The ObTiMA System – Ontology-based Managing of Clinical Trials

**Holger Stenzhorn[a], Gabriele Weiler[b], Mathias Brochhausen[c], Fatima Schera[b],
Vangelis Kritsotakis[d], Manolis Tsiknakis[d], Stephan Kiefer[b], Norbert Graf[d]**

[a] *Department of Paediatric Oncology and Haematology, Saarland University Hospital, Homburg, Germany*
[b] *Fraunhofer Institute for Biomedical Engineering (IBMT), Sankt Ingbert, Germany*
[c] *Institute for Formal Ontology and Medical Information Science (IFOMIS), Saarland University, Saarbrücken, Germany*
[d] *Foundation for Research and Technology Hellas (FORTH), Institute of Computer Science, Heraklion, Crete, Greece*

Abstract

Clinical Trial Management Systems promise to help research-ers in managing the large amounts of data occurring in clini-cal trials. In such systems Case Report Forms for capturing all patient data can usually be defined freely for a given trial. But if database definitions are automatically derived from such trial-specific definitions then the collected data cannot be easily compared to or integrated into other trials. We ad-dress this interoperability issue with an approach based on ontology and semantic data mediation. This resulted in the development of the ObTiMA system which is composed of a component for setting-up clinical trials and another for han-dling patient data during trials. Both components offer data reusability by relying on shared concepts defined in an ontol-ogy covering the whole cancer care and research spectrum.

Keywords:

Clinical trial management system, Application ontology, Se-mantic data mediation.

Introduction

Clinical Trial Management Systems (CTMS) promise to help researchers in hospitals and biotechnology/pharmaceutical companies to better manage the tremendous amounts of data involved when conducting clinical trials [1]. Their goal is to simplify and streamline the various aspects of clinical trials, such as planning, preparation, performance, and reporting, by providing functionalities, like automatic deadline tracking for legal or regulatory approval, progress report issuing, keeping participant information up-to-date, or import/export data from/into other clinical information systems. For example, it is still a common yet tedious and error-prone practice to collect data at each trial site on paper-based Case Report Forms (CRF) and then to enter them manually into the trial database at the trial center. CTMSs are supposed to avoid this by pro-viding user interfaces that blend into clinical work settings and shield users from underlying data and system complexity.

But as standardized, commercial CTMSs are not yet widely deployed, trial databases and their entry interfaces are often developed in-house specifically for a given trial and therefore not readily reusable in other trials. This issue causes an addi-tional reimplementation burden and makes it difficult to com-pare or integrate data between different trials. But even if CTMSs are used, the following issue remains unresolved: Those systems allow a user to freely define the CRF items and structures without the need of any informatics skills. But al-though this is very desirable, it can create the same interopera-bility problems. If a database is derived from the trial-specific CRF definitions, the database in turn is again also trial-specific and data reuse in further research stays problematic. Thus, our work focuses on solving this interoperability issue through an approach based on ontology and semantic (data) mediation.

This work has resulted in the development of the Ontology-based Trial Management Application, or *ObTiMA* for short. Its development started as part of the European Union project *ACGT* (Advancing Clinico-Genomic Trials on Cancer) aiming at creating an open, semantic and grid-based technology infra-structure to support clinicians and scientists in post-genomic clinical trials in cancer research [2]. Its development continues now in the European Union project ContraCancrum intended to develop a platform for simulating tumor development and response to therapeutic modalities to optimize the disease treatment procedure in a patient's individualized context [3].

An early ObTiMA prototype was presented in [4]. Here we describe the current state of its development. To this end, we first cover the ontology and semantic mediation as the basis of the data management in ObTiMA. Then we describe the two main components of the system: the Trial Builder for design-ing clinical trials and the Patient Data Management System for handling patient data within a trial. We conclude by referring to related research and how ObTiMA is now being evaluated.

Relevant ACGT Components

The advent of innovate technologies, like high-throughput screening or pharmacogenomics, has lead to the creation of new data on a previously unknown scale. However, tools to automatically analyze these data are still missing or not yet in an applicable stage. This is the core issue ACGT wants to

solve by developing a unified technology infrastructure to facilitate seamless, secure access to clinical and genomic data: High-performance knowledge discovery techniques are being created to support multi-centric, post-genomic clinical trials.

Among the various ACGT components, the two most relevant ones for ObTiMA are the *Master Ontology*, providing a unified set of (logically defined) domain concepts necessary to describe all aspects of clinical trials, and the *Semantic Mediator* providing ontology-based mediation between data sources.

