Ontoverse: Collaborative Knowledge Management in the Life Sciences Network

Ingo Paulsen[1], Dominic Mainz[1], Katrin Weller[1], Indra Mainz[1], Jochen Kohl[1] and Arndt von Haeseler[2]

[1] Heinrich-Heine-University, Universitaetsstr. 1, 40225, Duesseldorf, Germany
[2] Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna, Medical University of Vienna and Veterinary University of Vienna, Dr. Bohr-Gasse 9, 1030, Vienna, Austria
email: {paulseni, mainzd, weller, kohlj}@uni-duesseldorf.de
email: mainzi@biophys.uni-duesseldorf.de
email: arndt.von.haeseler@univie.ac.at
phone: (+49 0211) 81 13716, *fax:* (+49 0211) 81 15767

Abstract

This paper regards the two aspects of knowledge networking: data networks for information integration and social networks for information sharing in communities. The importance of ontologies as a means for effective information integration is discussed and related to the current Web 2.0 trends. The Ontoverse ontology wiki is introduced as a tool for collaborative ontology engineering and knowledge management with particular focus on interlinking the research community within the life sciences.

1 Introduction

Knowledge networking comprises two different aspects: collaborative knowledge management in communities (human networks) and effective information integration (data networks). Thus, on the one hand we regard techniques that help to structure and interlink existing knowledge sources effectively with ontologies as the core technique. And on the other hand, we speak of knowledge networking in terms of people who share their knowledge via social networks.

The Ontoverse project aims at combining these two points of view: we are establishing a platform, that provides tools for designing ontologies for annotating and interlinking knowledge sources and that also helps people to build up social (scientific) networks. This platform will be called the Ontoverse ontology wiki. It comprises support for collaborative ontology engineering, an ontology based publication management system and solutions for knowledge exchange in scientific communities.

This paper will be divided into two major sections: At first we will discuss the importance of ontologies for knowledge representation and retrieval in scientific contexts with emphasis on the life science domain. Then we will introduce the Ontoverse approach for collaborative ontology development and describe our experiences in providing supportive tools for this purpose.

2 Ontologies for Shared Knowledge Representation

Ontologies in terms of knowledge representation are systems of concepts, defining sets of individuals and their relations within a defined domain of knowledge. They can unambiguously define domain semantics making use of the rules and axioms supported by ontology languages such as OWL[1]. This also enables computers to reason about the depicted knowledge.

Ontologies can thus be used for explicitly describing the semantics of information sources [7, 14] and knowledge itself. More than any previous type of concept system, ontologies allow semantic annotation of documents to put them into wider knowledge contexts [5]. They address the problem of information integration: information should not only be found (information retrieval), but correlations should be pointed out. Every new piece of information has to be interrelated to already existing knowledge structures.

2.1 Information Integration for Scientific Data

Recently, the optimization of storing, retrieving and integrating data is becoming a popular focus for the World Wide Web in general and a fundamental task for scientific contexts. For our focused domain of interest, the life sciences, the particular problem is the integration of heterogeneous data. For example, bioinformatics data not only consist of customary textual items (scientific publications), but also comprises nucleotide sequences, amino acid sequences, 3D structure of molecules and a manifold of other experimental results. Such diverse biodata need to be stored and structured effectively. Recent progress in the life sciences has already led to the accumulation of gigantic amounts of biodata that now demand classification, accessibility and visualization [13].

Additionally to the management of the vast amount of biodata there is also an urgency to collect scientific cognitions and make them multidisciplinarily accessible. The Gene Ontology [4] for example structures investigated genes and information about them. This addresses a main problem in genetics which is the multiple denotation of similar genes which were found in different organisms. This problem leads back to the rapid increase of knowledge. Scientists often specialize on single research domains, for most of them it is hard to keep track of all new developments even in these limited areas, even harder it is to recognize trends in other fields that might effect one's own research. That is why a growing interest in ontologies can be noticed e. g. in the bioscience community, resulting in different ontology projects such as the already mentioned Gene Ontology and the National Center for Biomedical Ontology [12] which includes the Open Biomedical Ontologies[2]. Such collaborative approaches offer the possibility to manage the increasing knowledge and data.

