an increase in FEV1 of >200 mL and 12% from baseline measure after inhalation of short-acting b2-agonist (SABA).

Use medical history and physical examination to determine that symptoms of recurrent episodes of airflow obstruction are present. Use spirometry in all patients >5 years of age to determine that airway obstruction is at least partially reversible. Use severity classification chart and asthma control chart to determine asthma severity and determine treatment.

Asthma is highly variable over time. The National Heart, Lung, and Blood Institute (NHLBI) asthma severity classification scale accounts for the progressive nature of asthma by measuring it across the dimensions of types of symptoms and lung function:

- Mild intermittent
- Mild persistent
- Moderate persistent
- Severe persistent

c) WHO definition

Reference:

WHO Chronic respiratory diseases – asthma definition:
http://www.who.int/respiratory/asthma/definition/en/

Asthma attacks all age groups but often starts in childhood. It is a disease characterized by recurrent attacks of breathlessness and wheezing, which vary in severity and frequency from person to person. In an individual, they may occur from hour to hour and day to day.

This condition is due to inflammation of the air passages in the lungs and affects the sensitivity of the nerve endings in the airways so they become easily irritated. In an attack, the lining of the passages swell causing the airways to narrow and reducing the flow of air in and out of the lungs.

d) Global Initiative for Asthma (GINA) case definition

Reference:

Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention, 2016

www.ginasthma.org

Asthma causes symptoms such as wheezing, shortness of breath, chest tightness and cough that vary over time in their occurrence, frequency and intensity. These symptoms are associated with variable expiratory airflow, i.e. difficulty breathing air out of the lungs due to bronchoconstriction (airway narrowing), airway wall thickening, and increased mucus. Asthma can be caused by viral infections, domestic or occupational allergens (e.g. house dust mite, pollens, cockroach), tobacco smoke, exercise, stress and some drugs. Asthma has two key defining features:

- a history of respiratory symptoms such as wheeze, shortness of breath, chest tightness and cough that vary over time and in intensity, AND

- variable expiratory airflow limitation.

DIAGNOSTIC FEATURE	CRITERIA FOR MAKING THE DIAGNOSIS OF ASTHMA
1. History of variable respiratory symptoms	
Wheeze, shortness of breath, chest tightness and cough Descriptors may vary between cultures and by age, e.g. children may be described as having heavy breathing	• Generally more than one type of respiratory symptom (in adults, isolated cough is seldom due to asthma) • Symptoms occur variably over time and vary in intensity • Symptoms are often worse at night or on waking • Symptoms are often triggered by exercise, laughter, allergens, cold air • Symptoms often appear or worsen with viral infections
2. Confirmed variable expiratory airflow limitation	
Documented excessive variability in lung function* (one or more of the tests below)	The greater the variations, or the more occasions excess variation is seen, the more confident the diagnosis
AND documented airflow limitation*	At least once during diagnostic process when FEV_1 is low, confirm that FEV_1/FVC is reduced (normally >0.75–0.80 in adults, >0.90 in children)
Positive bronchodilator (BD) reversibility test* (more likely to be positive if BD medication is withheld before test: SABA ≥4 hours, LABA ≥15 hours)	*Adults*: increase in FEV_1 of >12% and >200 mL from baseline, 10–15 minutes after 200–400 mcg albuterol or equivalent (greater confidence if increase is >15% and >400 mL). *Children*: increase in FEV_1 of >12% predicted
Excessive variability in twice-daily PEF over 2 weeks*	*Adults*: average daily diurnal PEF variability >10%** *Children*: average daily diurnal PEF variability >13%**
Significant increase in lung function after 4 weeks of anti-inflammatory treatment	*Adults*: increase in FEV_1 by >12% and >200 mL (or PEF† by >20%) from baseline after 4 weeks of treatment, outside respiratory infections
Positive exercise challenge test*	*Adults*: fall in FEV_1 of >10% and >200 mL from baseline *Children*: fall in FEV_1 of >12% predicted, or PEF >15%
Positive bronchial challenge test (usually only performed in adults)	Fall in FEV_1 from baseline of ≥20% with standard doses of methacholine or histamine, or ≥15% with standardized hyperventilation, hypertonic saline or mannitol challenge
Excessive variation in lung function between visits* (less reliable)	*Adults*: variation in FEV_1 of >12% and >200 mL between visits, outside of respiratory infections *Children*: variation in FEV_1 of >12% in FEV_1 or >15% in PEF† between visits (may include respiratory infections)

Source: http://ginasthma.org/wp-content/uploads/2016/04/GINA-2016-main-report_tracked.pdf

e) International Classification of Diseases definition

Reference:

http://www.icd10data.com/ICD10CM/Codes/J00-J99/J40-J47/J45-/

A chronic disease in which the bronchial airways in the lungs become narrowed and swollen, making it difficult to breathe. Symptoms include wheezing, coughing, tightness in the chest, shortness of breath, and rapid breathing. An attack may be brought on by pet hair, dust, smoke, pollen, mold, exercise, cold air, or stress.

Clinical Information

- A chronic disease in which the bronchial airways in the lungs become narrowed and swollen, making it difficult to breathe. Symptoms include wheezing, coughing, tightness in the chest, shortness of breath, and rapid breathing. An attack may be brought on by pet hair, dust, smoke, pollen, mold, exercise, cold air, or stress.

- A chronic respiratory disease manifested as difficulty breathing due to the narrowing of bronchial passageways.
- A form of bronchial disorder with three distinct components: airway hyper-responsiveness (respiratory hypersensitivity), airway inflammation, and intermittent airway obstruction. It is characterized by spasmodic contraction of airway smooth muscle, wheezing, and dyspnea (dyspnea, paroxysmal).
- Asthma is a chronic disease that affects your airways. Your airways are tubes that carry air in and out of your lungs. If you have asthma, the inside walls of your airways become sore and swollen. That makes them very sensitive, and they may react strongly to things that you are allergic to or find irritating. When your airways react, they get narrower and your lungs get less air. Symptoms of asthma include:
    - wheezing
    - coughing, especially early in the morning or at night
    - chest tightness
    - shortness of breath

  Not all people who have asthma have these symptoms. Having these symptoms doesn't always mean that you have asthma. Your doctor will diagnose asthma based on lung function tests, your medical history, and a physical exam. You may also have allergy tests. When your asthma symptoms become worse than usual, it's called an asthma attack. Severe asthma attacks may require emergency care, and they can be fatal. Asthma is treated with two kinds of medicines: quick-relief medicines to stop asthma symptoms and long-term control medicines to prevent symptoms.

- Form of bronchial disorder associated with airway obstruction, marked by recurrent attacks of paroxysmal dyspnea, with wheezing due to spasmodic contraction of the bronchi.

Use additional code to identify:

- exposure to environmental tobacco smoke
- exposure to tobacco smoke in the perinatal period
- history of tobacco dependence
- occupational exposure to environmental tobacco
- tobacco dependence
- tobacco use

f) Clinical asthma case definition

Reference:

Trepka MJ, Martin P, Mavunda K, Rodriguez D, Zhang G, Brown C. A pilot asthma incidence surveillance system and case definition: lessons learned. Public health reports (Washington, DC : 1974). 2009;124(2):267-79.

