Meeting Notes on Ontology Development and Maintenance
June 25-26, 2008, Saarbrücken
(Including Submission Process and Implementation Process)

Project Number: FP6-2005-IST-026996

Document name: Meeting Notes on Ontology Development and Maintenance

Submission Date: 04/08/2008

MODIFICATION CONTROL			
Version	Date	Status	Author
1.0	09/07/2008	Draft	M. Brochhausen
2.0	01/08/2008	Pre-Final Version	M. Brochhausen
3.0	04/08/2008	Final	M. Brochhausen

List of Contributors

    –   Mathias Brochhausen, IFOMIS

    –   Cristian Cocos, IFOMIS

    –   Martin Dörr, FORTH

    –   Catherine Dosch, IFOMIS

Contents

1 Introduction

Both the ontology development process (especially to the extent to which it concerns the issue of reality representation (*conceptual design*)), and the process of ontology maintenance, are issues that demand further clarification at this stage of the ACGT project. The latter has been on the agenda of WP7 for quite a while, and FORTH—namely Martin Dörr's group—has elaborated a general strategy which can be used by ACGT. With respect to reality representation, there are still some issues concerning the relation (*properties*) in the ontology, especially regarding the extent to which CIDOC CRM can guide the process of their representation. This approach appears to be quite promising, since the CIDOC CRM aims at representing scientific and documentation-related processes which certainly could be a model for the part of the ontology dealing with clinical trials.

A closely related issue concerns the possibility of rendering the ontology in a way that is easily accessible for clinical users. It is likely that the ontology be enhanced in a way that fosters the development of user-friendly ontology viewers.

In order to initiate and coordinate the efforts targeting these issues we decided to have a meeting. Due to practical reasons, we decided to have a small meeting preparing initial commendations on the topic to be discussed among all partners in WP7, project management, and clinical users. The meeting was held at IFOMIS in Saarbrücken, Germany, June 25-26, 2008. It was attended by the three contributors of this document.

We want to stress that this documents contains both, decisions with respect to implementation and common interests for future research. Throughout the document it will be made clear which elements will be implemented right now.

2 Meeting agenda

The issues that were addressed at the meeting had been clear to all parties involved from previous e-mail communications between the contributors, project partners in WP7, and project management. We took the decision to fix the agenda at the meeting in order to allow for maximum flexibility. The following aspects were discussed:

1) Status of the ontology

2) Modelling Scientific Explanation

3) "Context" or "Niche"

4) "Flagging"

5) Submission process

6) Implementation process

The results on each of these topics will be presented in one chapter of this document. IFOMIS was asked to provide a state-of-the-art review on submission systems in biomedical ontologies and terminologies. This was carried out by Catherine Dosch after the meeting and the result is given under chapter 7.

3 Status of the ontology

IFOMIS presented the current work on the ontology, which consists mainly of adding class restrictions, which are of huge importance to ObTiMA. We will not document the details of the work here, but plan to submit another document showing our efforts.

We agreed that it is necessary to represent properties of the highest possible order, viz. relations that hold among the highest possible classes. Yet, given the complexity of the domain, this proved to be rather difficult in many cases. We therefore conceded that, at least as a first step, a property should be attached to the lowest class to which it applies with certainty. Attempts should be made later on to move the relation upward in the hierarchy of the classes.

An important point that was discussed was whether ObTiMA should show all the restrictions attached to a class. While displaying them all might lead to displaying a number of rather abstract properties, we maintain that from a purely ontological point of view, ObTiMA needs to make use of all properties a class has, even those inherited from its super-classes.

Some other critical issues pertaining to the ontology were discussed, and some solutions were found:

Participates in: This, as well as similar relations, testifies to the complexity of biomedical processes, which often makes it difficult to figure out the exact function or role of an entity within the process at hand. Taking our inspiration from CIDOC CRM,[1] which contains the relation "was_present_at/in." IFOMIS agrees to keep this relation in mind. Hitherto no candidate has occurred in the ACGT MO, but it is agreed that such a relation might be useful.

Man-made qualities and classification: We agreed to avoid classes like "HighRiskBlastemaNephromblastoma." Man-made measurements for qualities, like "High Risk," or "Low Risk" cannot justify a new sub-class. Nevertheless, we might need those characterizations in the quality-branch.

"Primary tumor": "Primary tumor" has proved to be problematic in the way it is represented. The ontologically sound way would be to make this a quality, but for practical reasons–stemming mostly from the way the mediation and mapping processes are carried out–this option was discarded. The alternative was to allow multiple instantiation.

Solution:

Neoplasm has two is_a related groups: by histology (all types of tumors, e.g. Carcinoma, Sarcoma) and by history (PrimaryTumor, Metastasis). The members of these two groups are each mutually disjoint, but the groups themselves are not. Then we need a constraint that for ALL x, whare x is an instance of TumorHistoType1 (e.g.), x is an instance of PrimaryTumor or Metastasis.

It was decided to implement this solution, but there are still some issues to be pondered since the formalization of the constraint as given above is similar to a formal *is_a* relation (Every instance of A *is_a* instance of B). There are cases which might be problematic here, but this is an issue which will be solved within the next weeks.

