The Initial ACGT Master Ontology

Project Number:	FP6-2005-IST-026996
Deliverable id:	D 7.2
Deliverable name:	The ACGT Master Ontology
Submission Date:	31/05/07

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	ACGT
Project Full Name:	Advancing Clinico-Genomic Clinical Trials on Cancer: Open Grid Services for improving Medical Knowledge Discovery
Document id:	D 7.2
Document name:	The Initial ACGT Master Ontology
Document type (PU, INT, RE)	INT
Version:	4.0
Submission date:	13/06/07
Editor: Organisation: Email:	Mathias Brochhausen IFOMIS mathias.brochhausen@ifomis.uni-saarland.de

Document type PU = public, INT = internal, RE = restricted

ABSTRACT: This deliverable constitutes Major Project Milestone M5. It consists of the ACGT Master Ontology (OWL file) accompanied by comments regarding technical details, principles of ontology development employed the role of the ontology for other elements of the ACGT framework, and the possibilities of maintenance and extension of the ontology.

KEYWORD LIST: Ontology, Ontology Development, Ontology Maintenance

MODIFICATION CONTROL			
Version	Date	Status	Author
1.0	11/05/07	Draft	M. Brochhausen
2.0	18/05/07	Draft	M. Brochhausen
3.0	21/05/07	Draft	M. Brochhausen
4.0	31/05/07	Final	M. Brochhausen

List of Contributors

- Mathias Brochhausen, USAAR
- Cristian Cocos, USAAR
- Martin Doerr, FORTH
- Norbert Graf, USAAR
- Luis Martin, UPM
- Holger Stenzhorn, USAAR
- Gabriele Weiler, FHG (IBMT)

Contents

1 Executive Summary

This deliverable comprises the initial ACGT Master Ontology (MO) in OWL format. The ACGT MO constitutes Major Project Milestone M5. This document aims at giving directions regarding the usage and scalability of the ontology, and the challenges that have to be addressed in the coming time of the project.

In **3.1** the technical structure of the ontology is described. Sources that have been used, e.g., the top level ontology, are specified. **3.2** gives an account of the actual process of development of the ACGT MO. The process was based on clinical needs, which is documented by the close collaboration with the clinical partners in the ACGT consortium. The principles and methods applied in the ontology development are given in **3.3**. This chapter proves that the ACGT MO is compliant with the state of the art in ontology development. **3.4** identifies options of quality management in ontology development, and shows how they will be applied to the ACGT MO.

Chapter 4 details the use of the MO within the ACGT system, by employing the mediator (**4.1**), and the Trial Builder (**4.2**).

5.1 specifies future needs of the ACGT system which are relevant to the ontology. **5.2** describes plans regarding the maintenance of the MO, and spells out the future perspective on the matter. This topic will be addressed in the work period that follows the present deliverable. **5.3** shows how the MO can be extended using the Trial Builder. Extending the MO is crucial to the success of the ACGT project, as more clinical trials need to be covered in sequential manner.

2 Introduction

Purpose of this document

This document accompanies Deliverable 7.2 which represents Major Project Milestone M5. The deliverable consists of the initial ACGT Master Ontology. This document describes the technical features of the deliverable, and the principles and methods that were used to develop the ontology. The utilization of the ACGT Master Ontology in the technical setting of the project will be clarified. Last, the document will also address questions of maintenance and extension of the ontology in the future.

We understand this document as a means to inform the project partners of MO's features and potential. It is intended, among others, to facilitate usage of the ontology and the development of new tools which are based on it. Furthermore, it will clarify some aspects of ontology development and maintenance, which is needed in order to promote greater acceptance of this ontology among potential users from clinical practice.

3 The ACGT Master Ontology (MO)

3.1 Structure of the Deliverable

The intention of the ACGT MO is to represent the domain of cancer research and management in a computationally tractable manner. As such, we regard it as a domain

ontology. The ontology has been developed in the description logic based language OWL-DL [1], using the editor Protégé [2]. It contains around 1100 classes and more than 120 restrictions at the time of this writing.