[1] http://www.w3.org/TR/owl-features/
[2] http://obo.sourceforge.net/

2.2 Knowledge Sharing with Ontologies

Within non-scientific contexts, we currently observe the development, that user communities are indexing heterogeneous online content, such as pictures, publications or collections of bookmarks on the WWW. This process is referred to as social tagging and is one major aspect within the vision of a Web 2.0. The totality of unstructured tags is called folksonomy. Tags are used as content-descriptive metadata for improving retrieval facilities. However this unstructured tagging performs less effective compared to traditional techniques in knowledge representation [9, 11].

With folksonomies, user communities have made a first step to widespread content-indexing within the WWW, awareness for the use of metadata arises. This now has to be lead on to combine the social mechanisms with semantic annotations using highly-structured ontologies. What is needed, is appropriate support for user communities to collaboratively collect knowledge and represent it as shared ontologies. An ontology wiki (as described in the next section) will facilitate collaboration and ensure that these ontologies do indeed represent a shared view, as opinions and suggestions of a broad community of domain experts can be regarded.

Shared knowledge representations like ontologies are of enormous value, not only for information storage and retrieval purposes, but also as a basis for scientific discussions. The following chapter will describe which support Ontoverse gives to scientific communities in collecting and structuring knowledge.

3 Ontology Management in Scientific Communities

The Ontoverse approach is to implement an ontology wiki; i. e. an editor platform that supports distributed work on structural (ontological) data as well as informal discussions and annotations (proto-ontological data). This platform will be open to interested users. They can view and use ontologies, join existing ontology projects or plan and start a project anew. All different phases of ontology development like conceptualization, editing, maintenance and reuse will be supported (see [3] for a survey of published ontology design methodologies). This also connotes, that our wiki will become a platform for multiple different ontology projects that have to be administered diligently and provided in an easily accessible way.

3.1 User Community and Collaboration

Some approaches have been made to support collaborative ontology engineering e. g. in [1, 6]. What is new in Ontoverse, is the explicit support of a social (scientific) network closely combined with a web-based ontology editor. A focus will be placed on the support of a heterogeneous community. Potential users will differ in their fields of interest and skills: On the one hand knowledge and expertise is needed from domain experts (DEs). On the other hand ontology languages can only be fully exploited by knowledge engineers — in the following

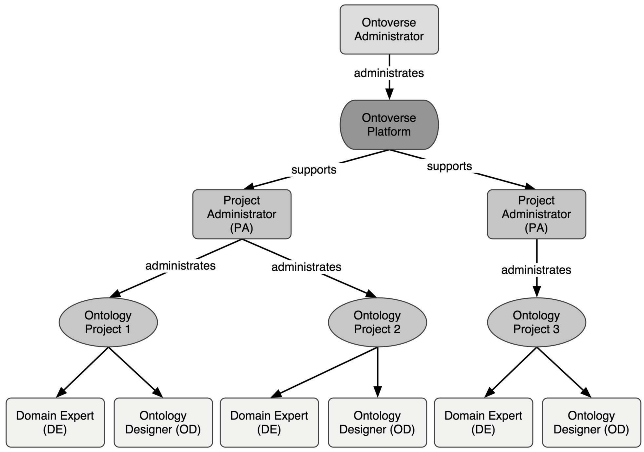

Figure 1: Every ontology project on the Ontoverse platform is administrated by a single group of project administrators composed of one or more members. Customarily, both domain experts and ontology designers are involved in an ontology development process.

called ontology designers (ODs). DEs are often unfamiliar with techniques of knowledge formalization, but contribute valuable data if they are provided with easy to use systems for knowledge input and additional help from ODs.

