OCRe: An Ontology of Clinical Research

Samson W. Tu, MS,[1] Simona Carini, MA,[2] Alan Rector, MD PhD,[3] Peter Maccallum, PhD,[4] Igor Toujilov, PhD,[5] Steve Harris,[6] PhD, Ida Sim, MD PhD[2]

[1]Center for Biomedical Informatics Research, Stanford University, Stanford, CA, USA
[2]Division of General Internal Medicine, UCSF, San Francisco, CA
[3]Department of Computer Science, University of Manchester, Manchester, UK
[4]CRUK Cambridge Research Institute, Cambridge, UK
[5]University College London, London, UK
[6]Oxford University, Oxford, UK

Abstract

Clinical research is generating research data at an increasing pace, yet there is no standard method for investigators to mine the sea of research data to find those that are relevant to their scientific hypothesis. Querying data and meta-data across clinical trials and observational studies is difficult because of the lack of semantic and terminology standards for describing the design and methods of human studies, and because of the variety of clinical terminology standards. The Ontology of Clinical Research (OCRe) is a formal ontology for describing human studies that provides methods for binding to external information standards (e.g. BRIDG) and clinical terminologies (e.g. SNOMED CT). It allows the indexing of research data across multiple study designs, interventions, outcomes, and health conditions. With such an indexing, investigators interested in the evidence pertaining to a particular question (e.g., what is the effect of A on B in people with C) will be able to locate relevant research data more easily.

Keywords: Ontology, Human Study, Clinical Trial

Introduction

Searching for studies that involve particular designs or interventions, for study components (e.g., case report forms), or for data about particular types of patients in particular types of trials are important tasks for clinical and translational research. Such queries are currently difficult to execute because: (1) clinical terms are rarely standardized across studies; (2) studies are run on different clinical trial management systems that often have different information models; and (3) there is no standard terminology or information model for the design and methods of human studies. To address the third difficulty, we developed the Ontology of Clinical Research (OCRe), a formal ontology, represented in OWL 2.0 (Motik, Patel-Schneider et al.), of the entities concerning the design and study of human experiments, and the logical relationships among them.

OCRe is a modular ontology of clinical investigation that includes (1) a representation of the structure of human studies and associated entities such as persons and organizations that play significant roles in studies, (2) informational entities such as study protocols and outcome variables, that are produced in the life cycle of studies, (3) terms for describing study characteristics, and (4) bindings to standard terminologies such as SNOMED CT and NCI Thesaurus. We import terms and relationships from existing models such as BRIDG (Fridsma, Evans et al.) whenever possible, and show that the ontology is consistent with the Basic Formal

Ontology (BFO) (Grenon, Smith et al.). We want to demonstrate that, with such a rich model of investigational studies, we can integrate multiple types of studies by making queries across them that cannot be done otherwise.

The paper is organized as follows: In the Background section, we describe the use cases and prior work that motivate this work; The Model Formulation section describes the strategies we used to scope and structure OCRe, and the Model Description section describes the modules and the entities and relationships in the ontology, and shows how they can be seen as extensions of BFO.

Background

Work on OCRe started with the need of the UK CancerGrid[1] project to establish a consensus regarding vocabularies to describe clinical trials and data elements for the purpose of annotating and integrating data. At the end of the CancerGrid project in 2008, the development of OCRe continues as part of the UCSF Trial Bank project, with collaborators from Stanford University, University of Manchester, and University College London.

Use cases

The scope of OCRe covers all interventional and observational human studies of any design (e.g., randomized, cross-sectional), any intent (e.g., therapeutic, preventive), in any clinical domain, and with any type of data (e.g., quantitative, qualitative, imaging, genomics, proteomics). To focus our work to a manageable scope, and to specify the competency the ontology should support, we defined the tasks and use cases that OCRe may satisfy. Among innumerable possible uses of an ontology of clinical studies, we identified three tasks that can help us to delimit the scope of OCRe and to contrast with existing ontologies.

An ontology of clinical research can provide the structure and terms to annotate and classify studies. The ontology should defines the essential properties and components of studies and, together with domain vocabulary, specify terms that characterize the studies. The annotated studies should have sufficient information to allow investigators to search for studies relevant to a research question. For example, an investigator may query a repository of annotated studies to find *all human studies contributing evidence to the effect of Drug A (Intervention or Exposure) on Mortality (Study Outcome) in patients with Diabetes (Health Condition, Eligibility Criteria)*. The ontology should provide uniform terms and structures that allow the query engine to return different types of studies that satisfy the query criteria. For interventional and experimental studies the query should return studies fulfilling "matches" for PICO (Patient/Eligibility, Intervention, Condition, Outcome), further classified by having control or not, randomized or not, and if has control, what the control intervention is/are. For Observational studies (e.g. case control, cross-sectional studies), the query should match intervention A to Exposure, Mortality to Study Outcome, Diabetes to Health Condition/Eligibility Criteria. If the study is an *analytic* study (i.e. there's a comparison group), the query should control exposure (e.g., no drug A, or something else). For secondary data analysis studies, the query should return the study if statistical hypothesis of the analysis matches our PICO, i.e., if the analysis was in a study subgroup (diabetics only), or outcome was a post-hoc outcome. The query results should be ranked by a hierarchy of

[1] http://www.cancergrid.org/

evidence. For example, double-blinded clinical trials should be ranked ahead case reports, and similarly pooled studies (e.g., meta-analysis) should be ranked ahead of single-source studies.

An ontology of clinical research can be extended to provide the methods for analyzing subject data for carrying out tasks in the life cycle of human studies. At the recruitment phase, for example, if a study's eligibility criteria can be precisely specified, they may be translated into executable queries on subject data to determine whether a subject is eligible for the study. At the analysis phase, if the outcome variables of a study are well-defined in terms of subject data, investigators may be able to compute them for individual subjects or for relevant subject populations.

An ontology of clinical research can be extended to support the management of subjects who are enrolled in the study. In addition to screening for eligible subjects, the management tasks may involve the scheduling of activities (e.g., drawing blood for various tests), computation of appropriate interventions (e.g., reducing the doses of chemotherapy in response to drug toxicity), or managing the workflow samples. The ontology can be the basis for detailed encoding of the activities prescribed by the study and computerized support for carrying out those activities. Alternatively, the ontology may provide the standard data structures that applications use when exchanging data among them.

We envision that OCRe's core niche lies in the first task, with extensions to support the second task. Scientists are most interested in the evidence pertaining to a particular question (what is the effect of A on B in people with C). OCRe should allow us to index into the scientific question studied across a variety of study designs, based on our modeling of study interventions, outcomes, and eligibility criteria. The value is in modeling these study design concepts broadly but rigorously across study designs, rather than modeling them deeply in one design.