In order to provide a consistent and sound representation, the ACGT MO employs the resources of a **Top Level Ontology** or **Upper Level Ontology**, which is, according to the **Standard Upper Level Ontology Working Group** of **IEEE**, "limited to concepts that are meta, generic, abstract and philosophical, and therefore are general enough to address (at a high level) a broad range of domain areas. Concepts specific to given domains will not be included; however, this standard will provide a structure and a set of general concepts upon which domain ontologies (e.g. medical, financial, engineering, etc.) could be constructed" [3]. We haven chosen the Basic Formal Ontology (BFO) [4] as Top Level for the ACGT MO since BFO has proven to be highly applicable to the biomedical domain [5]. Further information on BFO is given in 3.3 below.

The ACGT MO not only represents classes as linked via the basic taxonomical relation ("*is_a*"), but connects them via other semantic relations called "properties" in OWL terminology. OBO Relation Ontology (RO) [6] has been used as a basis in this regard, as RO has been specifically developed to account for relations in biomedical ontologies [7].

3.2 Representing clinical reality – basic steps and refinement

The need for the development of a new ontology for the field of cancer research and management was demonstrated by the state of the art review included in **Deliverable 7.1** [8]. Nevertheless, this does not preclude that, for some aspect of the ACGT domain, "heavyweight ontologies" (see below) with high usability could be identified and reused within ACGT. This, as a matter of fact, applies both to the **Foundational Model of Anatomy** (FMA) [9] as well as to the **Gene Ontology** (GO) [10] since they fulfil the quality requirements specified in 3.4.

The first challenge in developing the ACGT MO was the large scope of the project, integrating, among others, clinical studies, genomic research and clinical cancer management and care. This could, hence, easily be regarded as consisting of multiple domains. Nevertheless, we speak of the ACGT MO as a domain ontology in the sense that it represents the reality in the domain covered by the ACGT project in a uniform way. We have to be aware that the ontology could be broken down into a number of specific and highly reusable, particular domain ontologies, e.g. Clinical Trial Ontology or Patient Management Ontology. The use of FMA and GO shows how some existing ontologies are indeed used to cover restricted areas of reality or knowledge. In general, we renounced from the creation of several smaller ontologies in order to provide a unified representation of the complete domain in an orthogonal way.

Another challenge of the ACGT MO is to represent clinical reality in cancer management in a highly accurate and consistent way. This means that clinical reality has to be the basis of the representation, and that the result will prove highly usable in computer applications, like, e.g., the ACGT environment. Therefore, the ontology development team has to pursue the goal of active and extensive interaction with all of the clinical partners in the project.

The process that gave rise to the present state of the representation of clinical reality was rather convoluted and elaborated, requiring multiple recurring steps and a multifaceted approach. First, actual Case Report Forms (CRF) from ACGT trials were collected and analysed with respect to the universals (classes) more-or-less explicitly present in the information gathered. In parallel, basic aspects of cancer pathology and cancer management were studied by our researchers. The outcome of these activities provided the basic information on the universals and relations (properties) captured in the ACGT MO. This

ontology prototype was made available to all partners in the project. In addition, clinical partners were asked to review the prototype with respect to clinical accuracy, and technical partners for reviews on the usability. Based on the results of these reviews, the ontology was refined step by step keeping up the collaboration with all partners in the consortium and asking for their constant review of results.

We are also aware of the fact that the structure of the present ontology may not fit common clinical thinking or decision making; however, that is so for the sake of maintaining a concise and formal ontological approach [11]. Furthermore, we believe in keeping the ontology transparent to the end user. Therefore, user-friendly ways to access the ontology and make the information inside useable for the professional applying it have to be developed. The user will access the ontology only indirectly through specialised tools such as an ontology-based tool for the creation and management of new clinical trials which can easily be operated by clinicians without any background in ontological engineering or knowledge management. This tool is currently being developed by FHG in collaboration with USAAR and UHok. How this tool is employing the ontology is specified further in section 4.2. It is planed that this tool will become a major component of the extension of the MO, in order to open it to new diseases and new approaches (s. 5.3).