Master Ontology

The task of this ontology is to comprehensively represent the domain of research on cancer and its clinical management and care, with special emphasis on mammary carcinoma (breast cancer), nephroblastoma (Wilms' tumor) and rhabdoid tumor. Due to the multiplicity of entities and processes present in this domain, the ontology contains elements ranging from genetics, the medical and administrative field, up to the legal domain. Therefore it forms a cross-section between all of these subdomains with each one being a vital part of the overall domain.

Basically, domain ontologies represent a given domain by formally and univocally defining the types and their connecting relations, as used within that domain. Hence, the Master Ontology could effectively be seen not as "proper" domain but rather as application ontology since it is tailored towards the functionality requirements of the application services of ACGT. For example, "proper" domain ontologies exhibit a clear-cut, distinguishable domain that can be found in basic scientific disciplines, like anatomy or cytology. But since our ontology incorporates many different aspects, it is not possible to clearly delineate such a single, specific domain. However, in using ontologies the differentiation between domain ontologies and application ontologies blur considerably, anyway.

An ontology is thus a representation of the referents, in the linguistic sense of the word, in some domain or for some specific application providing references for the terms used in describing the domain. All naming and labelling in the Master Ontology was checked against actual term usage in the domain to ensure end-user usability [5] (which also contains examples from the ontology missing here due to space restrictions).

For the ontology development, the principles for state-of-the-art ontology engineering, as proclaimed by the OBO Foundry [6] have been followed closely. Those "best praxis" criteria for ontology design support in producing an artifact which is both sensible from the point-of-view of content as well as coherent regarding its logical, internal structure. As an example, all concepts are defined by formal subsumption ("is-a" relation) and logical constraints based on the relations between concepts. Hence concepts are clearly delineated from each other and ambiguity of pure natural language definitions is avoided.

Semantic Mediator

The process of semantic (data) mediation involves the matching, combination, and retrieval of data stored at disparate data sources to offer a unified view over them. It is semantic since the relationships between all parts of the data are made explicit by using ontological concepts and relations. The Semantic Mediator has initially been designed to fulfil this task as a part of the ACGT platform. It accepts queries to and retrieves results from disparate data sources, like databases or flat files, based on Master Ontology concepts and relations. Hence, users who want to retrieve data via the mediator do not have to know any technical details of the underlying data sources but employ ontological concepts and relations and combine them into queries to access all data sources in a unified way [2].

A new data source is integrated into the semantic mediator by creating a set of rules for mapping this data source onto the Master Ontology. For databases, this task is simplified by a graphical tool that assists in mapping from database tables and columns onto appropriate ontological concepts and relations. Still this process remains a complex task that needs to be performed by users who are experts both in the domain, the database, and the ontology. They must be able to realize the subtle differences between similar ontological concepts and how this is mirrored in the data sources (and vice-versa) [7]. On the upside, this mapping needs to be created only once when a data source is first added but can be reused for other tasks, so a genomic data source can be directly reused in other trials.

ObTiMA System Components

Trial Builder

The Trial Builder represents one of ObTiMA's two main components (cf. Fig. 1) and enables the user to specify the various aspects of a clinical trial. The trial outline and metadata can be defined in a master protocol based on templates for describing the trial goals and its administrative data, like start or end date. Treatment plans can be graphically designed to guide clinicians through the treatment of individual patients and particular treatment events, such as chemotherapy or surgery, can be defined with all necessary information. The particular order of treatments for individual patients can be defined by placing them on a timeline. Also, treatment stratifications and randomizations to be applied for a patient can be described. For each stage on the treatment plan a CRF can be assigned to collect the data documenting the treatment.
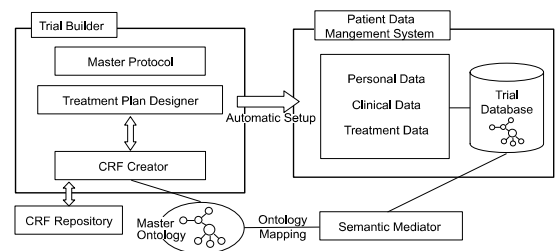

Figure 1 – ObTiMA System Components

Ontology-based CRF Creation

The creation of CRFs marks the core functionality of the Trial Builder. In a graphical user interface, the user can define the content, layout, and navigation of the CRFs which are used to

capture all patient data during a clinical trial, like the patient's history, medical findings, diagnostic data, or genomics data.