• Health-care professional diagnosis of asthma, reactive airway disease, hyperreactive airway disease, or wheezing-related respiratory illness (or chronic bronchitis if patient is pediatric)

• Symptoms (on symptom list) that improve with treatment at least once (see medication list) unless health-care professional has diagnosed an alternative diagnosis as causing symptoms (see list below)

• Medication: taking at least one rescue and one controller (see medication list)

• Laboratory criteria: 12% increase in FEV1 or FVC after the patient inhales a short-acting bronchodilator or 20% decrease in FEV1 after exercise challenge.

g) An ontological asthma case definition

Reference:

Afzal Z, Engelkes M, Verhamme KM, Janssens HM, Sturkenboom MC, Kors JA, et al. Automatic generation of case-detection algorithms to identify children with asthma from large electronic health record databases. Pharmacoepidemiology and drug safety. 2013;22(8):826-33.

For definite asthma patients, there was at least one entry in their medical record containing an asthma diagnosis confirmed by a specialist (paediatrician or pulmonologist).

For probable asthma patients, one entry contains evidence of asthma diagnosed by the GP, and there was at least one more entry in the patient record suggestive of asthma (ICPC code, free text, or use of specific bronchodilating drugs/anti-inflammatory drugs for the indication of asthma) within the next 12 months, or there are at least two additional entries in the patient record suggestive of asthma. Use of bronchodilating drugs only did not fit these criteria.

For doubtful asthma patients, there were one or more entries containing an indication or evidence of asthma, but they do not satisfy the criteria for a definite or probable asthma case.

A patient is a non-asthma case if there was no indication of asthma in any entry of the patient record.

2.2 Asthma ontology

The asthma ontology capturing the concepts from the asthma definitions can be found at the following web link:

Asthma ontology: http://rebrand.ly/mocha-asthma

The upper level concept organisation in the ontology is given in Figure 3.

Figure 3 – Upper level concepts of the asthma ontology

2.3 Asthma outcomes ontology

The asthma outcome measures have been organised according to the six quality domains suggested by the Institute of Medicine (given in detail in Chapter 1).

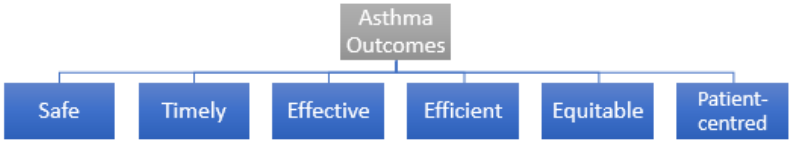

Figure 4 – Upper level concepts of the asthma outcomes ontology

2.4 Assessment instrument for understanding measures of health care quality for asthma

A selected subset of the asthma outcome measures have been used to develop a part of the study specific feasibility assessment instrument which can be used to assess the ability of a chosen database to provide the corresponding outcome. For example, the following grid will be completed by the database indicating if each outcome measure is “Not recorded”, “Partially recorded” or “Completely recorded”.

Outcome measures for asthma	Not recorded	Partial recording	Complete recording
Safety			
1. Asthma related deaths	☐	☐	☐
2. Medication adverse events	☐	☐	☐
Intervention efficacy			
1. Preventer/ reliever ratio	☐	☐	☐
2. Oral steroids	☐	☐	☐
3. Normal growth			
Patient centred approach			
1. Individual management plan	☐	☐	☐
2. Self-monitoring of peak flow	☐	☐	☐
Timeliness of care			
1. Patient waiting times for specialist review	☐	☐	☐
Efficiency			
1. Medication wasted (not taken, or not dispensed)	☐	☐	☐
Equitability			
1. Ethnicity	☐	☐	☐
2. Socioeconomic status	☐	☐	☐

Chapter 3 – Semantic model for epilepsy

In this chapter, we present the various case definitions that were considered for developing the ontology for epilepsy. A diverse range of definitions have been used to ensure that wide range of epilepsy related concepts are included in the ontology. The source from which the definition was obtained is also indicated along with each definition. The resulting ontology is hosted online as a web resource and the link to the ontology is provided at the end of this chapter.

3.1 Epilepsy case definitions

a) Clinical definition

Reference:

Fisher RS, Acevedo C, Arzimanoglou A, Bogacz A, Cross JH, Elger CE, Engel J Jr, Forsgren L, French JA, Glynn M, Hesdorffer DC, Lee BI, Mathern GW, Moshé SL, Perucca E, Scheffer IE, Tomson T, Watanabe M, Wiebe S. ILAE official report: a practical clinical definition of epilepsy. Epilepsia. 2014 Apr;55(4):475-82. doi: 10.1111/epi.12550.

In 2005, a Task Force of the International League Against Epilepsy (ILAE) formulated conceptual definitions of “seizure” and “epilepsy”:

An epileptic seizure is a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain.

Epilepsy is a disorder of the brain characterized by an enduring predisposition to generate epileptic seizures, and by the neurobiologic, cognitive, psychological, and social consequences of this condition. The definition of epilepsy requires the occurrence of at least one epileptic seizure

Epilepsy is a disease of the brain defined by any of the following conditions

1. At least two unprovoked (or reflex) seizures occurring > 24 h apart

2. One unprovoked (or reflex) seizure and a probability of further seizures similar to the general recurrence risk (at least 60%) after two unprovoked seizures, occurring over the next 10 years

3. Diagnosis of an epilepsy syndrome.

Epilepsy is considered to be resolved for individuals who had an age-dependent epilepsy syndrome but are now past the applicable age or those who have remained seizure-free for the last 10 years, with no seizure medicines for the last 5 years

b) Epidemiological definition

Definition by Thurman et al.

Reference:

Thurman DJ, Beghi E, Begley CE, Berg AT, Buchhalter JR, Ding D, et al. Standards for epidemiologic studies and surveillance of epilepsy. Epilepsia 2011;52(Suppl. 7): 2–26. http://dx.doi.org/10.1111/j.1528-1167.2011.03121.x.

Probable diagnosis of epilepsy:

One of the following 3 conditions are satisfied:

1. one medical encounter with a 3-digit code of G40.x (epilepsy);
2. ≥2medical encounters on separate days coded with G41 (status epilepticus)or with a 4-digit code R56.8 (other and unspecified convulsions);
3. and a single medical encounter coded as other and unspecified convulsions (R56.8) and an antiepileptic drug prescription for three or more months.

Suspected diagnosis of epilepsy:

A single episodes coded with R56.8 or G41.

Definition by Wilson et al.

Reference:

Tan M, Wilson I, Braganza V, Ignatiadis S, Boston R, Sundararajan V, Cook MJ, D'Souza WJ. Development and validation of an epidemiologic case definition of epilepsy for use with routinely collected Australian health data. Epilepsy Behav. 2015 Oct;51:65-72.