3.3 Principles and methodology used in the development of the ACGT Master Ontology

3.3.1 Principles

Lassila and McGuiness [12] categorised ontologies according to the information they need to express. Even though their classification ascribes the name “ontology” to nearly everything that is at least a finite controlled vocabulary with unambiguous interpretation of classes and term relationships and with strict hierarchical subclass relationships between classes, ontologies that meet more elaborate criteria contain a much richer internal structure, and have therefore been dubbed “heavyweight ontologies” [13]. Among the mentioned criteria are: formal “*is_a*” relation, properties, value restrictions, general logical constraints, and disjoints. In this sense we understand the ACGT MO indeed to amount to such a “heavyweight ontology” The ontologies that were subject of the review in **Deliverable 7.1** failed most of the more elaborate criteria.

The ontology is being developed by IFOMIS (in collaboration with FHG, FORTH and UPM) using the input received from the clinical partners in the project and available state-of-the art knowledge. The basic structure of the ontology as presented in **Deliverable 7.2**, was developed manually in order to secure the high standard of reality/knowledge representation mentioned above. For future maintenance and extensions, automated and semi-automated procedures have to be devised. This is further specified in **Section 5** of this document.

We have derived four of the principles used in the development of MO directly from BFO [4]. These principles are: *realism, perspectivalism, fallibilism, and adequatism* (see Glossary). We believe that these principles are a crucial part of the attitude one has to adopt regarding the development of any ontology. Besides those principles, several aspects of a more technical nature were taken into account by the developers of the ACGT MO such as fulfilling the basic needs of the developers of ontology-based applications/tools.

A basic principle of ontology development is that ontologies include only classes (types, universals) but not instances (tokens). Hence the ACGT MO does not include real world instances but only universals. The hierarchy of the universals is one of the major features of any ontology which may serve to prove its consistency and compliance with other formal standards. One of the gold standards to be followed in order to ensure a proper structure of the taxonomy of universals, is the avoidance of “informal *is_a*” relations [12].

In general, we embrace the belief that a properly constructed ontology should steer clear of a taxonomical tree that allows multiple parent classes for the same child class (i.e. one child that inherits from multiple parents). The central aim is to avoid polysemy that often results from multiple inheritances. In the ACGT MO we completely avoided multiple inheritances.

Another problematic case that can be found in quite a number of medical databases, terminologies and even "ontologies," is the presence of classes or types like "*UnknownX*" ("*UnspecifiedTumorStage*", "*UnknownAffiliation*"). However, "universals" like these do not, in fact, have any instances, they merely hint to a lack of data or knowledge. The alleged instances of those universals do not exhibit any shared properties, at least not necessarily. Therefore, we avoided such classes in the ACGT MO.

3.3.2 Methodology

Gómez-Pérez *et al* [13] describe the process of ontology development into three major components:

1) Ontology management activities
2) Ontology development oriented activities
3) Ontology support activities

Ontology management activities contain *scheduling*, *control* and *quality assurance*. The first two activities are not the subject to this deliverable, since they have been accomplished in close cooperation with **WP1** and the **Technical Management Board** of ACGT. Quality assurance will be addressed in more detail under section **3.4**.

This deliverable presents the result of the activities under 2) namely, environmental study and feasibility study (both pre-development), specification, conceptualization, formalization and implementation[1] (Development). Maintenance and use which constitute the post-development are not a major topic of this deliverable, although we give an outlook regarding maintenance under section **5**.

Regarding 3) the most important work that has been accomplished was knowledge acquisition.