Nevertheless, there is great interest in the clinical research community to establish a computer-interpretable language for standardizing eligibility criteria. The Clinical Research Special Interest Group of the American Medical Informatics Association and the Agreement on Standardized Protocol Inclusion Requirements for Eligibility (ASPIRE), a subgroup of the CDISC Protocol Representation Committee, for example, are groups that are attempting to develop standard eligibility criteria representations (Niland). Even though extending the OCRe ontology to support queries and analysis of data is not our primary focus, we want to demonstrate that the OCRe ontology is compatible with such extensions.

We chose to The amount of effort required to annotate a large number of studies is far less than what is required if tasks 2 and 3 are the primary goals. For example, instead of formulating an eligibility criterion *bronchoalveolar carcinoma with lobar or multilobar involvement* in terms of data (which may not be available) that can fully determine whether a patient satisfies the criterion, an annotation on the study may only indicates that *bronchoalveolar carcinoma*, or even just *carcinoma*, is the health conditions that help to define the eligibility criterion. Such coarse-grained annotation increases the false positive rate when an investigator is searching for more specific disease, but manual encoding of the full meaning of eligibility criteria and similar concepts such as outcomes are extremely labor intensive.

Several existing ontologies that focus on clinical trials aim to support task 3, management of subjects during the execution phase of a study. In the next section, we discuss the relationship between OCre and the prior work.

Prior Work

There are several ongoing efforts to model various aspects of human studies, including models specifically designed for clinical trials. How does OCRe compare with them and what are the chances of harmonizing with them?

BRIDG

BRIDG is a UML model of regulated clinical trial research that has gained adoption by HL7, caBIG, and CDISC. The BRIDG model has been developed with the goal of designing a model of the shared semantics of protocol-driven research (and its associated regulatory artifacts). This model is meant to support computable semantic interoperability between applications developed in caBIG and for message specifications developed in CDISC and HL7. As described in the previous section, OCRe borrowed liberally from BRIDG. BRIDG models planned, scheduled, and performed entities. OCRe incorporates most of BRIDG planned study entities. The `Clinical Event` component of OCRe correspond to BRIDG performed activities. The OCRe Observation class is a combination of BRIDG PerfomedObservation and ObservationResult classes. Furthermore, OCRe incorporates and re-interprets most of the BRIDG classes related to Entity, Role, and Participation.

Despite the close relationship between OCRe and BRIDG, there are also significant differences. OCRe's use cases involve querying studies of different types, including observational and retrospective data analysis, and, through the use of ERGO, specifying the eligibility criteria and outcome variables. In contrast, the use cases of BRIDG involve interoperability of applications. The applications, such as Patient Study Calendar developed at Northwestern University, uses BRIDG to define data structures that can be shared with other applications. OCRe, in contrast, will be used primarily for annotation of studies.

There are also differences in conceptualization: BRIDG: `PlannedStudy` is_a `StudyProtocol`, which is_a `Document`. OCRe makes a distinction between Study as an entity that exists in the world and the study protocol which is an informational entity that can be realized as a document. A study has a `StudyProtocol`, which contains study-specific descriptions such as eligibility criteria and outcome variables, and which has a `PlannedStudy`, a business process description of the planned activities in the study. OCRe is not designed for protocol management, hence not interested in scheduling of activities except to the extent needed for recognizing before and after temporal sequencing.

Finally, the differences between OCRe and BRIDG manifest themselves in the modeling languages used. OCRe is formalized in OWL, where the semantics of and interrelationships among terms are explicit in terms of logical axioms that allow automated inferences. BRIDG, in contrast, is an UML model that lends itself to integration with software development process.

Ontology of Clinical Investigation (OCI) and Ontology for Biomedical Investigation (OBI)

The OCI project focuses on the curation and organization of terms used in clinical investigation. Its goal is to create a generic hierarchy of terms independent of particular use cases. To accomplish that goal, it collect terms from multiple sources and organizes them as part of an upper-level ontology. As of 2008, the plan is to incorporate OCI as part of OBI, and, when OBI has an official release, extract an OCI view. In contrast to OCI, the development of OCRe is

very much use-case driven. The use cases inform the conceptualization of what a research study is and the scope of what should be included in the ontology. Furthermore, OCRe is not just a collection of terms, but contains classes that define the structure and properties of the studies.

Epoch

Epoch (Shankar, O'Connor et al.) is a clinical study ontology that, in its current form, focuses on providing support for the management of trial tasks. For example, Epoch allows detailed specification of the timing constraints among trial activities and the workflow involved in managing subjects' clinic visits, specimen collection, and shipping and receiving of bar-coded specimen containers. It has not attempted to model study design and characteristics that would facilitate queries across studies or detailed eligibility criteria.

Model Formulation Process

Given our primary goal of developing a light-weight ontology suitable for annotating human studies, it is imperative that we reuse, as much as possible, ontologies, information models, and terminologies, such as SNOMED CT and BRIDG, that have already covered relevant domains in great details. To develop OCRe, we follow the basic strategy of conceptualize, modularize & bind. First we articulate a conceptualization of human investigations that we want to model and clarify the roles of information model and terminologies in our ontology, then we create a core set of modular, related ontologies about the design and methods of human investigations, provide defined interfaces for binding to external ontologies and vocabularies as needed and finally import subsets of concepts from external models (e.g., BRIDG) when relevant. The development of OCRe follows best practices and “normalization” principles for ontology construction (Rector; Bittner and Smith) and is guided by iterative testing against queries defined by our use cases.

Model Description

An ontology is the formal specification of a conceptualization of domain (Gruber). In the first section, we describe the conceptualization of the entities and relationships in designing, performing, and analyzing human studies. Next we clarify the roles that information models and terminologies play in the ontology, thus allowing the creation of bindings to external terminologies such as SNOMED CT and NCI Cancer Thesaurus and of mappings to information models such as BRIDG. After describing the components of OCRe, we show how OCRe terms relate to terms in BFO, a upper-level ontology increasingly being adopted by the biomedical informatics community.

Conceptualization of Human Study

A study is a real-world entity whose properties and components parts evolve during the life-cycle of the study. At the design stage, a study is little more than a set of documents that spell out the scientific hypothesis being studied, the design of the study, and, if it is a prospective study, planned activities of the study. At the execution phase, participants of a study carry out activities that, from the point of the view of the study, result in a body of collected data. In the analysis phase, investigators transform the data and perform statistical analysis on them, resulting in publications and other artifacts (e.g., submissions to GEO).