We therefore distinguish the roles of DEs and ODs within the platform. Main differences lie in the user interface and different rights to edit data. For example, users without knowledge in ontology engineering can log in as DE and will be provided with a less complex user interface that does not include options for editing the actual ontological data — although they might be viewed and commented. Principally, every user will be able to choose whether to act the role of an OD and/or DE. Both roles are modifiable and are always coupled to single projects. Every ontology project will also have at least one user taking over the position of a project administrator (PA), who is responsible for final decisions and for coordinating discussions of other project members (see Fig. 1). Functions of the different user groups can be roughly summed up as follows:

Domain Experts

- Collect knowledge, provide it informally (proto-ontological data) and discuss it.
- Collect publications relevant to actual topics of interest and tag them.
- Answer domain-specific questions posed by ODs.
- Reflect and benchmark existing ontologies.
- Populate ontologies and continuously contribute new scientific findings.

Ontology Designers

- Exploit the input contributed by DEs and transfer it into formal ontology languages (ontological data).
- Discuss formal knowledge representations, edit and maintain them.

Project Administrators

- Define aim and domain of a new ontology.
- Coordinate discussions and monitor the collaboration process.
- Release versions of an ontology.

Considering the user roles and potential workflows as well as general problems in collaborative systems, the following requirements for the platform can be derived:

- Elaborated communication channels, adjusted for discussions between DE-DE, OD-OD, DE-OD, PA-DE and PA-OD.
- User profiles and networks to show fields of expertise and search for experts.
- Collaboration guidelines and principles of ontology engineering.
- Community awareness features, tracing of changes.
- Information on authorships (using signatures from an certified authority), copyright guidelines.

3.2 Knowledge Management in the Ontoverse Wiki

The typical workflow on this platform can be described as follows: A scientist needs an ontology of a specific domain. She checks the Ontoverse platform for existing ontology projects. If one of them already fits her needs, she may join the project as OD and/or DE and either actively contribute to the project or passively view and use the collected data[3]. If none of the projects was suitable, she initiates her own ontology project from scratch, becoming its PA automatically. This is the first step of the multilevel ontology engineering process.

One major focus of the Ontoverse platform is the support of all phases in the ontology life-cycle process as illustrated in Figure 2. In collaborative ontology engineering, the conceptualization phase as first step of every newly started ontology project is particularly important. During this phase the scope of an ontology is clarified, as well as its intended goal and a target group that should use it. This phase is complemented by an ontology requirement specification document (ORSD; modified from [15]) where fundamental determinations should be documented precisely. An ORSD is a guideline for the whole following engineering process, therefore a PA may prohibit further editing and changing if she thinks that a usable state has been reached.

Furthermore thematic discussions and the collection of proto-ontological data in terms of knowledge acquisition are important steps within the early ontology

[3] An alternative scenario might be, that a user is invited to join an existing project by project members or administrators.

development process. All planning and conceptualization steps are hardly considered in current ontology editors, they usually start right with the formal editing of ontologies. In Ontoverse ways for entering non-formalized knowledge are particularly necessary to support the DEs. We provide a wiki, where DEs can enter their domain knowledge in different ways. These proto-ontological data may be textual, e. g. in form of glossaries, single concepts enriched with explanations or definitions, unstructured collections of concepts or other notes and references to external sources. Like in conventional wikis, documents can be uploaded and integrated; in the first instance that will be pictures, e. g. schematic depictions of a subject. Throughout the whole wiki, changes can be traced and assigned to their authors. All this collected domain knowledge is needed as a basis to set up a formal ontology. Optimal results will be achieved if ODs pose concrete questions regarding the domain directly to the DEs, ideally even express their information needs concretely within the ontology wiki.