In the following some specific techniques and modes of work are described which were followed in the development of the ACGT MO. In accordance with Uschold and King [14] we think that the development of an ontology should not be centered around a decision whether bottom-up, or top-down. Nevertheless, we did not decide in favour of the middle-out approach proposed by Uschold and King, but decided to proceed from both directions combining a bottom up approach, which was focused on the CRFs from SIOP trial and the TOP Trial in ACGT, with a top-down approach, which utilized the reality representation in BFO.

In [15] Noy and Musen described the technique of alignment that uses the virtues of existing ontologies along with a new one, while retaining separate ontologies. The ACGT MO should be in alignment with existing ontologies, e.g. the FMA [9] and GO [10]. Some existing medical classifications will be slightly modified and added to the ontology. An example of this type is the TNM system [16, 17].

[1] Implementation in this context "builds computable models in an ontology language" [13]. Therefore, the ACGT MO is implemented.

3.4 Quality assessment in ontology development

Even though means to classify ontologies regarding their complexity and the amount of information represented in them do exist, ontology development needs to appeal to a standard of quality in light of theoretical soundness and conformity with reality. This aspect is addressed by aligning the ACGT MO with the OBO Foundry, which introduces "*a new paradigm for biomedical ontology development by the establishment of gold standard reference ontologies for individual domains of inquiry*" [18].

ACGT, hence, concurs with the concept of the OBO Foundry, and the ontology developers within the project aim at meeting the criteria of this organization. Our objective is for the ACGT MO to become a member of the OBO Foundry once it is fully completed. We, further, contend that all ontologies used by the mediator should also subscribe to the standard of OBO Foundry. This is already the case for FMA and GO.

4 Utilizing the ACGT Master Ontology in the technical setting of the project

4.1 The Master Ontology and the Semantic Mediator

During the last years, ontologies have been widely used as domain models for solving heterogeneities among different databases. In database integration, an ontology can act as a common framework to create virtual views of the data sources providing homogeneous access, and enabling database integration. There exist two main approaches to perform database integration, namely *Data Transformation* and *Query Translation*. In *Data Transformation*¸ actual data from different databases are converted and stored in one centralized repository (e.g. Data Warehouses), while in *Query Translation* data stay in their original sources, and a virtual view represents the integration. Given the nature of data in ACGT, a *Query Translation* approach has been adopted. Within *Query Translation*, there are two main ways to approach the problem of database integration: Global as View and Local as View [19]. The ACGT Mediator follows a Local as View-based approach. In Local as View, single representations of each of the data sources to be integrated are created by means of a global model. In ACGT, the Master Ontology acts as the global model, as it represents the domain of the underlying data.

The ACGT Semantic Mediator requires the creation of a view for every single data source using terms and relationships from the Master Ontology. These views are created through the *Mapping* process. During the *Mapping* process, correspondences between elements in the data sources and terms and relations in the Master Ontology are created. These correspondences are used to carry out the query translation. Different types of mappings can be performed depending on the expressiveness of the ontology language and the mapping format. Taking into consideration the power of OWL-DL, there is a wide range of possibilities, which we believe that, at the present stage of our analysis, are sufficient to cover user requirements.

Once a query is performed, the mediator splits it into the necessary queries dedicated to the underlying data sources. Every one of these queries passes through the mapping filter, which converts the terms and relationships from the Master Ontology to the original database vocabulary, generating the final queries in SPARQL to be sent to the database wrappers. The results are obtained in the database wrappers result set format. The mediator annotates them using the Master Ontology and finally retrieves an integrated set of results in OWL.

The role of the ACGT Master Ontology in database integration is twofold, 1) it supports the creation of homogeneous views representing the underlying data sources (the *mapping*

process), and 2) it serves as a vocabulary server to annotate the results of the queries, aiding to generate semantics-compliant result sets.

A suitable part of an ontology in a suitable encoding can be used or interpreted as target schema. The MO will be used as our Enterprise or Target Model in order to support the appropriate mappings from our local data schemata (Source models). These mappings will enable the integration under a common knowledge representation model (LAV approach) where data source relations are defined in terms of a global schema [20].