OCRe focuses on the design and analysis phase of studies. The objects in the ontology includes `Study` and physical entities such as `Person` and `Material`. Aggregates of objects include `Population` and `Organization`. Objects and aggregates of objects can take up `Roles` such as `Study Subject` and `Healthcare Provider`. A study includes the scientific hypothesis being tested, the study protocol, investigators, subjects, data sets, attributes such as start date and end date, and may include one or more study sites. It can be characterized in terms of qualities, such as study purpose (e.g., prevention, diagnostic, treatment), study type (e.g., interventional study or cohort observational study) or study status (e.g., planning, enrolling, or completed).

A particular study (e.g., a NCI sponsored clinical trial) is an instance of OCRe `Study`.

The design of a study is embodied in a `Study Protocol`, which describes details related to activities planned to achieve the objectives of the study. The study protocol specifies, for example, the subject to be enrolled, activities to be performed, data to be collected, outcomes to be analyzed, and method of analysis. A study protocol is an informational entity that is ontologically distinct from events that occur as part of the study.

Events are occurrences that happen to study subjects (called `Clinical Events` in OCRe) or to studies as a whole (e.g., `Management Events` such as adding sites or stopping recruitment). Most important clinical events are observations made (including assessments, diagnosis, and adverse events) and treatment performed (e.g. procedures and pharmacological treatment) on a subject. Additional research-related events may include enrollment, assignment, and sample collection.

Ideally the study protocol should constrain the events that occur as part of the study. However, the study that is executed may violate the constrains intended by the study protocol. In OCRe, we have not attempted to model the constraints expected to hold between the specification of a study protocol and the events that occur as part of a study.

Ontology, Information Model, and Terminology

We formalize the conceptualization of human studies in an ontology. Following (Smith, Kusnierczyk et al.), we view an ontology as a logical specification of the universals and defined classes in a specific domain and, at minimum, of the is-a relationship among the classes. The universals are types of entities that share some intrinsic characteristics. The domain may include informational entities, such as content of protocol documents and clinical statements. An information model, in this view, is the portion of the ontology that specifies the structure of the informational entities in the domain.

Conventional information models, such the Health Level Seven Reference Information Model (HL7 RIM) (Health Level Seven) and BRIDG, are best seen as data structures that have some correspondence to entities in the ontology. A clinical statement, recorded by some clinician, about a patient having a rash is a data structure that holds some information content: the rash that a patient is experiencing at a certain time. Thus, use of ontologies and conventional information models requires careful mapping (Rector, Qamar et al.). Some classes in an ontology of human studies, such as protocol and trial data, are informational entities that map directly to conventional information model classes in BRIDG or HL7 RIM. Others, such as `Organization` and `Person`, requires a shift from ontological view (e.g., seeing an instance

of `Person` as denoting a real person) to informational view (e.g., seeing the instance as an information record about a person).

A terminology is a collection of terms that refer to entities in a specific domain. The set of names or codes that refer to universals or defined classes in a particular ontology gives rise to a reference terminology. An ontology of diseases, for example, has classes that represent disease types. Instances of a disease are specific manifestations of the disease in different persons. The set of disease names and the is-a relationship of the diseases in the ontology give rise to a disease terminology. A clinical diagnosis, as recorded in an information model, is a statement, made by some clinician, that associates a patient identifier with a reference (a disease name or code) to a disease. The is-a relationships in the ontology correspond to the taxonomy of the disease names or codes.

In general, the relations (object properties and data properties of OWL) in the ontology correspond to structural attributes and relations in conventional information models. The values of structural attributes (e.g., the code attribute in the HL7 Observation class) are constrained to terms that refer to the ranges of the underlying relations in the ontology.

OCRe Components

OCRe is a set of modular components organized by their import relationship (Figure 1). The core modules are *clinical* (containing shared upper-level entities) and *research* (containing terms and properties for characterizing a study). The *criterion* and *study_protocol* modules are extensions, based on external models (ERGO (Tu, Peleg et al. 2008) and BRIDG respectively), that allow us to encode complex computer-interpretable expressions and criteria (e.g., eligibility criteria and outcome variables) and to specify the temporal aggregates (e.g., epochs and arms in clinical trials) and sequencing relationships among activities. The *research*, *study_protocol*, *criterion*, and *clinical* modules are applicable to any clinical domain. They are designed to be used in conjunction with domain specific models of health conditions, interventions, and measurements which conform to the shared ontology of clinical concepts to provide specific models of studies. For the purpose of annotating the set of sample studies (the *test-trials* module) we use SNOMED CT terms that are spelled out in the *snomed_interface* module. Finally, the *bfo-mapping* module contains the connecting relationships that map OCRe entities to BFO entities.

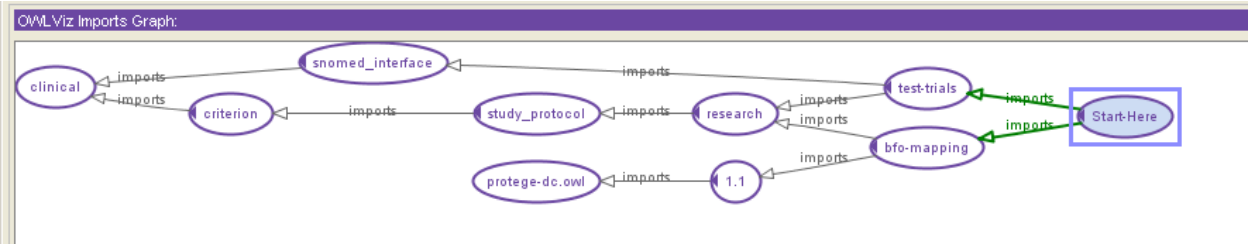


Figure 1 Modules of OCRe Ontology. Clinical and Research modules contain the core OCRe concepts. Criterion and Study Protocol modules contain extensions to allow modeling of decision criteria and scheduling of study activities. The bfo-mapping module maps OCRe to BFO so that OCRe becomes an extension of BFO.

Terminology interface

In OCRe, we try to distinguish concepts and relationships that are part of the OCRe ontology and those that should be imported from the external ontologies and terminologies such as NCI Thesaurus and SNOMED CT. We assume that these external ontologies and terminologies define concepts that represent a typology of entities in the world and terminology codes that denote the concepts.