Potential diagnosis of epilepsy

Patients coded with:

- epilepsy (ICD-10AMG40.xx)
- status epilepticus (G41.xx)
- other and unspecified convulsions (R56.8x), and
- acquired aphasia with epilepsy (F80.3x).

c) Epilepsy definition – using claims data

Reference:

Bakaki PM, Koroukian SM, Jackson LW, Albert JM, Kaiboriboon K. Defining incident cases of epilepsy in administrative data. Epilepsy Res. 2013 Sep;106(1-2):273-9. doi: 0.1016/j.eplepsyres.2013.05.005.

Individuals were identified as having epilepsy if they met all of the following criteria:

1. At least 1 visit with an epilepsy diagnosis; or at least 2 visits, on different dates, with a diagnosis of non-febrile convulsions. The epilepsy onset or epilepsy index date was determined as the date of the first diagnosis of epilepsy or the second diagnosis of non-febrile convulsion.
2. At a minimum of 30 days after epilepsy index date, there was at least 1 more visits related to epilepsy or non-febrile convulsions.
3. A minimum of 2 pharmacy dispensing claims, at least 30 days apart subsequent to the epilepsy index date, for any of the following AEDs: carbamazepine, ethosuximide, felbamate, gabapentin, lamotrigine, levetiracetam, oxcarbazepine, phenobarbital, phenytoin, pregabalin, primidone, tiagabine, topiramate, valproicacid, or zonisamide.

d) Case definitions based on data source

Reference:

Thurman DJ, Beghi E, Begley CE, Berg AT, Buchhalter JR, Ding D, Hesdorffer DC, Hauser WA, Kazis L, Kobau R, Kroner B, Labiner D, Liow K, Logroscino G, Medina MT, Newton CR, Parko K, Paschal A, Preux PM, Sander JW, Selassie A, Theodore W, Tomson T, Wiebe S; ILAE Commission on Epidemiology.. Standards for epidemiologic studies and surveillance of epilepsy. Epilepsia. 2011 Sep;52 Suppl 7:2-26. doi: 10.1111/j.1528-1167.2011.03121.x.

Using data obtained by trained health care provider (interview or medical records)

Definite

• clear evidence of two or more unprovoked epileptic seizures that have occurred over interval(s) exceeding 24 h, OR

• confirmed diagnosis of epilepsy by a health care provider with appropriate specialized training in the recognition and treatment of epilepsy.

Probable

• documentation of a diagnosis of epilepsy by a trained non-specialist health care provider without specific documentation of definite criteria above.

Suspect

• data suggest a possibility of epilepsy but criteria for definite or probable epilepsy are not met. The information provided is inadequate to confirm or refute the diagnosis of epilepsy.

Using population survey data collected by non-clinician interviewers

Probable

• respondent (subject or proxy) reports that a physician or trained health care provider has diagnosed epilepsy(probable).

Suspect

• information provided suggests a possibility of epilepsy but is inadequate to confirm or refute the diagnosis of epilepsy.

e) Epilepsy definition - using existing coded health data (International Classification of Diseases)

Reference:

Thurman DJ, Beghi E, Begley CE, Berg AT, Buchhalter JR, Ding D, Hesdorffer DC, Hauser WA, Kazis L, Kobau R, Kroner B, Labiner D, Liow K, Logroscino G, Medina MT, Newton CR, Parko K, Paschal A, Preux PM, Sander JW, Selassie A, Theodore W, Tomson T, Wiebe S; ILAE Commission on Epidemiology.. Standards for epidemiologic studies and surveillance of epilepsy. Epilepsia. 2011 Sep;52 Suppl 7:2-26. doi: 10.1111/j.1528-1167.2011.03121.x.

Note:

(1) The specificity and positive predictive values of ICD-coded medical encounter data have been shown to vary among studies of epilepsy in different localities (Holden et al., 2005a; Jette et al., 2010). The following scheme is suggested as rough guidance where only coded data are available. An evaluation of the specificity and predictive values of the following codes and code combinations in each study locality is advised if possible, with appropriate modifications of the following scheme as needed.

(2) In most localities, adequate sensitivity may be expected only when complete data can be linked for both inpatient and outpatient medical encounters in order to rule out acute symptomatic seizures.

Probable

• a single medical encounter assigned an ICD-9-CM diagnostic code 345.xx or ICD-10 code G40.x, OR

• two or more medical encounters on separate days each assigned ICD-9-CM diagnostic codes 780.39 or ICD-10 codes G41.x or R56.8, OR

• a single medical encounter assigned ICD-9-CM diagnostic codes 780.39 or ICD-10 code R56.8 AND an AED is prescribed for outpatient use for 3 or more months.

Suspect

• a single medical encounter is assigned ICD-9-CM code 780.39, or ICD-10 codes R56.8 or G41.x

3.2 Epilepsy ontology

The epilepsy ontology capturing the concepts from the epilepsy definitions can be found at the following web link:

Epilepsy ontology: http://rebrand.ly/mocha-epilepsy

The upper level concept organisation in the ontology is given below.

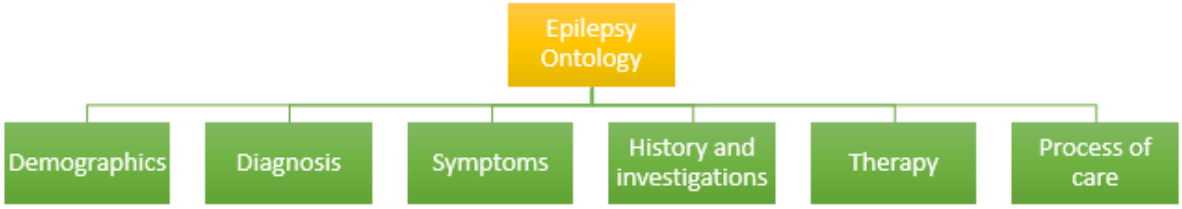

Figure 5 – Upper level concepts of the epilepsy ontology

3.3 Epilepsy outcomes ontology

The epilepsy outcome measures have been organised according to the six quality domains suggested by the Institute of Medicine (given in detail in Chapter 1).

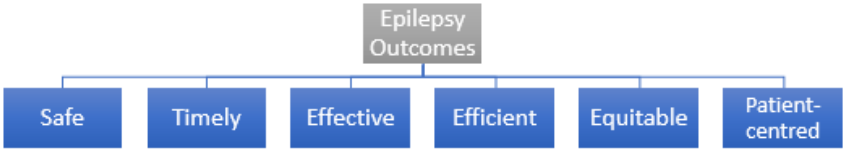

Figure 6 – Upper level concepts of the epilepsy outcomes ontology

3.4 Assessment instrument for understanding measures of health care quality for epilepsy

A selected subset of the epilepsy outcome measures have been used to develop a part of the study specific feasibility assessment instrument which can be used to assess the ability of a chosen database to provide the corresponding outcome. For example, the following grid will be completed by the database indicating if each outcome measure is "Not recorded", "Partially recorded" or "Completely recorded".