Mapping specifications should be given by domain experts and should be expressive enough to allow an IT-expert to configure the respective wrapping and mediation services without further help from the domain expert. A tool is, therefore, required in order to assist the mapping specification process. In order to support a domain expert in the mapping specifications, it is beneficial to mark a layer in the MO which is *adequate* to the ontological level of detail of characteristic data structures in the domain. Further examples of mappings of characteristic schema constructs can be helpful. It may also be beneficial to mark subsets of the MO by context of application to generate personalized views of the MO.

4.2 The Trial Builder and the ontology based clinical data management system

A data source has to be mapped to the ACGT master ontology in order to be integrated into the mediator. The mapping process is an error prone and tedious task, but necessary for legacy data sources [21].

Because of this fact, the ACGT project also wants to explore the approach that databases of newly developed data management systems can be set up during creation in an ontology compliant way to allow a seamless integration of the data collected in these systems into the ACGT mediator architecture. Currently a user-friendly tool, called Trial Builder is under development, which will enable a chairman of a clinical trial to set up a clinical data management system with comprehensive metadata in terms of the ACGT-MO. This tool will allow clinical trial chairmen to capture data definition and further design specifications for a clinical trial in a standardized way, based on the ACGT Master ontology. Furthermore, tools will be provided, that allow to set up the clinical data management system for the trial from these definitions automatically.

This approach will make it attractive for clinical trial leaders to utilize the ontology in their trial since the ontology will be seamlessly integrated into user-friendly tools to design their trial and to set up the appropriate clinical data management system.

In the following we will briefly describe how the Trial Builder will allow a clinician to define easily all information needed to set up the CRFs and databases for the clinical data management system, as well as the metadata from the ACGT-MO. (ontology view).

A clinician aiming to design a trial will naturally want to focus on designing the CRFs and integrating them into the workflow of the specific clinical trial he wants to perform. He does not want to be bothered with databases or ontology views. The trial chairman will build up the later by designing the CRFs for his trials. With the help of the Trial Builder the trial chairman can define the questions on the CRFs, the order in which the questions will be queried and constraints on the answer possibilities. In principle for each question the following attributes have to be defined: the question itself, data type of the answer and optionally possible data values, range constraints and a measurement unit. Additionally in order to describe the answer of the question semantically, a description from the ontology has to be chosen. This description has to be a path from the ontology starting at the class patient since this class is always the context of the CRFs. The trial chairman has to choose only the appropriate path for an item from the ontology. As far as possible, all other attributes will be determined

automatically from this selected path, but can later be changed by the trial chairman to a certain extend.

Although an ontology is 'human understandable' by providing natural language definitions of entities and relationships it is described by definitions not based on practical or clinical perceptions of reality. Therefore the Trial Builder will provide an application specific view on the ontology in a way that a clinician can understand it and will guide the trial chairman to select appropriate paths to describe the questions on his CRFs from the ontology. The integration of other ontologies is described in section 5.3.

With this process the clinician will define the structure of the databases as well as the ontology view for the database. An appropriate clinical data management system can be set up automatically and the data collected in that system can be seamlessly integrated into the mediator architecture.

5 Maintenance and extension of the ACGT MO

5.1 Future perspectives

The aim of the ACGT project is clearly future oriented. The result of the IP will be a GRID-based infrastructure enabling the sharing of data, tools and ultimately scientific knowledge on cancer, with the objective to support translational research and the transfer of research results to the bedside for the benefit of patients..

In 3.2 we reported that, as a first step, we agreed to restrict the MO to the content of the trials which are part of the ACGT project, i.e. SIOP and TOP. Nevertheless, the aim is to operate the ACGT system for all other relevant domains of cancer, its biology, its pathology and its care. This means that there are two vital aspects to discuss in order to keep the system running and up to date, for as long as possible: *maintenance* and *extension*.