In OCRe's interface to external terminologies, the world of concepts and the world of codes are related through the `has_code` annotation property on concepts and `is_code_of` object property of codes. Thus, for the `acute_myocardial_infarction` concept in SNOMED CT disease hierarchy, we have an `acute_myocardial_infarction_code` that `is_code_for` `acute_myocardial_infarction` (Figure 3). To allow subsumption reasoning within the Protégé editor, we imported a small fragment of the SNOMED CT hierarchy. Terms in the hierarchy are annotated with their corresponding codes (Figure 3). If the ontology is used together with a terminology service, the subsumption relationship will be provided by the terminology service and it won't be necessary to construct such terminology hierarchies in Protégé.

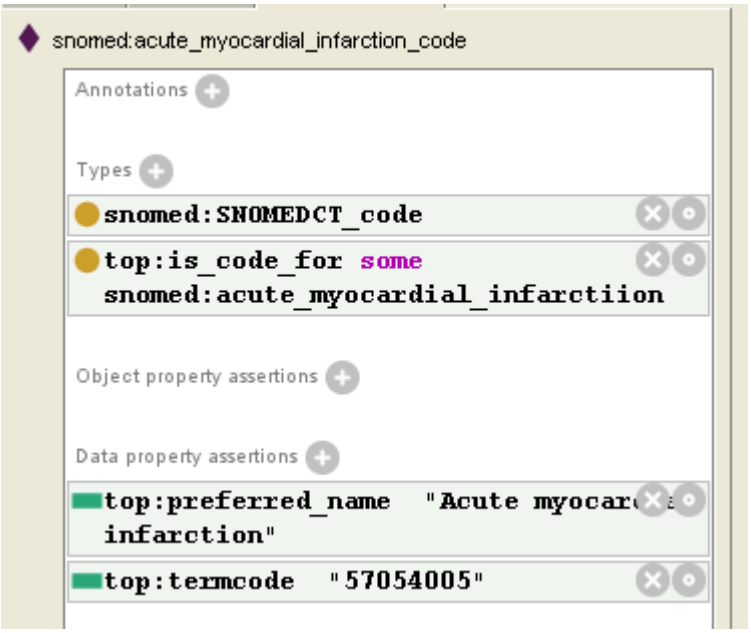


Figure 2 OCRe's interface to SNOMED CT. The terminology code snomed:acute_myocardial_infarction_code has the SNOMED CT code and preferred name, and an optional reference to the class for which it is a code.

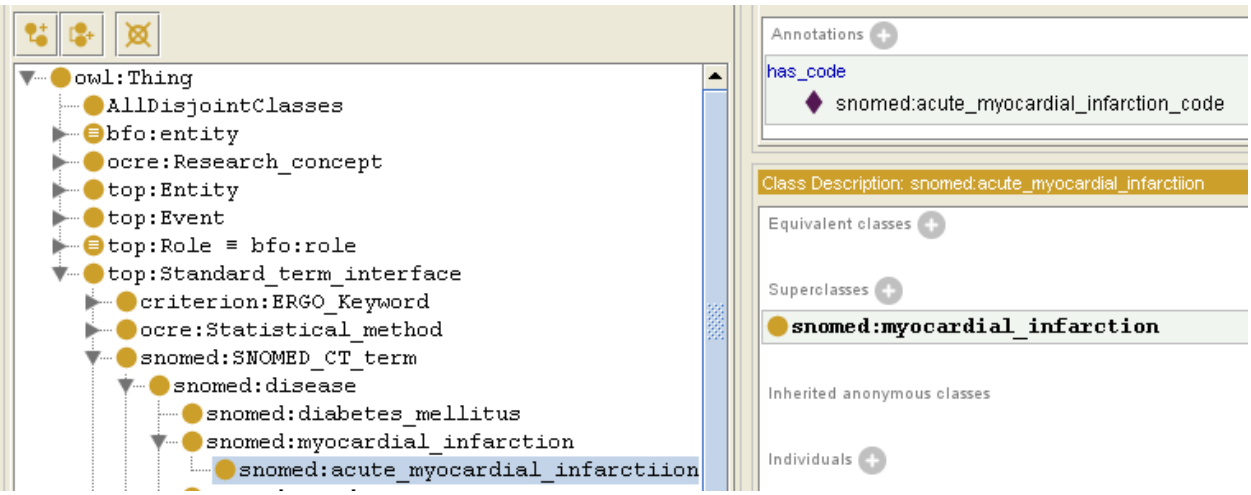


Figure 3 The term snomed:acute_myocardial infarction is annotated with a snomed:acute_myocardial_infarction_code.

To encode a study, we create an individual of the `Study` class and make property assertion `has_health_condition_studied acute_myocardial_infarction_code`. There is no need to assert the restriction `"has_health_condition_studied some acute_myocardial_infarction"` because with the `is_code_of` property, we can always navigate from the code to the concept, and from there, check subsumption relationships with other concepts. Furthermore, when we want to write class expressions to query for the health condition studied in a study, we need to write the query expression in terms of the terminology code denoting the health conditions, not the class `acute_myocardial_infarction`. We will never get the concept class as a subclass of the query expression, for the simple reason that there may be instances of the class that are not studied in the study.

One way to think about assertions like `thrombolytic_therapy_vs_PCI_protocol has_health_condition_studied acute_myocardial_infarction_code` is that they are assertions in the study_protocol, an information entity, of a study. A study has other things, like study start and end dates, specimens and data collected, that are not part of the study protocol.

Study and its properties

A study is the root individual in the OCRe ontology. It is characterized by a set of study characteristics, has attributes such as the date of first enrollment, and object properties to other entities such as the study protocol, healthcare organizations that function as study sites of the study, and populations of subjects who are screened and enrolled.

We model study characteristics as OCRe-specific hierarchy of qualities (Figure 4). The hierarchy of qualities is being refined continually. It constitutes a small terminology for characterizing studies. As such, just like the terminology interface to SNOMED CT, we associated terminology codes with the concepts. For a particular study (e.g., Thrombolytic_Therapy_versus_Primary_Percutaneous_Intervention), we place restrictions like `has_study_characteristics some Interventional` and property assertion like

`has_study_characteristic Interventional_code` to denote that the `study_type` is `Interventional`.