It is important that all information can be defined here which are necessary for the data integration, i.e., each CRF item is described based on ontology concepts together with metadata, like data type and measurement unit, to set-up the trial database. However, the internal CRF (data) representation is not the focus of clinicians but their "user interface" (layout) and their adaption and integration into the specific workflow of the planned trial: clinicians are not to be bothered with the underlying aspects of the trial database or the ontological metadata. Thus all these aspects are made transparent to the user through a graphical user interface which hides the actual complexity yet gathers all required information for automatically creating the trial database. This interface is derived also automatically from the content and structure of the Master Ontology but presents a simplified ontology view, adapted to the task of creating items (Figure 2). It comprises the following sections:

In the *Ontology View* (1), the user selects concepts from the ontology to create a CRF item. Here, the interface tries to overcome the gap between clinical practice and the actual logical representation of ontology concepts: Although the ontology provides natural language descriptions for its concepts/relationships (in addition to the logical definitions), those often do not fully mirror the needs of practical or clinical perception of reality. In order to meet this need, we do not present the full Master Ontology here but rather a simplified clinical view which contains a trial-independent basic classification of CRF contents from a clinician's point of view.

It is by intention that the clinical view is far less detailed as the actual Master Ontology and since this allows the possibility to provide a much easier entry point for the user. The interface of the clinical view is implemented as a tree always that starts at node of the concept "Patient" as focus of any clinical study (and hence CRF) and only presents those concepts that are directly reachable from this concept, like "Weight" or "Tumor" (indicating a patient's tumor). Only when a concept is selected then also the concepts directly reachable from this one are shown, such as "Laterality" in the case "Tumor" was initially chosen (indicating the laterality of the patient's tumor).
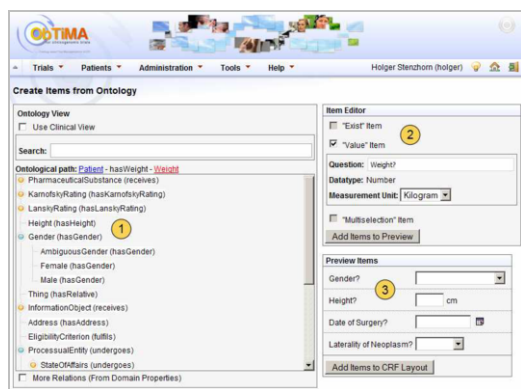

Figure 2 – Ontology Viewer while Creating CRF Items

When a concept is chosen in (1) then a corresponding item is automatically created and shown in the *Item Editor* (2) together with its attributes determined automatically based on the chosen concept, such as label, data type, or answer possibilities, and which can be manually adopted. For example, the concept "Weight" has a numerical data type and a list of suitable measurement units attached. So, when the CRF with this item is used in a clinical trial then the measurement units are offered as selection possibilities (in a drop down menu). The specified value (entered into a text field) is automatically tested to be of numerical type and also to be non-negative (since a weight cannot be negative). Finally, *Preview Items* (3) presents all created items in the order in which they are intended to appear on the CRF. Single items can be reordered by simple drag and drop and subsequently transferred to the interface where the overall layout of the CRF is then defined in turn.

CRF Repository

Revisiting the reuse and interoperability issue discussed in the introduction, in many trials similar or equal data are collected, yet stored differently because of different data(base) definitions. Applying the Master Ontology already improves this situation through using standardized concepts when creating CRFs. Going a step further, the situation would be further improved by partial or complete reuse of existing CRF in case similar data is collected. This idea realized by creating a unified CRF Repository as crucial part of ObTiMA. This repository allows the storage and retrieval of entire ontology-based CRFs and single CRF items or components for reuse and adaption in subsequent trials: When setting-up a clinical trial, fitting CRFs can either be directly reused or new ones quickly created by "plugging together" existing CRF items and components. This in turn fosters the standardization of CRFs even more, since CRFs can now be compared not only on the level of single items (through their basis on ontological concepts) but also on the level of larger components or in their entirety.

Patient Data Management System (PDMS)

The PDMS supports clinicians when conducting a clinical trial and is automatically set-up based on the master protocol and CRFs defined in the Trial Builder. The PDMS guides the clinicians through the actual treatment of patients according to their individual treatment plans and provides a graphical user interface to fill in the CRFs relevant to the patient's current treatment situation. The interface is adjusted to everyday clinical needs: As with the Trial Builder, the complexity of the underlying ontology is hidden from the user, yet its logic-based concept definitions are used to provide direct validity checking when CRF are filled in. The basic look of the data entry interface corresponds to section (3) on Fig 1 with each input element providing on-the-fly feedback about its current state based on the just mentioned checking, i.e., in case a negative value is specified for a weight then this error is immediately highlighted along with an explanation of the error.