Section 5.2 addresses aspects concerning the update of the ontology, the inclusion of relevant information stemming from new scientific discoveries, as well as the reflection of ever-changing user needs into the system. Section 5.3 specifies the application of the Trial Builder as a means to extend the ACGT MO by making it available to medical researchers setting up and managing clinical trials on cancer.

5.2 Maintenance of the ACGT MO

We have to be aware that maintenance of an ontology does not only raise technical questions but also touches administrative and political aspects regarding the project.

From the technical point of view the following convictions are relevant: We expect that the amount of new information that needs to be integrated in the ontology will far exceed the amount that could be manually handled by a single curator. Yet, practice has shown that the best way to curate an ontology is curator-driven. Therefore, we aim at developing a network of MO maintenance tools, without, however, relinquishing the need for curator-validation of all transactions. It is in this way, we believe, that MO will remain concise and formal, as well as up to date. The technical means required to achieve such a goal concerns efforts to be started from month 16 onward.

From the administrational point of view it must be demonstrated that the ontology is effective, and that the maintainers are, as a matter of principle, open to any modification destined to settle inconsistencies and varying requirements with potential supporters, within the quality standards of the OBO Foundry. Nevertheless, we propose to mobilize a sufficient number of experts to contribute, by proving that it is in their own interest to do so. *Social tagging* has emerged as a buzzword in this respect, and refers to the process of eliciting engagement

and wide feedback. There must be an effective business plan to distribute expertise about the ontology in a hierarchy similar to that of the ontology itself. The particular workflow of distributed terminology maintenance is a well established practice in library and information science since for many decades.

We are aware of the necessity to provide a state-of-the-art solution for the maintenance of the ACGT MO. In order to meet best practice we will prepare an extensive state-of-the-art review.

5.3 Extension of the ACGT MO

ACGT sees itself as a pioneering approach to knowledge sharing between researchers and clinicians. Our goal, hence, must be to find as many researchers undertaking ACGT-based clinical trials as possible. In order to achieve this objective, ACGT will provide a Clinical Trial Builder that will be developed by FHG, USAAR and UHok. The ontology-grounded structure of the Trial Builder calls for extensions of the ontology in light of new trials. We envision the process of ontology extension described below to be executed by (or through) the trial builder application.

The Trial builder application will be used by clinicians to set up new clinico-genomic trials within the ACGT platform. With the help of the trial builder application new CRFs for a trial can be created. The clinician will have the ability to enter every question/item on a CRF. This can be done in two different ways:

1. A clinician can enter the item into an entry field. This entry field is directly connected to the MO and is indeed a search field of the MO. The MO will be searched by this item and the corresponding Thesaurus. Such a Thesaurus is needed and implemented in the system. Having found the item in the MO a clinical view of the MO is presented to the clinician depicting the item and the dependencies in a tree diagram. This would allow the clinician to validate the correctness of the item and to even copy not only the searched item to the CRF but a whole branch or a part of it.
2. A clinical view of the MO is presented to the clinician in a graphical way (tree with branches). The clinician can, hence, easily parse the MO and select a single item or a whole branch of the clinical view in order to copy directly to the CRF.

In both cases the clinician is not aware that he is actually dealing with the MO.

Three possibilities may occur in case a clinician wants to add a new item to a CRF in the above described way.

1. The item (or a synonym of the item (found in an implemented Thesaurus)) is found in the MO:

   No further steps are necessary. By copying the item (or even a whole branch) to the CRF the link to the MO is automatically created. The end user is only confronted with the clinical view of the MO.

2. The item or a synonym of the item is not found in the MO, but in another existing Ontology fulfilling the quality requirements of the MO. For this task we will use search functions and browsers which are available for most ontologies and vocabularies. These Ontologies are called Linked Ontologies (LO). The creation of that list is part of the work to be done from month 16 onward.