We have to use both restrictions and property assertion because, just as in the case of acute myocardial infarction described earlier, if we want to query for the study type of a study, an OWL query like

```
Study_type and is_study_characteristic_of some (Study and
(scientific_title value "Thrombolytic Therapy versus Primary
Percutaneous Intervention" ))
```

will not give you `Interventional`. We have to make the query for the code,

```
is_code_for some Study_type and is_study_characteristic_of some
(Study and (scientific value "Thrombolytic Therapy versus
Primary Percutaneous Intervention" )),
```

to get `Interventional_code`.


ocre:Study_characteristic
ocre:Recruitment_status
ocre:Active_not_recruiting
ocre:Completed_recruiting
ocre:Enrolling_by_invitation
ocre:Not_yet_recruiting
ocre:Recruiting
ocre:Suspended
ocre:Terminated
ocre:Withdrawn
ocre:Study_design
ocre:Allocation_concealment_type
ocre:Allocation_type
ocre:Deterministic_allocation
ocre:Randomized_allocation
ocre:Restricted_randomization
ocre:Simple_randomization
ocre:Analysis_of_cost_flag
ocre:Include_analysis_of_cost
ocre:No_analysis_of_cost
ocre:Assignment_characteristics
ocre:Intervention_cardinality
ocre:Multiple_intervention
ocre:Single_intervention
ocre:Investigator_assignment_role
ocre:Investigator_assigns_intervention
ocre:Investigator_assigns_no_intervention
ocre:Multiple_intervention_timing
ocre:Sequential_assignment
ocre:Simultaneous_assignment
ocre:Blinding_type
ocre:Investigator_blinding
ocre:Outcomes_assessor_blinding
ocre:Statistician_blinding
ocre:Subject_blinding
ocre:Control_group_characteristics
ocre:Control_group_cardinality
ocre:Control_group_locality
ocre:External_control_group
ocre:Within_subject_control_locality
ocre:Control_group_type
ocre:Case
ocre:Matched_cohort
ocre:Nested_case-control

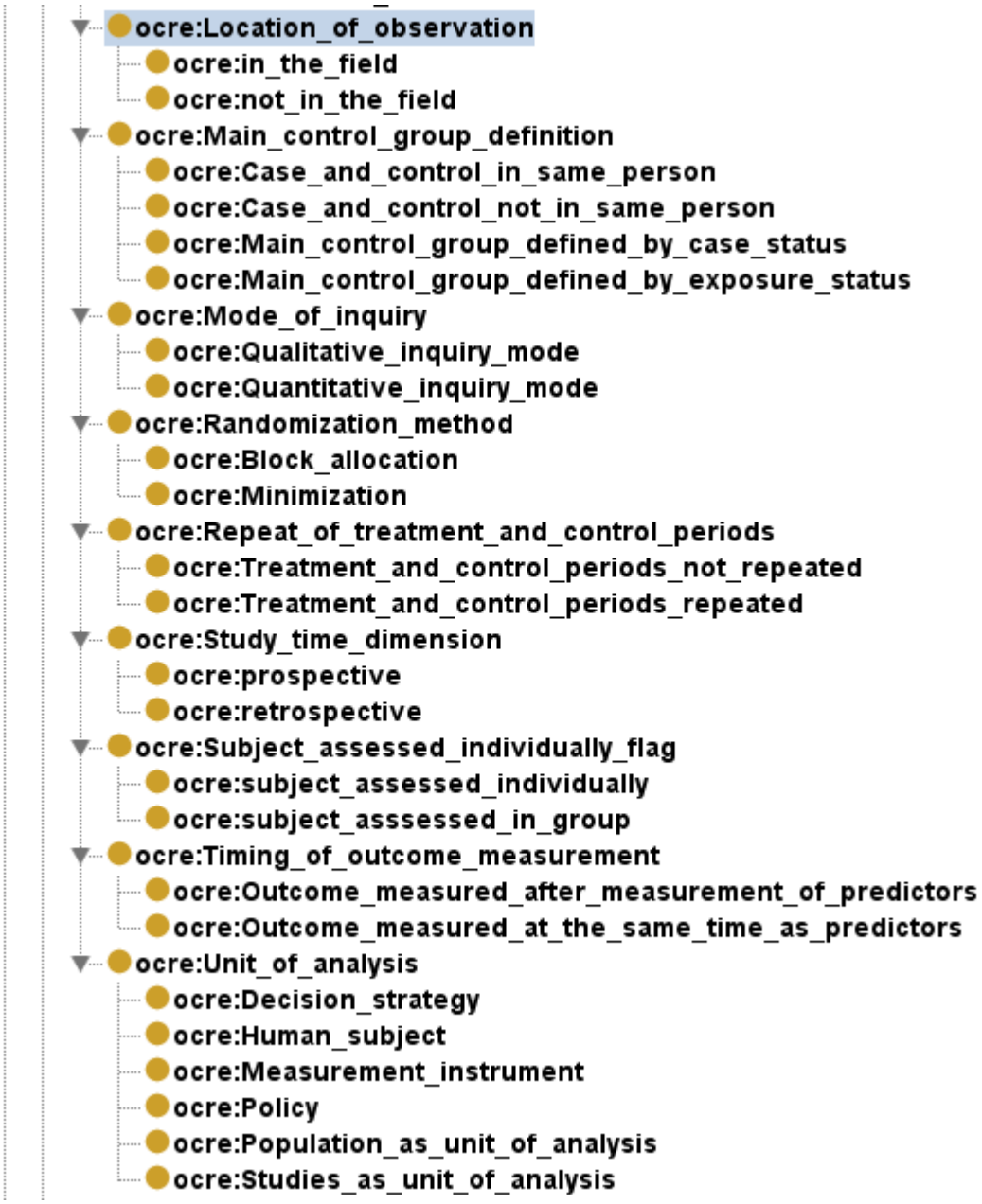


Figure 4 Study characteristics (see also StudyDesignTypeClassification.6.ppt)

Study Protocol

A study protocol is an informational entity that specifies how a study should be carried out. We model the following components of a study protocol:

Eligibility Criteria

Eligibility criteria are clinical statements that are potentially applicable to a subject. A study protocol usually records the criteria in abstract, data-model independent language (e.g., "Creatinine < 1.5 mg/dL," "has diagnosis of breast cancer"). A clinician interprets the criteria based on his or her background knowledge. "Creatinine < 1.5 mg/dL," for example, can be

interpreted to mean that a measurement of the density of creatinine in a blood sample should be less than 1.5 mg/dL.

To enable the possibility of scanning a database for potentially eligible subjects, instead of modeling the criterion "Creatinine < 1.5 mg/dL" as a constraint on the density of creatinine in a blood sample, we formulate it as a query expression, in some expression language, written in terms of a data model and a standard terminology. Depending on the data model, for example, "Creatinine < 1.5 mg/dL" can be encoded as a *variable* “Creatinine_in_mg/DL” whose value is less than 1.5 or an instance of *Observation* whose code is the SNOMED term 113075003 and whose value is a Physical_Quantity where the value is less than 1.5 and the unit is the SNOMED term 258797006.