The following principles for ontology development were mentioned and confirmed:

---

[1] The CIDOC CRM (ISO 21127:2006): http://cidoc.ics.forth.gr.

- ▪ Categorization into the class that is clear, maybe the super-class. If it is unclear whether it is A or B, given A and B are subclasses to C, it should be classified as C. Thereby, we avoid NOS (Not otherwise specified) classifications[2]

- ▪ A restriction should always be moved up to the highest relevant super-class.

- ▪ User's categorization of the domain might not fit ontological design principles. Therefore, it should not constitute the guideline for ontology development, but we should support efforts providing user specific views.

The participants agree that future research activity should include research on "facets". M. Dörr proposes that there are different facets to *is_a* relations. The following facets are proposed:

- ▪ Genesis/History
- ▪ Morphology/Histology
- ▪ Location
- ▪ Behaviour/Function

"Flagging" (s.b.), or making these four different types of *is_a* relation visible might be one way of dodging the multiple instantiation mentioned above. Yet, it is still an open question whether facets can be formalized at all. Obviously they go beyond the set theoretic framework which is the foundation of formal *is_a*.

Aside from the aspects mentioned above, a conference paper by former IFOMIS research Ingvar Johansson might be interesting in order to explore these issues in more detail.[3]

4 Modelling Scientific Explanation

We agree that scientific explanation provides an important framework which clarifies the structural aspects of the ontology, due to that fact that a huge number of the processes which have to be reported e.g. within clinical trials are scientific by nature. These considerations resulted in the addition of "ScientificObservation" as a sub-class of "Process." "MeasuringProcess" was moved as a sub-class of "ScientificObservation."

Of central interest is the representation of diagnosis. With respect to that, the following changes were agreed upon:

Move "Diagnosis" to "Document".

"Diagnosis" outcomeOf some "DiagnosticProcess".

---

[2] B. Smith, M. Brochhausen, „Establishing and Harmonizing Ontologies in An Interdisciplinary Health Care and Clinical Research Environment," in: B. Blobel, P. Pharow, M. Berlich (eds.), *eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge,* IOS Press, Amsterdam, 2008: 219-234.

[3] I. Johansson, "Four Kinds if "Is_A" Relation," WSPI 2006, Saarbrücken, May 3-4, 2006, http://hem.passagen.se/ijohansson/information7.pdf.

In implementing these changes we agreed to review the sub-classes of "InformationObject" which had been moved from "Object" to "GenericallyDependentContinuant" prior to the meeting.

The question was raised whether "Life" should be represented as "Process" or "ProcessAggregate." Final decision on this matter was, however, postponed, due to some discussion with the OBI group Cristian Cocos reported. Due to the multigranularity of ontologies, OBI decided to not make use of any Aggregate class ("ProcessAggregate" and "ObjectAggregate"). Yet it is not clear when this decision will be implemented into the BFO. Generally, we agreed that "Life" is an "OrganismalProcess," and we dropped "BiologicalProcess" and "ExtraorganismalProcess."

5 "Context" or "Niche"

ACGT supplies clinical users with a view of the ontology that is appropriate for their needs, and for their particular way of approaching the clinical trial domain. Aside from providing a patient-centred and relation-based view via ObTiMA, and aside from designing and implementing an Ontology Viewer which facilitates customization of the view, we have not found a formal way to provide knowledge about "views" or "contexts" in the very ontology. We generally agree that this is new research important to the field of ontological engineering, especially in the field of biomedicine. These problems have been contemplated some years ago by Rector *et al*[4]., but no general cure has been found and implemented yet.

We subscribe to the meaning of the term "niche" as construed by Berman and Semwayo:[5] The potential of this paper should be exploited further as a necessary step towards achieving the aims sketched above.

6 "Flagging"

Another strategy that can offer promise in meeting the aim described under 5 is the flagging of certain MO classes, which represents a view on the ontology with respect to a certain user. In Protege this can be achieved via the comment field.

Yet before we embark on this activity we need to clean up the properties in the ACGT MO. They are key in identifying the classes that should be flagged.

The flagged part of the ontology has to be functionally complete with respect to the questions we want to answer.

The following optional process was outlined in order to create the flagging:

Take two databases, take their schema, use Martin's mapping (Charis), analogous mapping.

Formulate query as path. Check whether you can answer the question based on the ontology and secondly whether you can answer the question based on the ontology using relation that you automatically take from the schema.

---

[4] A. L. Rector, P. E. Zanstra, W. D. Solomon, J.E. Rogers, R. Baud 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), 1999: 229-242.

[5] S. Berman, T.D. Semwayo, „A Conceptual Modeling Methodology Based on Niches and Granularity," C. Parent, K.D. Schewe, V.C. Storey, B. Thalheim (eds.), *Conceptual Modeling –ER 2007*. LNCS 4801, Springer, Heidelberg, 2007: 338-358.

There is an ambition to find a Generic Schema for all CRFs. It is independent from the domain of observation.