Outcome measures for epilepsy	Not recorded	Partial recording	Complete recording
Safety			
1. Monitoring blood levels	☐	☐	☐
2. Seizure related accidents	☐	☐	☐
Intervention efficacy			
1. Days seizure free	☐	☐	☐
2. Able to participate in full range of school activities	☐	☐	☐
Patient centred approach			
1. Individual management plan	☐	☐	☐
2. Ready access to epilepsy nurse	☐	☐	☐
Timeliness of care			
1. Patient waiting times for specialist review	☐	☐	☐
Efficiency			
1. Medication wasted (not taken, or not dispensed)	☐	☐	☐
Equitability			
1. Ethnicity	☐	☐	☐
2. Socioeconomic status	☐	☐	☐

Chapter 4 – Semantic model for attention-deficit hyperactivity disorder (ADHD)

4.1 ADHD case definitions

a) MOCHA ADHD case definition

Reference:

MOCHA - Work Package 1: Identification of models of children's primary care: Systematic Review and Meta-analysis of the Literature – Part 2

ADHD is a psychiatric disorder of the neurodevelopmental type marked by an ongoing pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development (American Psychiatric Association, 2013). The symptoms of inattention and/or hyperactivity-impulsivity must be chronic or long-lasting, impair the person's functioning, and cause the person to fall behind normal development for his or her age.

Three types of ADHD:

1. **Combined** - All three core features are present and ADHD is diagnosed when ≥6 symptoms of hyperactivity/impulsivity and ≥6 symptoms of inattention have been observed for ≥6 months
2. **Inattentive** - Diagnosed if ≥6 symptoms of inattention (but <6 symptoms of hyperactivity/impulsivity) have persisted for ≥6 months
3. **Hyperactive/impulsive** - Diagnosed if ≥6 symptoms of hyperactivity/impulsivity (but <6 symptoms of inattention) have been present for ≥6 months

b) International Classification of Diseases ADHD case definition

Reference:

http://www.icd10data.com/ICD10CM/Codes/F01-F99/F90-F98/F90-

A behavior disorder in which the essential features are signs of developmentally-inappropriate inattention, impulsivity, and hyperactivity, originating in childhood. (At least some of the symptoms must be present before the age of 7 years.) Symptoms last more than 6 months and cause problems in school, at home and in social situations. Adhd is more common in boys than girls. The disorder may be caused by genetics and/or environmental factors.

F90-Hyperkinetic disorders

A group of disorders characterized by an early onset (usually in the first five years of life), lack of persistence in activities that require cognitive involvement, and a tendency to move from one activity to another without completing any one, together with disorganized, ill-regulated, and excessive activity. Several other abnormalities may be associated. Hyperkinetic children are often reckless and impulsive, prone to accidents, and find themselves in disciplinary trouble because of unthinking breaches of rules rather than

deliberate defiance. Their relationships with adults are often socially disinhibited, with a lack of normal caution and reserve. They are unpopular with other children and may become isolated. Impairment of cognitive functions is common, and specific delays in motor and language development are disproportionately frequent. Secondary complications include dissocial behaviour and low self-esteem.

F90.0-Disturbance of activity and attention

Attention deficit:

disorder with hyperactivity

hyperactivity disorder

syndrome with hyperactivity

F90.1-Hyperkinetic conduct disorder

Hyperkinetic disorder associated with conduct disorder

F90.2 - ADHD, combined type

Symptoms of both types are present, but neither is predominant. Most diagnoses of ADHD are this type.

F90.8-Other hyperkinetic disorders

F90.9-Hyperkinetic disorder, unspecified

Hyperkinetic reaction of childhood or adolescence NOS

Hyperkinetic syndrome NOSF84-Pervasive developmental disorders

c) **American Psychiatric Association - Diagnostic and Statistical Manual of Mental Disorders DSM-5 ADHD case definition**

Reference:

http://www.dsm5.org/Pages/Default.aspx

http://www.dsm5.org/Documents/ADHD%20Fact%20Sheet.pdf

The definition of attention-deficit/hyperactivity disorder (ADHD) has been updated in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). ADHD is characterized by a pattern of behavior, present in multiple settings (e.g., school and home), that can result in performance issues in social, educational, or work settings. As in DSM-IV, symptoms will be divided into two categories of inattention and hyperactivity and impulsivity that include behaviors like failure to pay close attention to details, difficulty organizing tasks and activities, excessive talking, fidgeting, or an inability to remain seated in appropriate situations. Children must have at least six symptoms from either (or both) the inattention group of criteria and the hyperactivity and impulsivity criteria, while older adolescents and adults (over age 17 years) must present with five. Descriptions will help

clinicians better identify typical ADHD symptoms at each stage of patients' lives. Using DSM-5, several of the individual's ADHD symptoms must be present prior to age 12 years, compared to 7 years as the age of onset in DSM-IV.

d) **US National Institute of Mental Health (NIMH - NIH) ADHD case definition**

Reference:

https://www.nimh.nih.gov/health/topics/attention-deficit-hyperactivity-disorder-adhd/index.shtml

ADHD is defined as a persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development. ADHD begins in childhood and is considered a developmental disorder. Attention-deficit/hyperactivity disorder (ADHD) is a brain disorder marked by an ongoing pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development.

• Inattention means a person wanders off task, lacks persistence, has difficulty sustaining focus, and is disorganized; and these problems are not due to defiance or lack of comprehension.

• Hyperactivity means a person seems to move about constantly, including in situations in which it is not appropriate; or excessively fidgets, taps, or talks. In adults, it may be extreme restlessness or wearing others out with constant activity.

• Impulsivity means a person makes hasty actions that occur in the moment without first thinking about them and that may have high potential for harm; or a desire for immediate rewards or inability to delay gratification. An impulsive person may be socially intrusive and excessively interrupt others or make important decisions without considering the long-term consequences.

For people with ADHD, these behaviours: are more severe, occur more often, interfere with or reduce the quality of how they functions socially, at school, or in a job.

e) Clinical ADHD case definition

Reference:

Huss M, Holling H, Kurth BM, Schlack R. How often are German children and adolescents diagnosed with ADHD? Prevalence based on the judgment of health care professionals: results of the German health and examination survey (KiGGS). European child & adolescent psychiatry. 2008;17 Suppl 1:52-8.

We define individuals to be affected with ADHD if the diagnosis was provided by a medical doctor or a psychologist. Potential ADHD is evident if individuals reach a clinically significant score of ‡7 on the hyperactivity- inattention subscale of the SDQ and have not yet been given a diagnosis by a medical doctor or psychologist. Additionally, those 3- to 11-year-olds who reach an overall symptom score ‡6 in the behavioural observation but have either not yet been diagnosed or not reached a clinically significant score on the hyperactivity-inattention subscale of the SDQ are considered abnormal with respect to the cardinal symptoms of ADHD.

4.2 ADHD ontology

The ADHD ontology capturing the concepts from the ADHD definitions can be found at the following web link:

Epilepsy ontology: http://rebrand.ly/mocha-adhd

The upper level concept organisation in the ontology is given below.

Figure 7 – Upper level concepts of the ADHD ontology

4.3 ADHD outcomes ontology

The ADHD outcome measures have been organised according to the six quality domains suggested by the Institute of Medicine (given in detail in Chapter 1).