Data Export

To integrate ObTiMA into real-world clinical settings, the system must be capable to interface with other existing CTMS

and be able to exchange data in a format they understand. To meet this requirement, ObTiMA allows to import and export trial metadata, CRF descriptions and patient data through an extended version of the CDISC Operational Data Model (ODM) format [8]. This platform-independent, quasi-standard for exchanging and archiving clinical trial data is supported by many current CTMSs. Observing CDISC's extension guidelines, we enriched this format by allowing the additional inclusion of (metadata) descriptions based on Master Ontology concepts. In the case other CTMSs want to import data generated by ObTiMA, they can chose to interpret the supplemental descriptions but if this is not feasible the resulting data is still "ODM complete" and can sensibly be used by those systems.

Administration, Security and Pseudonymization

To administer multicentric clinical trials, ObTiMA contains several advanced facilities for managing the multitude of institutions, researchers, and patients usually participating in such trials. An elaborated, fine-grained security architecture has been implemented to handle the rights and roles that can be attached to the system's users in order to guarantee that they can only perform the tasks which they are fully authorized for. It is also straightforward to dynamically react to changes within a running clinical trial, since new institutions and users can always be added or extra security roles and rights be defined.

It is also indispensable that ObTiMA, as a system holding real patient data, securely stores all of the data which could possibly identify some patient to non-authorized persons in pseudonymized and encrypted form. To foster security even more, such personal data is physically separated from the actual clinical research data through the use of two distinct database servers: One server holds the database for storing the personal data of the patients, such as their names and addresses (which must never be shared, e.g., via the Semantic Mediator). The protection of this database strictly follows all current legal regulations for data protection in clinical environments. The other server hosts the database that contains the actual research data collected in a clinical trial (through the use of the CRFs). It is possible within the Trial Builder to mark certain CRF items as personal which results in this data being stored in the database for personal data and not in the one for research data.

Advantages of Using Ontology and Semantic Mediator

Built-in Semantic Trial Interoperability

As pointed out before, when clinical trials are designed with the Trial Builder (and thus linked to the Master Ontology) then this means that all items defined on the CRFs are also attached to the corresponding ontological concepts. Therefore, when data is entered for some CRF item then this data is also directly linked to the ontology based on the item's attached concept. No manual and error-prone data annotation happening subsequently and using biomedical terminologies is necessary. The advantage can be easily seen when looking at recent studies highlighting that the accuracy of SNOMED annotations exceeds 50% only slightly for three different scenarios [9] and hence annotated data cannot be reliably compared at all.

The automatic link of the collected data to the ontology makes it further simple to "publish" collected data or "blend in" ex-

ternal data into a current clinical trial via the Semantic Mediator: As ObTiMA automatically generates the mapping rules needed by the mediator from the concept-based item definitions, its (research) database can be readily added as data source to the mediator. Other trials running on ObTiMA (and therefore also based on the Master Ontology) or other applications based on that ontology, can perform concept-based queries on the trial data using the Semantic Mediator. The opposite direction to integrate other, external data sources into the current clinical trial is also made possible: for biomedical or biomolecular data sources containing, e.g., genomic data or data collected in related clinical trials (but not using ObTiMA) a mapping based on the Master Ontology can be created for the mediator (see above). Then by using the same concepts (combined with using the same interface) to query external data and data collected in the current clinical trial, it becomes straightforward to perform cross-trial meta-analyses.

Increased Data Quality

Continuing the above, by basing the data collection on the shared Master Ontology which has been developed by clinical domain experts in cooperation with ontology experts, the data becomes consistent to the knowledge of the underlying domain and hence its quality increases. The Trial Builder in ObTiMA ensures, mostly transparently to the user, that during the creation of CRF items only concepts from the ontology are chosen and logical restrictions attached to the concepts, like domain and range restrictions, are satisfied. However, currently not all of the restrictions encoded in the ontology, such as number restrictions, can be guaranteed automatically. Therefore we are currently investigating novel algorithms to support the user in further improving the data quality and consistency [10]. As with the ontology integration, those algorithms will be applied "below the surface", in order to support the user and improve quality but without exposing their intrinsic complexity.

Conclusion

In this paper we have described ObTiMA, the Ontology-based Trial Management Application and presented the details of its two main components to design clinical trials and to manage the patient data within them, as well as the Master Ontology and Semantic Mediator as the foundation of the system.