7 Submission process/Release Process

In the following the submission process will be specified. The process as given here can be extended. It is necessary to identify the roles of users in the systems. Initially, the following roles were identified::

- End users (clinicians, in the case of ObTiMA, all users are subscribers)
- Ontology experts (Maintenance team), can be remote: the financing is still an open issue
- Administration (Maintaining information and documentation items, homepage, communication with users, etc.)

Submission process

Submission is sent (contains a) term or relation, b) critique of term c) critique of relation),

for new term: name, definition, context of use, justify need, example

for new relation: domain, range, name, definition, context of use, justify need, example

re-submission: is a reference to a previous submission needed or useful? Is a submission ID per root node, date and submitter sufficient?

Receipt of submission (automated)

Goes into Submission Pool

Priority list sorted submission date (oldest first)

Alert to coordinator

Subject definition by coordinator (one or more submission)

Directly accepted or

Reformulation

Redundant submission (standard phrases to submitter, give path)

Request for clarification (standard letter to submitter: does re-submission occur?))

Request for domain expert opinion

Receive clarifications (from both submitter or domain expert) request status

Implementation proposal (directly in workspace, or view) (Versioning system SVN Tortoise) Question is if subject definition and implementation proposal is distinct.

Validated by coordinator (message to submitter (standardized letters, including implementation information) or back to Implementation proposal or back to Subject Definition, or Submitter (standardized letter).

Implementation in workspace

Process end.

Release process (follows submission process)

Release definition: Freeze workspace (can be done for partitions seperately), for each frozen partitions of the workspace a new workspace may be opened, or better after global quality control (Stop submission processing, or change to local implementation definition)

Global quality control
> global adjustments,
> deferred decision by sending submissions
> revision of implementation decision (note to submitter)

(Bulk submissions might take place here)

Release decision (publication event)

Automated release note to all submitters.

Process end.

Status list (must be given to submitter):
Submitted
Subject definition
Implementation
Wait for release (changes may still occur)
Released

Implementation view

Is needed to report implementation of submission to submitter

Find root node of change (domain or subclass)

Show all immediate relations

Questions: Can we define a root node for OWL rule definitions?

If not we need a rule view.

generate Root Nodes (all classes, all symmetric rules)

assign persistent ID system

change submissions refer to ID system

XML database maintain submission history based on ID system

Special cases:

Rename: change ID

Delete: attach history to super-class

Merge/Split: branching history

After implementation we have to check the differences.

Assign implementation views to submissions.

Diff mechanism has to check for which IDs the view was changed.

Special cases cannot be recognized automatically. We can detect the disappearance of an ID, though. Split cannot be recognized AT ALL if one name is preserved! We ignore this case in the first run, since there is a nominal successor.

Graphical report of the consequences/history of each submission.

All change requests (change of node, change of relation) are traced back to root nodes.

Maybe we will need overall comments/requests. (changes not traceable to root node).

8 State of the Art in Submission Systems to Biomedical Terminologies and Ontologies

The review used the list of biomedical OTD given in Deliverable 7.1.

UMLS	Submission by Email. A list with questions that must be answered when submitting material is provided. Some of the criteria for a successful submission are given.
SNOMED CT	Submission by proposal to the „National release center". Neither criteria for the proposal nor a detailed account of the submission method is given.
GALEN	Users registered in the forum may point out mistakes and propose new terms.
LOINC	Submission by email. A list of the criteria for a successful submission is provided. A manual with examples for submissions in different file formats is available.
MeSH	A form with criteria for a successful submission is provided.
ICNP	Submission - Questionnaire on the Homepage. Filled out forms can be submitted by mail or e-mail.
GO	Registered members can use the „Curator Request tracker" to submit. Unregistered Users may send e-mail containing proposals to the GO Helpdesk.
GOA	One form for the submission of new terms, proposal of improvements and corrections exists, but no special submission form.
IntAct	Successful submission is possible after the publication of the proposal in a peer-review journal. Submissions can be sent to IntAct by e-mail before publication. Formal criteria to be observed are given.
dbSNP	Submission process in up to four steps: 1. Application for a NCI-abbreviation (‚Handle') for the submitting lab. 2. Submission files are sent by e-mail. A guide with formal requirements is provided. 3. Submission report is sent to the submitting lab after review of the proposals.

4. Improved proposals, which weren't accepted in the first run, can be submitted again

MedDRA   Submission is possible as one point of general feedback, but no special submission tool exists.

The following OTDs do not provide submission tools (at least no submission process can be found on their web pages.):

FMA   No Submission-tool. It is possible to send e-mails to the curator.

NCI Thesaurus   No Submission-tool.

ICD   There's a survey for users, but no submission tool.

ICF   see ICD.

FDA   Users (registered producers) are welcome to report errors.

OMIM   No submission-tool.

UNIPROT   No submission-tool.

SCOP   No submission-tool.

REACTOME   No submission-tool.

KEGG   No submission-tool.

HGP   No submission-tool.

rRNA   Page is not updated anymore.

JSNP   No submission-tool.