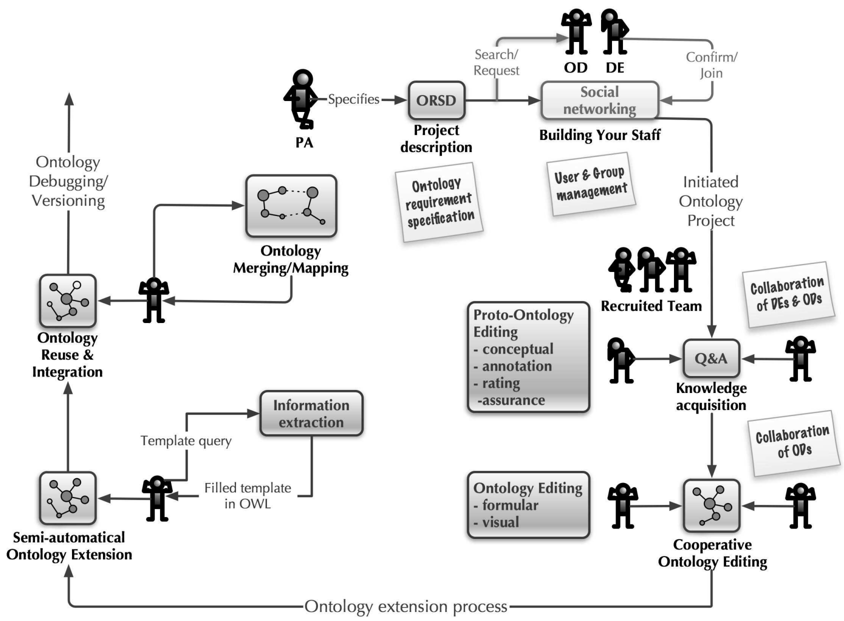

Figure 2: Schematic illustration of the Ontoverse ontology life-cycle process

ODs then build up the actual ontology via another channel: the web-based, multi-user ontology editor. The high-level support of formal ontology engineering actually constitutes a major part of the Ontoverse project, but should not be discussed very detailed in this paper (see [8] for more information). Main problems to be regarded in the design of the editor are:

- *Access conflict resolution.* During the commission of modifications, conflicts with other users, who are currently editing the ontology, may arise. While a manual merging process is quite functional working with textual data, it is insufficient for semantic rich languages with extensive interconnections. Merging of semantic data has to be extended by methods to identify the consequences of alterations.
- *Ambiguity disintegration and debugging processes.* In order to permit alternative formalizations of domain knowledge in an ontology project until a consensus has been found, an ambiguity disintegration process has to be supported by the platform. This process ensures a definite ontology state before the inconsistency debugging and quality control steps. To optimally support this procedure we currently develop an ambiguity labeling mechanism and its implementation into the Ontoverse platform. Ambiguity disintegration and the ensuing debugging phase are part of the release process described next.

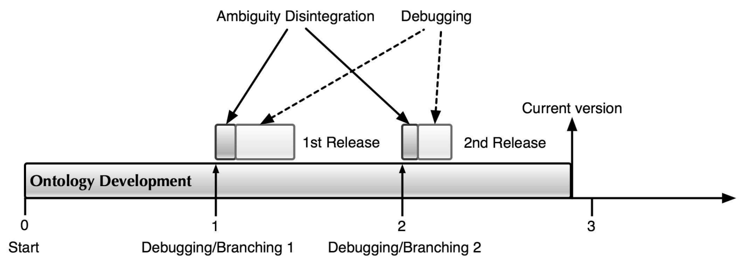

Figure 3: Ontology debugging and release process. After certain time intervals an ontology should be submitted to a disambiguity resolution process followed by a debugging stage where inconsistencies are removed. The final stable ontology gets a fixed release number.

- *Validation and version releases.* A project's ontology is subject to successional changes. To enable periodical consistent and stable releases the system incorporates a release process. If the PAs and/or the project community that it is time for a new release the current state of the ontology is copied into a debugging branch. This branch is used for ambiguity disintegration and debugging of the data and has a consistent and identifiable release as outcome. The final stable ontology gets a fixed release number and can be applied directly or stored in an internal archive as a source for potential ontology reusage (see Fig. 3).
- *Elaborated visualization techniques.* These are needed to present complex ontologies and concept relations in human readable form, e. g. with conceptual graphs. This will show the current state of the ontology to DEs, who can then help to evaluate it.