   In this case the branch of the LO, in which the item was found, will be linked to the MO. By walking through the clinical view of the MO the clinician labels the parent branch of the MO to which the branch of the LO will be automatically added. The added branch will be automatically visualized as a new branch in the MO. The Clinician can accept or change the connection side of both Ontologies (MO and LO). This results in an extended MO. After doing this, the clinician can copy the

item or a whole branch of the extended MO to the CRF. In the background a process will start immediately to find inconsistencies in the extended MO. The curator of the MO will be automatically informed by email about the extension of the MO. If no inconsistencies exist, the extended MO will be accepted as the new MO. If inconsistencies exist, the inconsistencies will be removed by the curator of the MO. In this case the link to the item on the CRF has to be maintained in every case. After clearing all inconsistencies the extended MO will be the new MO. The next time a clinician needs the same item in a new CRF, he will run through point one of this section.

The clinician himself will obviously not be aware of this process, which is taking place completely behind the scene.

3. The item or a synonym of the item is not found in the MO, nor in any other Linked Ontology:

   In this case the clinician adds this item to a selected branch of the MO, wherever he believes that it is the correct place. This is done by walking through the clinical view of the MO and labelling the parent branch of the MO. By labelling the item will be automatically added to the MO. This would result in an extended MO. After doing this, the same process will start as described in point two, to clear inconsistencies. After this process is completed, the so extended MO will be the new MO.

It is of utmost importance to realize that this process will avoid that the MO will be neither a needle eye for new clinico-genomic trials within ACGT, nor the flow of ACGT processes in general. It is no longer necessary to wait for an extended version of the MO including a complete set of all items for a new trial. One can start with creating new CRFs for a trial regardless of the content of the MO. By using the trial builder application in the above described way, the MO will automatically be extended, and only meaningful items will be added; namely those that are needed and used in clinico-genomic trials. The more clinico-genomic trials will use the trial builder via the ACGT platform, the faster the MO will grow. Again, this identifies one more task for the work to be done in the next months and we plan to provide a review on common practices in this area.

Appendix 1 - Abbreviations and acronyms

ACGT	Advancing Clinico-Genomic Trials on Cancer
BFO	Basic Formal Ontology
CRF	Case Report Form
CTO	Clinical Trial Ontology
FMA	Foundational Model of Anatomy
GO	Gene Ontology
LAV	Local-as-view
LO	Linked Ontology
MO	Master Ontology
OBO	Open Biological Ontologies
OWL-DL	Web Ontology Language – Description Logic
RO	Relation Ontology
SIOP	International Society of Paediatric Oncology
TOP	Test of Principle (Breast Cancer Tria)l

Appendix 2 - The Initial ACGT Master Ontology OWL-file

The file containing the Initial ACGT Master Ontology in OWL format can be downloaded from the internet. The URL is:

http://www.ifomis.org/acgt/1.0

Glossary

Adequatism This is the position that a good theory of reality must do justice to all of the different phenomena that reality contains. In opposition to the tendency to attempt to reductively explain higher level macroscopic phenomena in terms of "more basic" or fundamental components of reality, adequatism entails that the entities in any given domain of reality be taken seriously on their own terms first. Thus, just as an ontology of physics should be about atoms and sub-atomic particles, and an ontology of chemical reactions should include the existence of various kinds of elements and compounds, so an ontology of biological phenomena should include the existence of, at various levels, cells, organs, biological systems and organisms, as well as populations and environments. The goal of adequatism is to do justice to the vast array of different kinds of entities that exist in the world, in different domains and at different levels of granularity, rather than ignoring them or attempting to explain them away.

Fallibilism Fallibilism involves commitment to the idea that, although our current scientific theories are the best candidates we have for the truth about reality, it may nevertheless be the case that portions of our current knowledge are incorrect, hence our current purported reality representations are *not* representations after all. The fallibilist maintains that it is a matter of empirical investigation what the facts of reality are, and recognizes that empirical investigation is an ongoing, open-ended, experimental process.