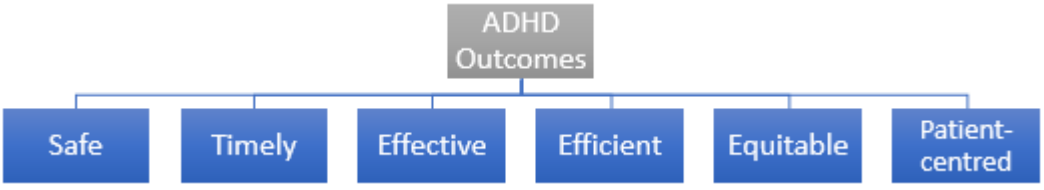

Figure 8 – Upper level concepts of the ADHD outcomes ontology

4.4 Assessment instrument for understanding measures of health care quality for ADHD

A selected subset of the ADHD outcome measures have been used to develop a part of the study specific feasibility assessment instrument which can be used to assess the ability of a chosen database to provide the corresponding outcome. For example, the following grid will be completed by the database indicating if each outcome measure is “Not recorded”, “Partially recorded” or “Completely recorded”.

Outcome measures for ADHD	Not recorded	Partial recording	Complete recording
Safety			
1. Symptom reduction	☐	☐	☐
Intervention efficacy			
1. Educational progression	☐	☐	☐
2. Able to participate in full range of school activities	☐	☐	☐
Patient centred approach			
1. Use of mental health services	☐	☐	☐
2. Ready access to mental health nurse	☐	☐	☐
Timeliness of care			
1. Patient waiting times for specialist review	☐	☐	☐
Efficiency			
1. Medication wasted (not taken, or not dispensed)	☐	☐	☐
Equitability			
1. Ethnicity	☐	☐	☐
2. Socioeconomic status	☐	☐	☐

Chapter 5 – Semantic model for immunisation

We have followed a different approach for this tracer as immunisation is a generic study area and not a specific condition for which case definitions can be defined. Additionally, we have attempted to reuse the extensive work carried out in the biomedical research field to develop immunisation related ontologies.

We have created an immunisation ontology by merging several existing published ontologies which are described below. Merging these ontologies was possible since they were derived from the standard upper level ontology named as Basic Formal Ontology (BFO). Merging of ontologies created an extensive collection of immunisation related concepts. For this reason, we will consider only a relevant subset of concepts within the scope of MOCHA immunisation studies during the feasibility assessment stage and the creation of study specific assessment instruments.

Basic Formal Ontology

BFO is a highest-common-denominator upper ontology that is designed to support interoperability between domain ontologies for shared use of scientific research data across disciplinary boundaries.[9] BFO is used extensively in biomedical research and does not contain physical, chemical, biological or other terms which would properly fall within a specific research area.

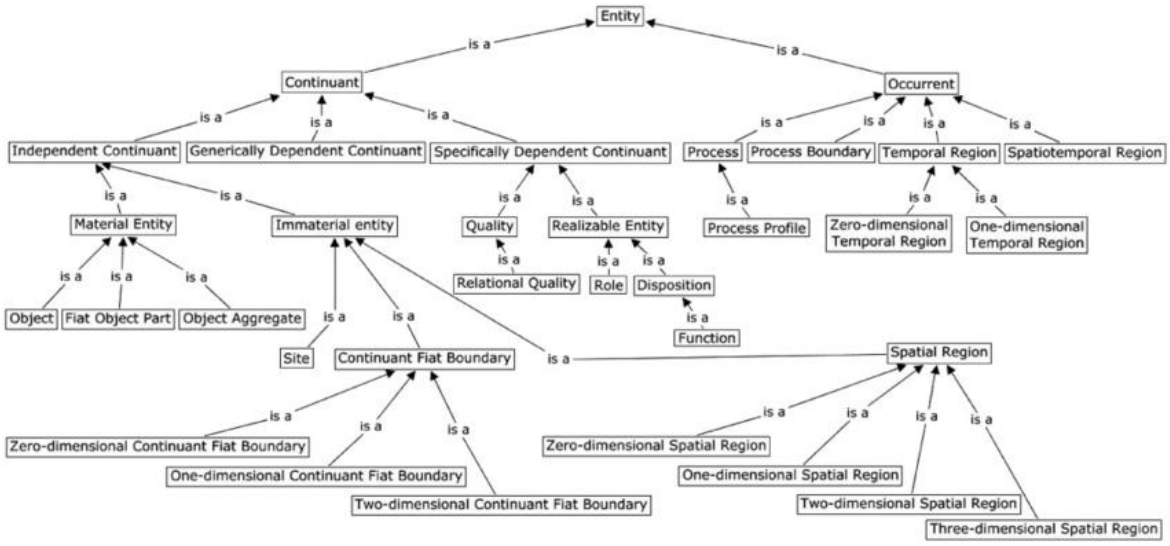

Figure 9 – BFO 2.0 structure
(Source: https://www.researchgate.net/figure/252325553_fig4_Figure-2-Draft-BFO-20-is_a-Hierarchy-21)

5.1 Vaccine ontology (VO)

The Vaccine Ontology was developed to overcome issues of not having a standard vocabulary in the vaccine research field for data integration and analysis. [10] The Vaccine Ontology uses the

Basic Formal Ontology (BFO) to align with other ontologies in biomedical research. The ontology focuses on concepts corresponding to vaccine categorization, vaccine components, vaccine quality, and vaccine-induced host responses. For the immunisation ontology we will benefit from the recording of an extensive list of vaccines and vaccines components.

5.2 Ontology of Adverse Events (OAE)

The Ontology of Adverse Events is a standardised ontology created to semantically integrate data on biomedical adverse events. [11] The adverse events concerned include vaccine and drug adverse events. Similar to the VO described above, the OAE is also based on the BFO standard. Additionally, the OAE described concept types such as adverse events (for vaccines, drugs, medical devices and nutritional products), drug administration, medical device usage and nutritional product usage. The immunisation ontology will mainly utilise the vaccine adverse events given in this ontology.

5.3 Vaccine data source ontology

The vaccine data source ontology is an application ontology developed to describe the semantic concepts of the AIRR (ADVANCE international research readiness) instrument. This instrument was used to profile vaccine data sources in the IMI ADVANCE project.[2] This ontology specifically describes the characteristics of a database including population, coverage, linkage and data access. This ontology is also aligned with the BFO standard.

5.4 Immunisation ontology

The Immunisation ontology developed by merging the above mentioned ontologies can be found at the following web link:

Immunisation ontology: http://rebrand.ly/mocha-immunisation

5.5 Immunisation outcomes ontology

The Immunisation outcome measures have been organised according to the six quality domains suggested by the Institute of Medicine (given in detail in Chapter 1).