The aspects mentioned so far all regard the *intra-ontological* level of ontology engineering, i. e. the conceptualization and construction of a single ontology. Within Ontoverse as a platform for multiple ontologies, the *inter-ontological* level of ontology engineering also has to be considered. It comprises all actions and considerations that effect more than one ontology, i. e. interaction and interrelation of different ontologies. Note that no universal definition exists on what is exactly *one* ontology; therefore confusions may arise if someone speaks of subontologies belonging to one general ontology and someone else does regard all of them as independent ontologies. Inter-ontological editing includes aligning, mapping and merging of two or more ontologies. Such aspects of reusing and interrelating ontologies are highly relevant in a working environment, where different ontology projects do exist in parallel. Existing ontologies may be reused in new projects or different ontologies have to be aligned in order to enable complex knowledge based workflows in applications.

3.3 Experiences in Developing a BioInformatics Ontology for Tools and Methods

Within our project we are designing an ontology to gather experiences which are incorporated into the development of the platform. For that purpose we have chosen the domain of bioinformatics tools. This BioInformatics Ontology for Tools and Methods (BIO2Me) is needed, because at the present state the plethora of existing bioinformatics programs cannot be overviewed with reasonable effort. Bioinformatics provide tools for the efficient processing of the amount of experimental data (e. g. from genome sequencing), sequence analysis, and structure prediction and visualization (amongst others) and by now there is a variety of tools available which handle similar biological questions with different computational methods. Without a structured overview, even for experts in bioinformatics it is hard to decide which tool fits their individual requirements best. Therefore, the knowledge about these tools will have to be collected and organized, for otherwise the full potential of existing tools cannot be utilized. The goal of BIO2Me is on the one hand to enable search for bioinformatics tools that meet the user's needs and on the other hand to provide information about certain tools and computational methods.

The domain of BIO2Me eminently points out the need for a collaborative ontology engineering. To represent bioinformatics tools with their applications, the whole bioinformatics research field and biology itself have to be displayed adequately in a structured way, as the application range of every tool has to be set into this context. Furthermore, existing bioinformatics tools have to be collected and described, computational methods have to be modelled as well to relate them to the tools in which they are utilized. It is hardly possible to manage this within a small group, experts from different fields are needed, among them the developers of bioinformatics tools, who can characterize them best, as well as their users, who can rate the programs or report their experiences. It will be necessary to constantly keep track of developments in bioinformatics tools, e. g. as new tools or new versions of existing tools will be published. Ideally, the developers would directly report this to the community concerned with the ontology.

The development of this ontology needs an extensive planning and conceptualization phase. We started with capturing the requirements and went on with a collection of relevant knowledge resources. We checked for existing and publicly available ontologies, but have not found reusable ones for our purpose. Therefore the ontology has to be built from scratch. For a start, we have built the core structure of the ontology and started to model the branches concerning our own fields of expertise in more detail. Basing on that structure, a community will be able to extend and refine the ontology. As a next step we started to identify DEs and to invite them to join BIO2Me. They should now discuss the current state and continue the conceptualization and collection of informal knowledge, which will then enable us to go on with formalizing more parts of this ontology.

4 Conclusion

The necessity for ontology engineering, annotating, integrating is uncontested. Ontologies are the core element for an upcoming Semantic Web. Furthermore, the so-called Web 2.0 initiatives aim at interconnecting communities on the web and enabling fruitful collaboration for private and business use as well as for scientific work in various research areas.

Ontoverse will take part in this process by developing a new, internet-based application for collaborative and interdisciplinary ontology building in terms of an ontology wiki. This solution will serve as a basis for cooperative knowledge management and knowledge structuring in form of domain ontologies and will at the same time become a platform for scientific dialog to support institutions and professional users in confidential exchange.