Perspectivalism Perspectivalism involves the recognition that reality is a complex and variegated phenomenon. While not all purported representations of reality are good, because some are accurate to the facts of reality and some are not, there are nevertheless many different equally good representations (good in the sense of being true), precisely in that they capture different and important features of one and the same reality, that is, they capture competing angles of investigation.

Property An OWL property is a binary relation. Two types of properties are distinguished: first, datatype properties, relations between instances of classes and RDF literals and XML Schema datatypes. Secondly, object properties, relations between instances of two classes [22].

Property Restrictions When properties are defined there are a number of ways to restrict those relation. The domain (i.e. subject) and range (i.e. object) of the properties can be specified. The property can be defined to be a specialization (subproperty) of an existing property and also more elaborate restrictions (e.g. cardinality restrictions) are possible [18].

Realism 'Realism' can be defined as the view according to which reality and its constituents exist independently of our (linguistic, conceptual, theoretical, cultural) representations. Realism is the thesis that the things that scientific knowledge is about are in fact real, mind-independent things. Thus, ontologies are representations of reality, not representations of people's concepts or mental representations of reality.

References

[1] http://www.w3.org/2004/OWL

[2] http://protege.stanford.edu

[3] http://suo.ieee.org

[4] http://www.ifomis.org/bfo

[5] Grenon P, Smith B, Goldberg L "Biodynamic Ontology: Applying BFO in the Biomedical Domain." in: Ontologies in Medicine, D. M. Pisanelli, Ed., Amsterdam: IOS Press, 2004, pp. 20-38.

[6] http://obofoundry.org/ro

[7] Smith B, Ceusters W, Klagges B, Kohler J, Kumar A, Lomax J, Mungall CJ, Neuhaus F, Rector A, Rosse C "Relations in Biomedical Ontologies." *Genome Biology, 2005, 6:R46*

[8] ACGT "Consolidated Requirements on ontological approaches for integration of multi-level biomedical information." Project Number: FP6-2005-IST-026996, Deliverable id: D 7.1, 2007

[9] http://sig.biostr.washington.edu/projects/fm

[10] http://www.geneontology.org

[11] Rector AL, Zanstra PE, Solomon WD, Rogers, Baud R, et al. "Reconciling Users Needs and Formal Requirements: Issues in developing Re-Usable Ontology for Medicine." IEEE Transactions on Information Technology in BioMedicine 2(4), pp. 229-242, 1999.

[12] Lassila O, McGuinness D "The role of frame-based representation on the Semantic Web." Technical Report KSL-01-02. Knowledge System Laborartory. Stanford University, Stanford

[13] Gómez-Pérez A, Fernández-López M, Corcho O "Ontological Engineering." Springer, London, 2004

[14] Uschold M, King M "Ontologies: Principles, Methods, and Applications. Knowledge Eng. Rev., vol. 11, no. 2, pp. 93-155, 1996.

[15] Noy NF, Musen MA „SMART: Automated Support for Ontology Merging and Alignment." Twelfth Banff Workshop on Knowledge Acquisition, Modeling, and Management, Banff, Alberta, Canada, 1999.

[16] Wittekind C, Meyer HJ, Bootz F "TNM. Klassifikation maligner Tumoren" 6. Aufl., Springer, Heidelberg, 2005.

[17] http://www.uicc.org/index.php?id=508

[18] http://obofoundry.org

[19] Cali A, De Giacomo G, Lenzerini M "Models for information integration: Turning local-as-view into global-as-view." In: Proc. of Int. Workshop on Foundations of Models for Information Integration (10th Workshop in the series Foundations of Models and Languages for Data and Objects) (2001).

[20] Calvanese, D., Giacomo, G., Lenzerini, M., Vardi, M.: View-Based Query Answering and Query Containment over Semistructured Data. DBPL (2001a), 40-61

[21] Kalfoglou Y, Schorlemmer M "Ontology mapping: the state of the art." *The Knowledge Engineering Review* 18(1) pp. 1-31, 2003.

[22] www.w3.org/TR/owl-guide