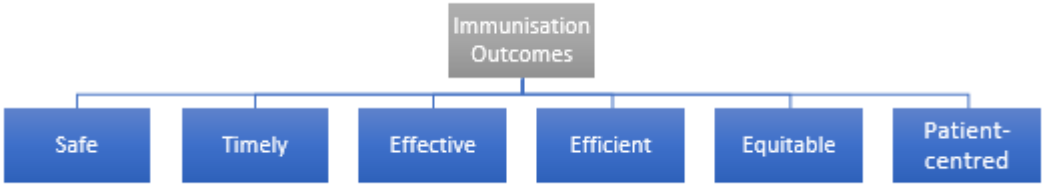

Figure 10 – Upper level concepts of the immunisation outcomes ontology

[2] ADVANCE – Accelerated development of vaccine benefit-risk collaboration in Europe www.advance-vaccines.eu

5.6 Assessment instrument for understanding measures of health care quality for immunisation

A selected subset of the immunisation outcome measures have been used to develop a part of the study specific feasibility assessment instrument which can be used to assess the ability of a chosen database to provide the corresponding outcome. For example, the following grid will be completed by the database indicating if each outcome measure is "Not recorded", "Partially recorded" or "Completely recorded".

Outcome measures for immunisation	Not recorded	Partial recording	Complete recording
Safety			
1. Adverse events following immunisation	☐	☐	☐
2. Precise recording of batch number	☐	☐	☐
Intervention efficacy			
1. Incidence of vaccine preventable disease	☐	☐	☐
2. National/international studies to explore impact, especially sub-groups	☐	☐	☐
Patient centred approach			
1. Adherence to immunisation schedule	☐	☐	☐
2. Personalisation of regimes based on history of allergy and health	☐	☐	☐
Timeliness of care			
1. Timely administration of vaccines	☐	☐	☐
Efficiency			
1. Medication wasted (not taken, or not dispensed)	☐	☐	☐
Equitability			
1. Herd immunity	☐	☐	☐
2. Appropriate level of staff involved in immunisation	☐	☐	☐

Chapter 6 - Conclusion

This deliverable presented a process for systematically developing semantic models for clinical conditions and related outcome measures. The methodology initially examined the evidence-base to identify concepts used to describe features and treatment of clinical conditions from published case definitions. The identified concepts were then used to build semantic models (i.e. ontologies) that captured the semantics of the concepts of the clinical conditions. These ontologies were developed according to the Ontology Web Language standard in order to enable interoperability with other ontologies and developed in the biomedical research community.

The ontologies can be used to annotate data stored in health databases without having any dependency to the coding system used for recording the clinical data. These annotated databases can subsequently be assessed for their feasibility for contributing to MOCHA studies based on chosen clinical conditions. The feasibility assessment process will provide answers to data quality inquiries such as “What clinical concepts are recorded in a database?”, “What outcome measures are recorded?”, “What is the completeness level of data recorded for the clinical concepts?” etc.

In the next phase of MOCHA WP5, we intend to use the semantic models developed to conduct feasibility assessment of databases for the chosen clinical conditions. The process will inform study teams as to which databases are most suitable to contribute data to their studies that compare health systems models in various countries.

Acknowledgements

The MOCHA WP5 team would like to extend their gratitude to Prof. Richard Parish, Prof. Michael Rigby, Prof. Mitch Blair and Dr. Denise Alexander for their valuable comments and suggestions to improve the deliverable.

Appendix A: Using Protégé (Desktop) to view ontology files

Protégé is open-source software used to create, edit and view ontologies developed in OWL (ontology web language). The software can be freely downloaded from the website: http://protege.stanford.edu/.

Ontologies and outcome measures for each of the four conditions (ADHD, asthma, epilepsy and immunisation) have been collected into .owl files provided as an attachment to this report. Users can open and view .owl files in the Protégé software by downloading and opening the software and opening the file. Clicking on the Entities tab will display the ontology classes; clicking on a class will expand the class and display sub-class (i.e. child concepts) if available. When a class is selected, the information related to the class is displayed in Class Annotations on the right of the class panel.

Detailed user guidance is available at:
http://protegewiki.stanford.edu/wiki/Protege4UserDocs

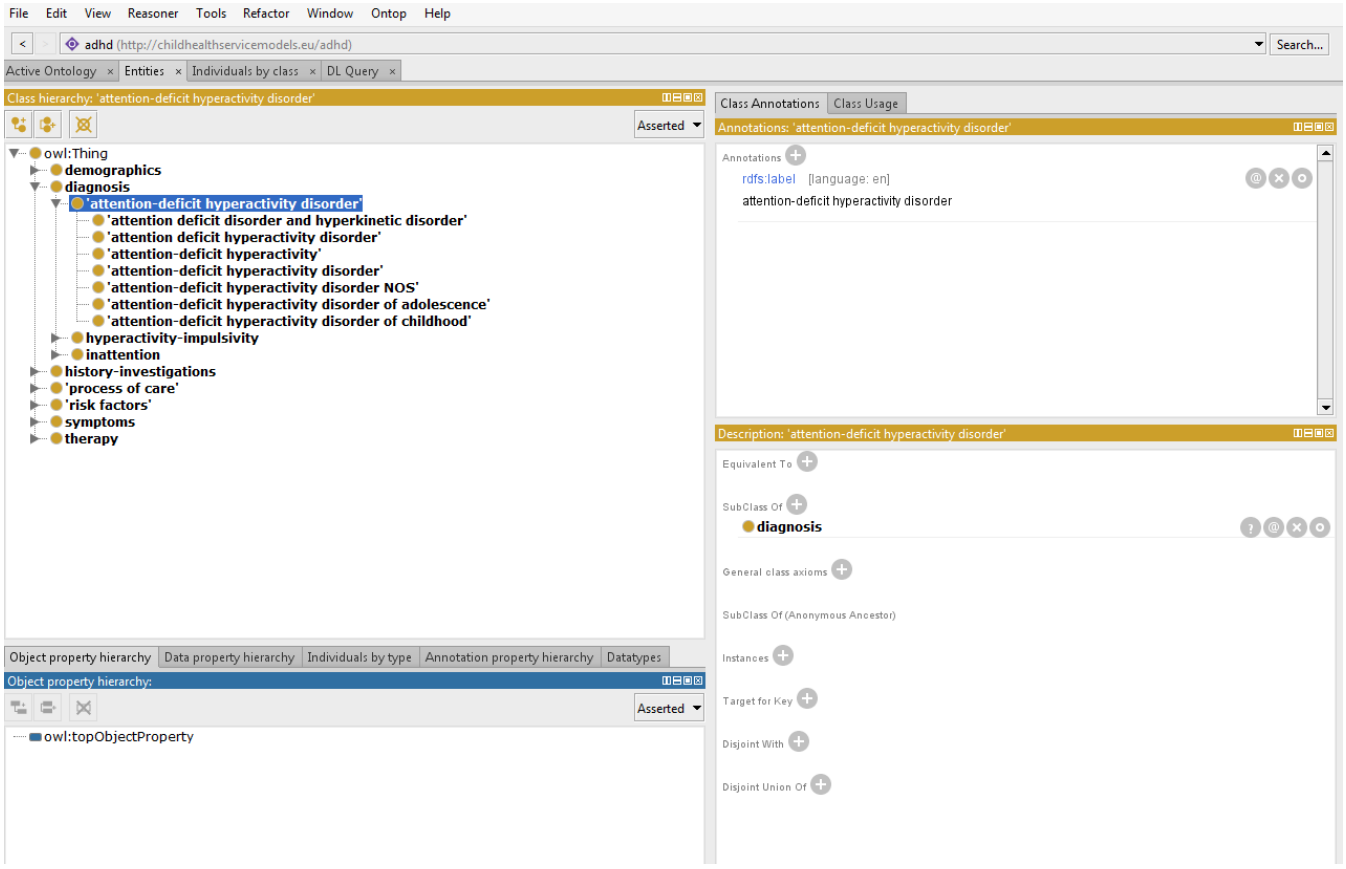

Figure 11 – Screenshot of an .owl file viewed in Protégé software

Appendix B: Using Web Protégé to browse ontologies

Web Protégé is a web-based collaborative ontology development environment. It allows creation and sharing of ontologies developed according to the OWL (ontology web language) standard.