Recognizing that the expression specifying an eligibility criterion can be very complex, the OCRe ontology incorporates expression templates similar to those in other medical knowledge modeling projects (Tu and Musen 2001; Tu, Campbell et al. 2007). These expression templates come from ERGO, which models the structure of the information content of the criteria . They consist of patterns of *clinical statements* that are composed from *expressions* and *noun phrases*. For example, the criterion "Creatinine < 1.5 mg/dL" is a *comparison clinical statement* composed of (1) a comparison operator "<″, (2) a query expression for the value of Creatinine among instances of Observations, and (3) the Physical Quantity constant 1.5 mg/dL. Multiple statements can be connected with logical operators to form compound statements. The expression templates include functions, variables, queries to a patient information model, noun phrases, and complex data types. Noun phrases can either be primitive terms from standard vocabularies or constructed using modifiers.

In this conception, eligibility criteria are informational entities whose semantics is not formally captured in OCRe. It is possible to specify a partial semantic constraint relating a criterion to events modeled in OCRe. Thus, the instance of comparison clinical statement that expresses "Creatinine < 1.5 mg/dL" has the semantic constraint (in Manchester OWL syntax)

```
has_semantic_constraint some
     (top:Observation and
     has_finding some snomed:creatinine_measurement_serum
 and has_value some(Physical_quantity
                        and has_unit some snomed:mgPerdL
                        and realvalue some float[<= 1.5]))
```

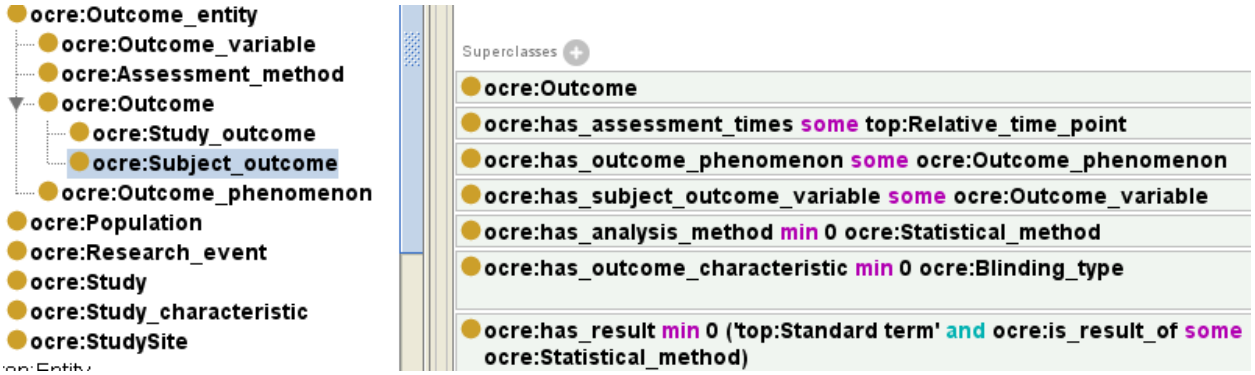
Study intervention and Health condition studied

A study protocol holds statements about study intervention and health condition studied. We model them as object properties `has_study_intervention` and `has_health_condition_studied` that relate a study protocol to the terminology codes that represent the interventions and health conditions.

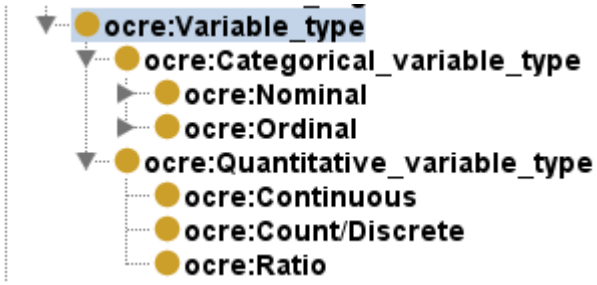
Outcome

Outcomes are informational entities specified in a study protocol. It can be divided into two types: Subject_Outcome and Study_Outcome.

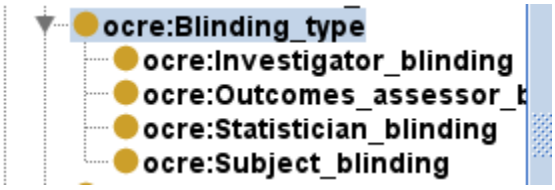
A subject outcome is a description of an outcome for a particular subject participating in the study. An outcome has the following properties:



1. Outcome phenomenon: A description of the phenomenon (e.g., clinical phenotype such as high blood pressure) being assessed, and the method of assessment. An outcome may involve the assessment of multiple phenomena (e.g., stroke, death, and MI) and define an outcome variable based on all phenomena assessed for the outcome. An Assessment_Method has some coded annotation (e.g., Hamilton scale), a text description, and one or more assessment_characteristics (reliability, reproducibility, inter-rater variability, etc.; subclasses of ocre:Performance_characteristics). How to deal with cost of treating MI (is the phenomenum 'cost'? or 'treating MI' or 'MI?
2. Assessment time: The time points at which the outcome phenomena should be assessed
3. Subject outcome variable: An Outcome_variable has an expression that specifies how the value of outcome should be computed from the assessed outcome phenomena (e.g., the presence of "first event of either death, stroke, MI within 6 months after index MI") and a set of outcome variable characteristics such as descriptive statistics (e.g., mean) and variable type (e.g., continous).



4. Outcome characteristics: A characterization of the outcome. Possible outcome characteristics include Blinding_type



5. has_analysis_method 0 or more Statistical method

6. has result (0 or more) a standard term (e.g., slope) that is_result_of some Statistical method (e.g., linear regression).

A Study outcome is the outcome aggregated over the subject outcomes.

1. Statistical method used to analyze the outcome
2. Study outcome measure: outcome aggregated over all subjects (e.g., count of subject outcome of specific kind, average of subject outcome), and possibly aggregated over other study outcome measures (e.g., proportional hazard model, really any multivariable analysis)

A study may distinguish between primary and secondary outcomes or characterize adverse events as a kind of outcome.

Examples of outcomes (TGN412 protocol)

Safety:

- Number and phenotype analysis of lymphocyte subsets
- Serum levels of selected inflammatory cytokines
- C5a as a marker of complement activation
- Anti-TGN1412 antibody formation
- Epstein-Barr viral load
- Rheumatoid factor (RF), anti-nuclear antibodies (ANA)

Pharmacokinetics

predicted time of TGN1412 systemic exposure (as determined from animal studies). The peak concentration of TGN1412 in the serum (C_{max}) and the time to reach peak concentration (t_{max}), overall systemic exposure (AUC_{0-t}, $AUC_{0-\infty}$), the apparent terminal elimination half life ($t_{1/2}$), volumes of distribution at steady state (V_{ss}) and associated with the terminal rate constant ($V_{\lambda z}$), total clearance (CL) and the mean residence time (MRT) will be determined.