While we have mainly dealt with community structures, communications and knowledge input in this paper, there are more aspects we are working on in the project. To assist the community in populating the ontologies, Ontoverse integrates information extraction (IE) technologies [2, 10] that can propose new concepts and instances extracted from scientific publications. IE screens textual data to fill predefined templates with facts that can be used to extend knowledge bases after being curated by ODs (see Fig. 2). A publication database including domain-specific articles is being built up as a corpus. In return, the newly developed ontologies themselves will help to retrieve relevant documents from the database.

Acknowledgements

Financial support was provided by the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF), Bonn, of the Federal Republic of Germany, through the Projektträger im Deutschen Zentrum für Luft- und Raumfahrt e.V. (DLR), Neue Medien in der Bildung + Fachinformation (PT-NMB+F), Registration Numbers 01C5972 to 01C5978. Arndt von Haeseler is also supported by the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF). The authors would like to thank Gerhard Steger for his support with the design of BIO2Me.

References

1. J. Bao, V. Honavar. Collaborative Ontology Building with Wiki@nt. A Multi-agent Based Ontology Building Environment. *In: Proceedings of the 3rd International Workshop on Evaluation of Ontology-based Tools (EON)*, Hiroshima 2004, S. 1-10.
2. P. Buitelaar, P. Cimiano, S. Racioppa, M. Siegel. Ontology-based information extraction with SOBA. *Proceedings of LREC 2006*, 2006.
3. M. Fernández López. Overview of methodologies for building ontologies. *In Proceedings of the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods (KRR5)*, 1999.
4. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25: 25-29, 2000.
5. S. Handschuh, S. Staab. Annotation for the Semantic Web. *Frontiers in Artificial Intelligence and Applications*, Vol. 96, Amsterdam 2004.
6. M. Hepp, D. Bachlehner, K. Siorpaes. OntoWiki — Community-driven Ontology Engineering and Ontology Usage based on Wikis. *Proceedings of the 2005 International Symposium on Wikis (WikySym)*, San Diego 2005.
7. U. Leser, F. Naumann. Informationsintegration. Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen, Heidelberg 2007.
8. N. Malzahn, S. Weinbrenner, P. Hüsken, J. Ziegler and H. U. Hoppe. Collaborative Ontology Development — Distributed Architecture and Visualization. *In this volume.*
9. A. Mathes. Folksonomies? Cooperative Classification and Communication Through Shared Metadata. Retrieved March 31, 2007 from `http://adammathes.com/academic/computer-mediated-communication/folksonomies.html`, 2004.
10. M.T. Pazienza (ed.). Information Extraction in the Web Era. Springer-Verlag: New York, 2003.
11. E. Quintarelli. Folksonomies: power to the people. Paper presented at the ISKO Italy-UniMIB meeting. Retrieved March 31, 2007 from `http://www.iskoi.org/doc/folksonomies.htm`, 2005.
12. D. L. Rubin, S. E. Lewis, C. J. Mungall, S. Misra, M. Westerfield, M. Ashburner, I. Sim, C. G. Chute, H. Solbrig, M.-A. Storey, B. Smith, J. Day-Richter, N. F. Noy, M. A. Musen. National Center for Biomedical Ontology: advancing biomedicine through structured organization of scientific knowledge. *OMICS*, Vol. 10, No. 2 : 185-198, 2006.
13. L. D. Stein. Integrating biological databases. *Nature Rev. Genet.*, 4(5):337–345, 2003.
14. H. Stuckenschmidt, F. van Harmelen. Information Sharing on the Semantic Web, Berlin, Heidelberg 2005.
15. Y. Sure. A Tool-supported Methodology for Ontology-based Knowledge Management. In H. Stuckenschmidt, E. Stubkjaer and C. Schlieder (eds.), *The Ontology and Modelling of Real Estate Transactions*, pp. 115-126. Ashgate, 2003.