Related Work

We are aware of other research initiatives aiming to achieve data integration by utilizing ontologies in the clinical trials. Ontology based data integration frameworks, such as the Epoch project [11], are a very active research field. The difference between the latter and ObTiMA lies in the fact that these frameworks focus on only integrating existing data sources (instead of creating CRFs to gather new shareable data). Other initiatives focus on applications for creating standardized CRFs by using ontologies, like TermTrial [12], which also allows automatic database creation based on those CRFs. ObTiMA combines the advantages of the two research strains enabling both a user-friendly ontology integration during the trial design and the automatic set-up of the PDMS as well as the seamless integration of external data and the "publication"

of its own database. It is also attractive that all ontology and semantic mediation integration is transparent to the user and so any additional complexity that might be introduced by those technologies/methodologies is hidden, but still their application provides a strong support and added value to the user.

Evaluation

ObTiMA is currently being evaluated within the SIOP 2001/GPOH study [13] at the Saarland University Hospital as that study's trial center. The evaluation is performed by two trial administration experts who, until now, have been using a Microsoft Access database created in-house along with self-defined templates to enter the data collected on paper-based CRFs. Both ObTiMA's trial design and execution facilities are evaluated, accompanied by a user study where users are asked to report their experiences and propose improvements. In the first step, the CRFs have been designed together with the Master Ontology developers. For this, the existing paper-based study CRFs were taken and "translated" into their electronic, ontology-based pendant. The Trial Builder's user interface was found to be mostly straightforward and quick to understand after a short tutorial. Yet some possible improvements, e.g. regarding performance and user guidance, were identified and are now being worked on. Also some minor concepts were initially missing from the ontology necessary to fully model all details of the original CRFs but which have been added since. In a second step, ObTiMA's PDMS is evaluated by entering a large collection of CRFs. Feedback is also very encouraging here since it was expressed that the given functionality and its interface suit the clinical audience and are easy and quick to use. Still, a scientifically valid evaluation on a larger user basis is needed and therefore an evaluation including several clinical centers (based on a Rhabdoid tumor study) is now on its way. (The feasibility of ontology-based data integration via the Semantic Mediator has been successfully proven in ACGT [7].)

Acknowledgements

References

[1] Choi B, Drozdetski S, Hackett M, Lu C, Rottenberg C, Yu L et al. Usability comparison of three clinical trial management systems, AMIA Annu Symp Proc. 2005:921.

[2] Tsiknakis M, et al. A semantic grid infrastructure enabling integrated access and analysis of multilevel biomedical data in support of post-genomic clinical trials on cancer. IEEE Trans Inf Technol Biomed. 2008 Mar;12(2):205-17.

[3] Marias K, Sakkalis V, Roniotis A, et al. Clinically oriented translational cancer multilevel modeling: the ContraCancrum project. IFMBE Proc. Jan 2010;25(4):2124-7

[4] Weiler G, Brochhausen M, Graf N, et al. Ontology based data management systems for post-genomic clinical trials within a European grid infrastructure for cancer research. Conf Proc IEEE Eng Med Biol Soc. 2007;2007:6435-8.

[5] Grigonyte G, Brochhausen M, et al. Evaluating ontologies with NLP-based terminologies - a case study on ACGT and Its Master Ontology. Proc Int Conf FOIS. 2010;in press.

[6] Smith B, Ashburner M, Rosse C, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007 Nov;25(11):1251-5.

[7] Kalfoglou Y, Schorlemmer Y. Ontology mapping: the state of the art. Knowl Eng Rev. 2003 Jan;18(1):1-31.

[8] Kuchinke W, et al. Extended cooperation in clinical studies through exchange of CDISC metadata between different study software solutions. Meth Inf Med. 2006;45(4):441-6.

[9] Andrews J, Richesson R, et al. Variation of SNOMED CT coding of clinical research concepts among coding experts. J Am Med Inform Assoc. 2007 Jul-Aug;14(4):497-506.

[10] Weiler G, Poetzsch-Heffter A, Kiefer S. Consistency checking for workflows with an ontology-based data perspective. Lect Not Comp Sci. 2009 Aug;5690:98-113.

[11] Shankar R, Martins S, O'Connor M, et al. An ontology-based architecture for integration of clinical trials management applications. AMIA Ann Symp Proc. 2007:661-5.

[12] Merzweiler A, et al. TERMTrial - terminology-based documentation systems for cooperative clinical trials. Comput Methods Programs Biomed. 2005 Apr;78(1):11-24.

[13] Graf N, Hoppe A, Georgiadi E, Belleman R, et al. 'In silico' oncology for clinical decision making in the context of nephroblastoma. Klin Padiatr. 2009 May-Jun;221(3):141-9.

Address for Correspondence

Holger Stenzhorn, Department of Paediatric Oncology and Haematology, Saarland University Hospital, Building 9, Kirrberger Straße, 66424 Homburg/Saar, Germany