Pharmacodynamic endpoints

PD endpoints considered appropriate for this first-in-man trial include:

- The absolute and relative numbers of lymphocyte subsets
- The systemic cytokine release profile
- Ex-vivo assessment of T-cell function in response to mitogens or recall antigens

Phase II Study of Allogeneic Large Multivalent Immunogen Vaccine and Aldesleukin in Women With Stable Metastatic Breast Cancer

http://www.cancer.gov/clinicaltrials/UMN-2007LS094

Outcomes

Primary Outcome(s)

Percentage of patients achieving complete response, partial response, or disease stabilization as assessed by RECIST criteria

Secondary Outcome(s)

Immune response as assessed by delayed-type hypersensitivity response to allogeneic large multivalent immunogen (LMI) vaccine; IFN-γ production by tumor-specific CD8+T cells; and CD8+T-cell binding to HLA-A2 multimer breast cancer-derived peptide complexes
Progression-free survival
Overall survival rate at 1 and 2 years
Safety and toxicity

RECIST Response Criteria

Evaluation of target lesions

* Complete Response (CR):	Disappearance of all target lesions
* Partial Response (PR):	At least a 30% decrease in the sum of the LD of target lesions, taking as reference the baseline sum LD
* Progressive Disease (PD):	At least a 20% increase in the sum of the LD of target lesions, taking as reference the smallest sum LD recorded since the treatment started or the appearance of one or more new lesions
* Stable Disease (SD):	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD, taking as reference the smallest sum LD since the treatment started

Planned activities and schedule of activities

We import a subset of BRIDG release 1.1.1 classes related to the planned activities and their schedules. Our intention is to import only the barest essential classes and attributes that are needed to specify important aggregations (e.g, planned epoch, planned subject study encounter, and planned arms) and the relative sequencing of planned activities (e.g., planned arm schedule of activities). Figure 5 is a schematic view of the classes and relations from BRIDG 1.1.1 that we model in OCRe. The UML aggregation relation is represented as '`is composed of`' relation in the figure. Figure 6 shows the OWL representation of the BRIDG classes. The UML relations are represented as OWL object properties.

The OCRe representation of the planned activities and their schedule follows that of BRIDG except that, instead of modeling a `PlannedStudy` as a subclass of `StudyProtocol`, as it is done in BRIDG, we model a study protocol as having a `has_planned_study` relation to an instance of `PlannedStudy`.

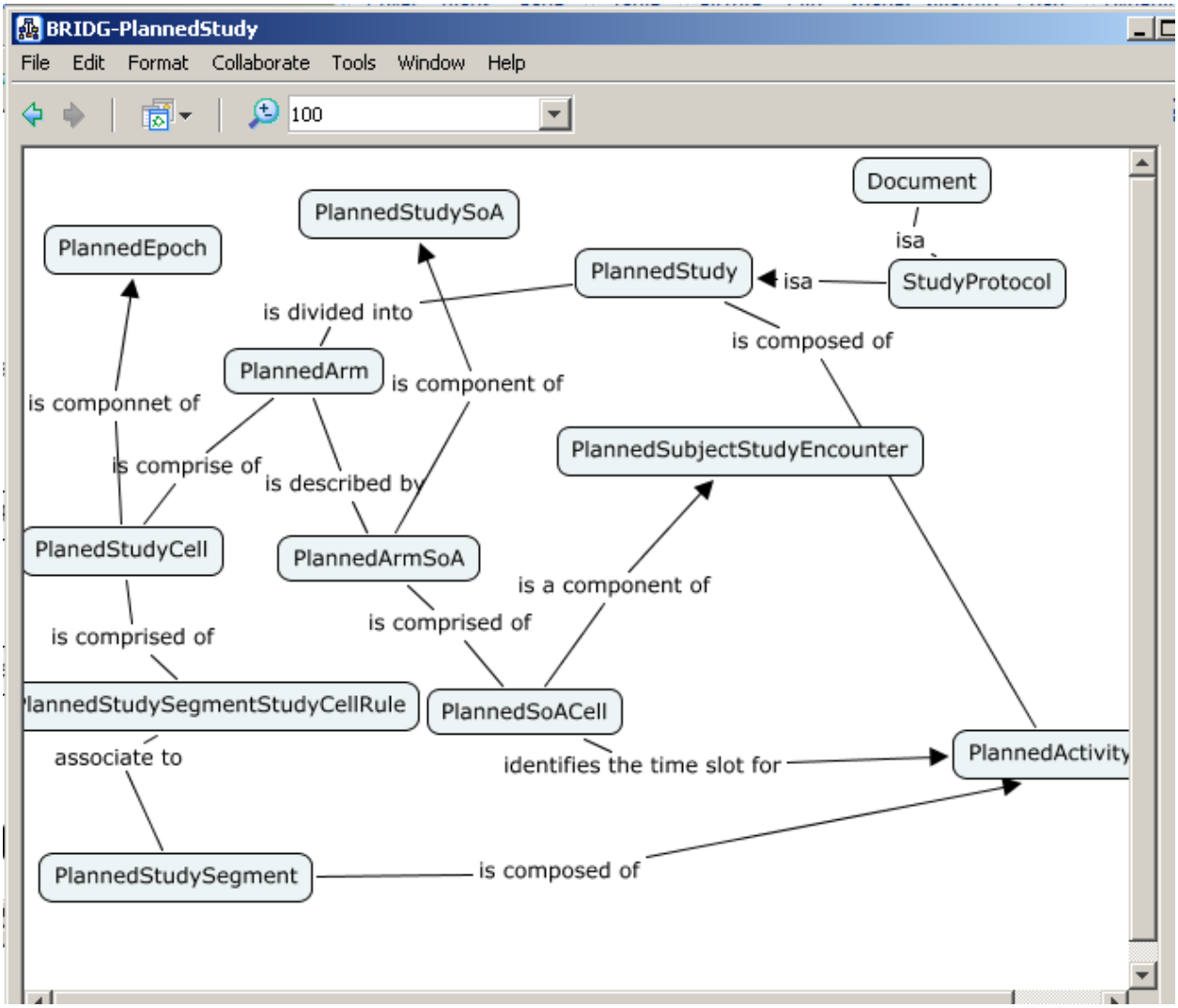


Figure 5 Schematic view of BRIDG planned entities modeled in OCRe.

Figure 6 OCRe representation of classes imported from BRIDG. The UML relations are represented as OWL object properties.