Clicking on the ontology links provided in this document will directly open the corresponding ontology in Web Protégé. The application allows the user to browse the hierarchy of concepts in the "Classes" panel. A concept in an OWL ontology is represented as a class. Clicking on a class will expand the class and display sub-class (i.e. child concepts) if available.

When a class is selected the information related to the class is displayed in Class description panel on the right of the class panel.

A detailed user guide is available at:
http://protegewiki.stanford.edu/wiki/WebProtegeUsersGuide.

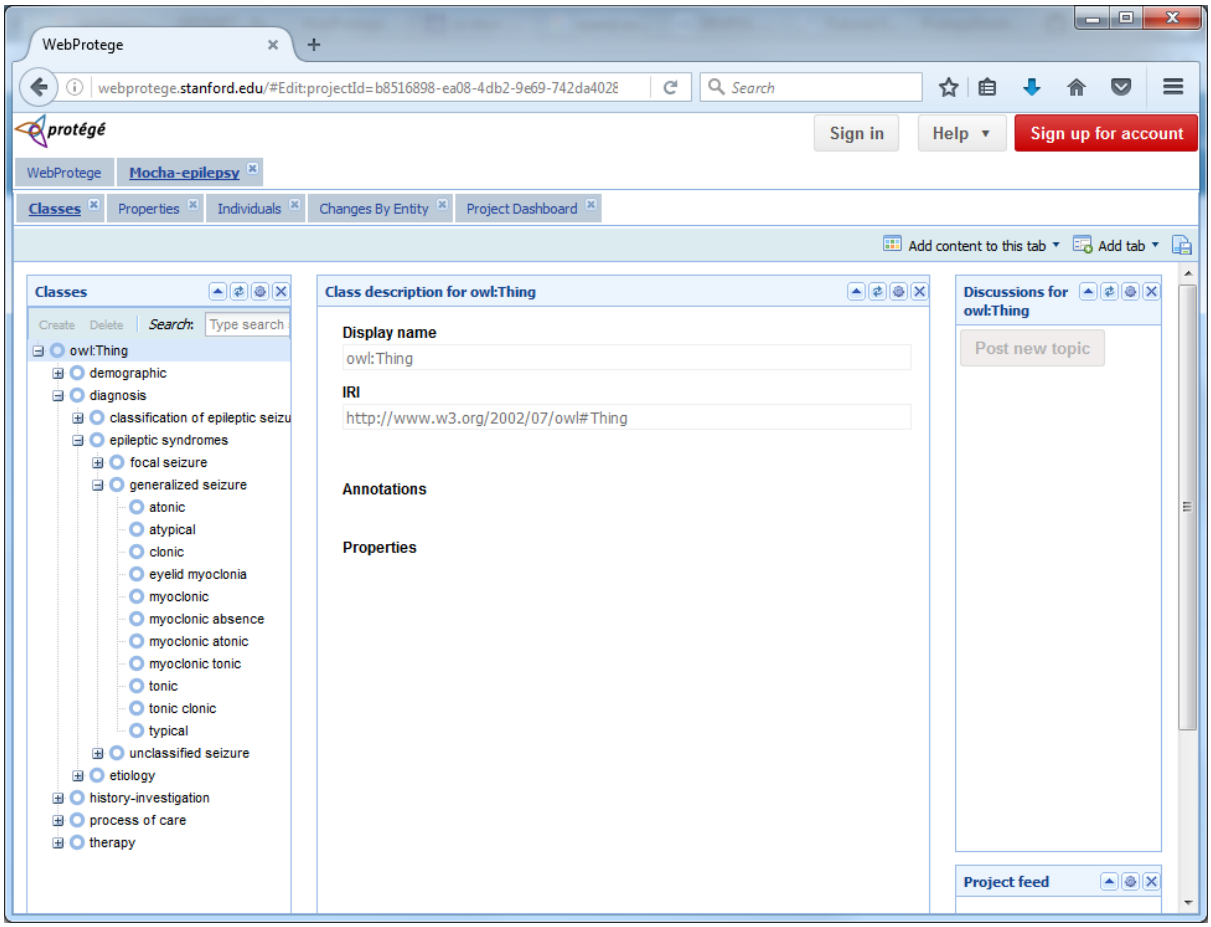

Figure 12 – Screenshot of the Web Protégé application

Appendix C: MOCHA International Research Opportunity Instrument

[Excerpt from D5.1 –Technical Requirements Analysis – Chapter 3]

The MOCHA International Research Opportunity Instrument (MIROI) Instrument was developed with the aim of identify candidate data sources in each participating MOCHA country, so that WP5 can endeavour to obtain comparable information from analysis of electronic health data sources that would indicate the effects and outcomes of the various different child health care models in Europe. The reported through this survey would enable researchers in the MOCHA project to access data sources that could provide data for secondary research purposes and participate in the model appraisal activities.

Survey questions relating to the readiness of data sources for conducting various research projects in generic health data were developed using experiences from previous / ongoing projects (e.g. TRANSFoRm, ADVANCE) exploring this area. The MIROI survey instrument consisted of 23 questions collecting basic information such as database contact details, population, data quality and governance. The questions featured in the questionnaire are given below.

Section A – Database Description

1. Country: ……..

2. Name of the database/register: ……

3. Database/register website (URL): ……..

4. Type of Database (by source of data upload): downloaded anonymised primary care dataset / hospital discharge data / insurance claim database / disease registry (specify) …….. / regular survey (specify) …….. / Census (specify)…….. / other (specify): ……

5. Please indicate how frequently the database is updated
   - ☐ Daily (ongoing data entry)
   - ☐ Weekly
   - ☐ Monthly
   - ☐ Three monthly
   - ☐ Six monthly
   - ☐ Annually
   - ☐ Not updated

6. Brief description about data custodian**:

** The data custodian is the entity managing the database.

……

7. Population covered: e.g. whole country / defined locality (specify) …… /

other (specify) ……

8. Age range covered: ……..

9. How up-to-date is the database: e.g. last update and period covered:

……

10. Time lag before data are released for analysis (e.g. do local preset analyses have to be published before the data are released for other analysis?):
……..

11. Appraisal of population representation, data accuracy or latent bias (e.g. are there any known exclusions/low coverage such as private patients, immigrant health services):
……..