Mapping to Basic Formal Ontology

We had not attempted to use an upper ontology at the beginning of the OCRe project. However, as our conceptualization evolved, OCRe top-level classes proved to be compatible with BFO. Figure 7 shows how top-level OCRe classes relate to the classes in BFO. OCRe informational entities such as `Study_Protocol`, `Outcome`, BRIDG `Planning Entity`, and ERGO criterion templates are kinds of `bfo:generally_dependent_continuant`; OCRe `Study` and `Physical_Entity` are `bfo:object` and OCRe `Population` and `Organization` are `bfo:object_aggregate` and OCRe `Event` is a kind of `bfo:process`.

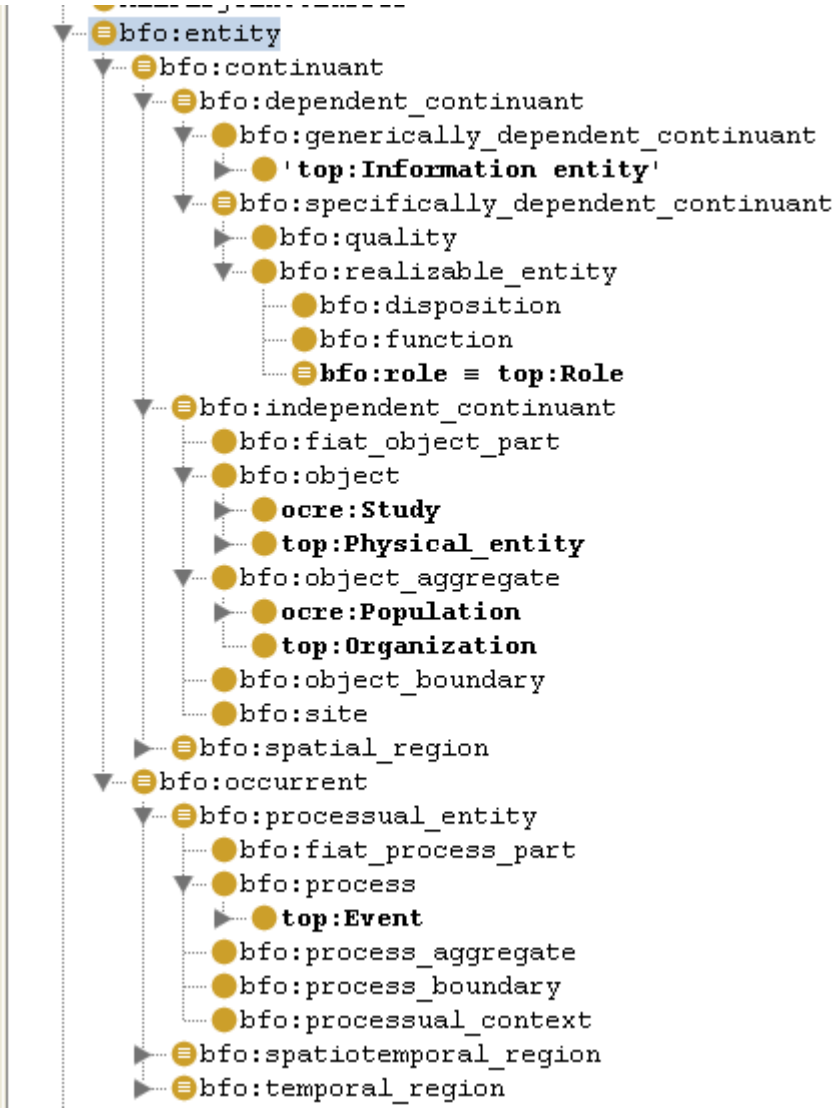


Figure 7 Embedding of top-level OCRe terms in Basic Formal Ontology. The terms marked by name-space prefixes are OCRe classes.

Discussion

Conclusion

OCRe formally defines an ontology for the design and methods of human studies. It has terminological components that specifies standard terms for characterizing studies and informational components that map to conventional information models. OCRe complements other technologies for cross-study queries across information systems for translational research.

References

Bittner, T. and B. Smith (2004). Normalizing Medical Ontologies using Basic Formal Ontology. GMS, Innsbruck/Tirol, Austria.

Fridsma, D. B., J. Evans, et al. (2008). "The BRIDG Project: A Technical Report." J Am Med Inform Assoc **15**(2): 130-137.

Grenon, P., B. Smith, et al. (2004). "Biodynamic ontology: applying BFO in the biomedical domain." Stud Health Technol Inform **102**: 20-38.

Gruber, T. R. (1993). "A Translation Approach to Portable Ontology Specification." Knowledge Acquisition **5**: 199-220.

Health Level Seven. (2006). "HL 7 Reference Information Model." 2008, from http://www.hl7.org/library/data-model/RIM/modelpage_mem.htm.

Motik, B., P. F. Patel-Schneider, et al. (2006). "OWL 1.1 Web Ontology Language: Structural Specification and Functional-Style Syntax." 2008, from http://www.w3.org/Submission/2006/SUBM-owl11-owl_specification-20061219/.

Niland, J. (2007). "ASPIRE: Agreement on Standardized Protocol Inclusion Requirements for Eligibility." Retrieved August 20, 2008, from http://hssp-cohort.wikispaces.com/space/showimage/ASPIRE+CDISC+Intrachange+July+10+2007+Final.ppt.

Rector, A. (2003). Modularisation of domain ontologies implemented in description logics and related formalisms including OWL. Proceedings of the 2nd international conference on Knowledge capture, Florida, USA.

Rector, A., R. Qamar, et al. (2006). Binding Ontologies & Coding Systems to Electronic Health Records and Messages. Proc of the Second International Workshop on Formal Biomedical Knowledge Representation (KR-MED 2006).

Shankar, R., M. O'Connor, et al. (2006). Epoch: an Ontological Framework to Support Clinical Trials Management. International Workshop on Health Information and Knowledge Management (HIKM 2006), Arlington, VA, ACM Press.

Smith, B., W. Kusnierczyk, et al. (2006). Toward a Reference Terminology for Ontology Research Development in the Biomedical Domain. KR-MED 2006.

Tu, S. W., J. R. Campbell, et al. (2007). "The SAGE Guideline Model: achievements and overview." J Am Med Inform Assoc **14**(5): 589-598.

Tu, S. W. and M. A. Musen (2001). Modeling Data and Knowledge in the EON Guideline Architecture. MedInfo, London, UK.

Tu, S. W., M. Peleg, et al. (2008). ERGO: A Template-Based Expression Language for Encoding Eligibility Criteria.