12. Which of the terms given below could be used to classify the data contents of the database?(select all that apply)

- ☐ Primary health care
- ☐ Outpatient electronic medical records
- ☐ Community /ambulatory care records
- ☐ Inpatient electronic medical records / hospital
- ☐ Health care reimbursement claims, including date and place of service, patient, diagnoses, treatment.
- ☐ Communicable / infectious disease surveillance
- ☐ Vaccination / immunisation registry or coverage data
- ☐ Population data (census and demographic)
- ☐ Vital records (birth and death registries)
- ☐ Pharmacy dispensing records
- ☐ Specialized care consultations
- ☐ Drug / vaccine adverse event reporting systems
- ☐ Specific registry (inc. chronic or rare disease, cancer registries)

- ☐ Population health surveys
- ☐ National health surveys
- ☐ Health care costs
- ☐ Biobank (e.g. genetics data)
- ☐ Pharmacovigilance systems
- ☐ Novel data sources(e.g. wearable devices, mobile applications)
- ☐ Other(Please specify:
)

13. Does the database contain data items that could assist in determining Equity of Access issues:

    Sex: Yes / No

    Ethnicity: Yes / No How assessed and recorded: ……..

    Socio-economic Group: Yes / No How assessed and recorded: ……

    Small area locator: Yes / No How assessed and recorded: ……..

14. Date range for which complete quality data is available

    From ___/__ /_____ (DD/MM/YYYY) to ___/__ /_____ (DD/MM/YYYY)

15. Total number of registered subjects, including adults (please provide population

    denominator data as well): ……

16. Total number of registered children (0-18 years of age) available for studies:

    …… (indicate if only partial child age range): ……

Section B – Database Access

17. Is there a written policy governing data access?

- ☐ Yes
- ☐ No

Comments:

18. Who can authorise access to the database? **

Title:

First name:

Last Name:

Organisation:

Job Title:

Affiliation:

Address:

City:

Postcode:

Country:

Phone number (inc. international code):

Alternative phone no:

Email address:

**If the authorisation is given by a committee, please provide details of a representative.

19. Scientific contact person details

Title:

First name:

Last Name:

Organisation:

Job Title:

Affiliation:

Address:

City:

Postcode:

Country:

Phone number (inc. international code):

Alternative phone no

Email address:

20. Before granting access to data, who needs to evaluate requests for data access?

……

21. Is a charge made for data access?

- ☐ Yes
- ☐ No

Comments on charge basis:

22. If enquiries and analyses of the database can only be undertaken by named scientists authorised as trusted, please give details of an appropriate contact:

Title:

First name:

Last Name:

Organisation:

Job Title:

Affiliation:

Address:

City:

Postcode:

Country:

Phone number (inc. international code):

Alternative phone no

Email address:

23. Would this agent make a charge made for processing analyses?

- ☐ Yes
- ☐ No

Comments on charge basis:

Please provide your details (respondent)

Respondent name:

Organisation:

Email address:

Appendix D: OWL (Ontology Web Language)

OWL is a language for expressing ontologies. The term ontology in informatics usually corresponds a certain kind of computational artefact.

An ontology is a set of precise descriptive statements about some part of the world (usually referred to as the domain of interest or the subject matter of the ontology). Precise descriptions satisfy several purposes: most notably, they prevent misunderstandings in human communication and they ensure that software behaves in a uniform, predictable way and works well with other software.

In order to precisely describe a domain of interest, it is helpful to come up with a set of central terms – often called vocabulary – and fix their meaning. Besides a concise natural language definition, the meaning of a term can be characterized by stating how this term is interrelated to the other terms. A terminology, providing a vocabulary together with such interrelation information constitutes an essential part of a typical OWL document. Besides this terminological knowledge, an ontology might also contain so called assertional knowledge that deals with concrete objects of the considered domain rather than general notions.

OWL is not a schema language for syntax conformance. OWL does not provide elaborate means to prescribe how a document should be structured syntactically. In particular, there is no way to enforce that a certain piece of information (like the social security number of a person) has to be syntactically present.

OWL is not a database framework although OWL documents store information.

- Based on OWL 2 Web Ontology Language Primer (Second Edition)

http://www.w3.org/TR/2012/REC-owl2-primer-20121211/

References

1 Kringos D, Boerma W, Bourgueil Y, Cartier T, Dedeu T, Hasvold T, Hutchinson A, Lember M, Oleszczyk M, Rotar Pavlic D, Svab I, Tedeschi P, Wilm S, Wilson A, Windak A, Van der Zee J, Groenewegen P. The strength of primary care in Europe: an international comparative study. Br J Gen Pract. 2013 Nov;63(616):e742-50.doi: 10.3399/bjgp13X674422.
2 de Lusignan S. In this issue: Ontologies a key concept in informatics and key for open definitions of cases, exposures, and outcome measures. J Innov Health Inform. 2015 Jul 10;22(2):170. doi: 10.14236/jhi.v22i2.170.
3 Liyanage H, Krause P, de Lusignan S. Using ontologies to improve semantic interoperability in health data. J Innov Health Inform. 2015;22(2):309-15. doi: 10.14236/jhi.v22i2.159.
4 Institute of Medicine . Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.
5 Martínez-Costa C, Cornet R, Karlsson D, Schulz S, Kalra D. Semantic enrichment of clinical models towards semantic interoperability. The heart failure summary use case. J Am Med Inform Assoc. 2015 May;22(3):565-76. doi:10.1093/jamia/ocu013.
6 de Lusignan S. Codes, classifications, terminologies and nomenclatures: definition, development and application in practice. Inform Prim Care. 2005;13(1):65-70.
7 Dentler K, Numans ME, ten Teije A, Cornet R, de Keizer NF. Formalization and computation of quality measures based on electronic medical records. J Am Med Inform Assoc. 2014 Mar-Apr;21(2):285-91. doi: 10.1136/amiajnl-2013-001921.
8 Centers for Disease Control and Prevention (2006, 2012). Principles of Epidemiology in Public Health Practice: An Introduction to Applied Epidemiology and Biostatistics.
9 Simon J, Dos Santos M, Fielding J, Smith B. Formal ontology for natural language processing and the integration of biomedical databases. Int J Med Inform. 2006 Mar-Apr;75(3-4):224-31.
10 He Y, Cowell L, Diehl AD, Mobley HL, Peters B, Ruttenberg A, Scheuermann RH, Brinkman RR, Courtot M, Mungall C, Xiang Z, Chen F, Todd T, Colby LA, Rush H, Whetzel T, Musen MA, Athey BD, Omenn GS, Smith B. Proceedings of the 1st International Conference on Biomedical Ontology (ICBO 2009), Buffalo, NY, USA. VO: Vaccine Ontology. The 1st International Conference on Biomedical Ontology (ICBO 2009), Buffalo, NY, USA. Nature Precedings. 2009. http://precedings.nature.com/documents/3553/version/1.
11 Yongqun He Y, Sirarat Sarntivijai, Yu Lin, Zuoshuang Xiang, Abra Guo, Shelley Zhang, Desikan Jagannathan, Luca Toldo, Cui Tao and Barry Smith. OAE: The Ontology of Adverse Events. Journal of Biomedical Semantics. 2014, 5:29 doi:10.1186/2041-1480-5-29.