Title	Dynamic Ontology Lifecycle Scenario in Translational Medicine
Author(s)	Nováek, Vít; Handschuh, Siegfried
Publication Date	2007
Publication Information	Vit Novacek, Siegfried Handschuh, Loredana Laera, Diana Maynard, Max Voelkel "Dynamic Ontology Lifecycle Scenario in Translational Medicine", Proceedings of the 5th European Conference of Computational Biology (ECCB 2006) - Book of Abstracts, in conjunction with 5th European Conference of Computational Biology, Oxford University Press, 2007.
Publisher	Oxford University Press
Item record	http://hdl.handle.net/10379/494

Dynamic Ontology Lifecycle Scenario in Translational Medicine

Vít Nováček[1,2]
Siegfried Handschuh[1]
[1]DERI, National University of Ireland, Galway
[2]Faculty of Informatics, Masaryk University, Czech Republic
email: {vit.novacek,siegfried.handschuh}@deri.org

October 25, 2006

Abstract

In this paper, we present a healthcare–oriented vision of dynamic ontology lifecycle that has been recently developed within *Knowledge Web* — EU Network of Excellence aimed at transition of Semantic Web technologies to industry. The core contribution of this paper is the proposal of methodologically and technically integral ontology lifecycle scenario, provided with extensive use case from the domain of translational medicine, showing the forthcoming impact on the biomedicine industry.

1 Introduction and Related Work

Although there already exist studies on integration of the Semantic Web and bioinformatics fields, they are rather abstract or strictly task–oriented. We can name especially [6] — a W3C report on possible adoption of Semantic Web technologies in the translational medicine. [7] describes incorporation of semantic underpinning in clinical documentation management using an ontology engineering framework. Semantics enabled approach to drug design is presented in [10]. The PROMPT tool [11] deals with semi–automatic ontology matching. However, similar approaches are tied with one particular problem and technology, even when they deliver particular application. They not reflect possibly very valuable results of recent research in complex Semantic Web scenarios.

To the best of our knowledge, there has been no deliberate effort dedicated to the deployment of coherent and complex knowledge management methodology and related tools in this field. Therefore we have developed a schema of integration of existing techniques in one holistic scenario of the ontology lifecycle. We provide it with clear demonstration of possible benefits of application in translational medicine.

The rest of the paper is structured as follows. Section 2 presents the general ontology lifecycle chart and its components. In Section 3, we describe dynamics of the scenario in use case from the translational medicine domain. Section 4 concludes the paper and gives an overview of future work.

2 Ontology Lifecycle Components

Figure 1 below contains the lifecycle chart that integrates all the interwoven phases and technologies that have been developed and/or suggested within the *Knowledge*

Web project[1].

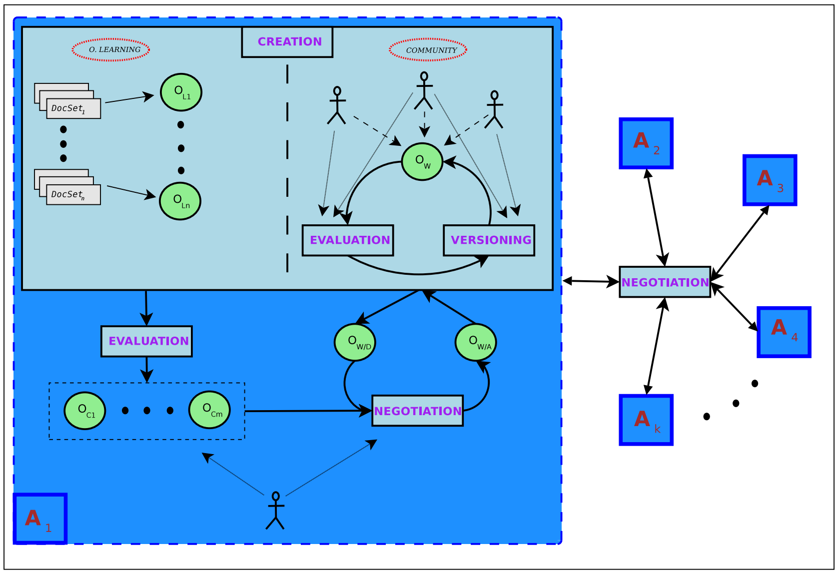

Figure 1: Dynamics in the ontology lifecycle

RDF format (see `http://www.w3.org/RDF/`) serves as a universal communication layer for all the cycle's phases, as the related tools process RDF or RDF–based data (like RDF/S or OWL). The particular components of the chart are described in the following list[2]:

- **Creation** — The *creation* phase of the ontology lifecycle has two major parts. These directly correspond to the sliced component of the chart in Figure 1. They differ in methodology, although the aim is same — to create and evolve an ontology. The first sub–component presents *manual* ontology creation. Within this phase of ontology lifecycle, people from a community are supposed to directly create and update an ontology using their expert knowledge. This is realised either by means of collaborative portals [8, 15], or using an interface based on a controlled language [13, 12]. Second component exploits methods of *automated ontology extraction* from textual resources, using NLP and machine learning techniques. Applicable ontology learning framework is *Text2Onto* [1], which builds on state–of–the–art *GATE* architecture [2].
- **Versioning** — Versioning (or change management) is no doubt considered as a must in the area of collaborative software development. The same need arises even in community driven ontology design. A universal (RDF) versioning system takes its place in the manual sub–component of the ontology creation phase. It supports commits by different users working on the same ontology, diffs

[1]The four main phases (*creation*, *versioning*, *evaluation* and *negotiation*)are indicated by the boxes annotated by respective names — either with sub–components like in case of *creation*, or without them. Ontologies or their instances in time are represented by circles and arrows express various kinds of information flow. The *A* boxes present actors (institutions, companies, research teams etc.) involved in ontology development, where A_1 is zoomed–in in order to show the lifecycle's components in detail.

[2]Note that the presented integration of concrete components and tools in the lifecycle conforms to and extends the standards of knowledge engineering methodology specified for example in [4].

between various ontology versions and other operations. RDF–centric versioning approach and its implementation in the *SemVersion* system [14] is perfectly suitable for this component of our ontology lifecycle scenario.

- **Evaluation** — The role of *evaluation* component in our scenario is twofold, following the inner split structure of the *creation* component. Firstly, we have to assess ontology quality within the cycle of collaborative ontology development (either during the production phase or when selecting preferred version of the ontology for a release dump). Assessment of learned ontology quality is needed in the ontology learning sub–component (aimed mainly at determination of appropriate candidate ontology to be passed further in the lifecycle (read more in Section 3 on concrete evaluation methods).
- **Negotiation** — The *negotiation* component resides in two different parts of ontology lifecycle again. It seeks for an optimal alignment between different ontologies [9]. Prominence of this approach resides in that it can facilitate various alignment preferences bound to different ontologies. Provided by set of possible mapping (using an ontology matching service, see [3] for details), *negotiation* component efficiently finds alignment optimally acceptable for all ontologies involved. In particular, *negotiation* triggers when different ontologies (e. g. manually created and learned) need to be aligned within one actor (institution, research team etc.), or when we need to match ontologies released by different actors.

3 Translational Medicine Use Case

Translational medicine studies methods and concrete processes that would allow direct transfer of biomedicine basic research results to clinical practice (from bench to bedside) and vice–versa. W3C consortium suggests adoption of robust knowledge management infrastructure for translational medicine (see `http://www.w3.org/2005/04/swls/`). However, no integral use case provided with tools for its implementation has been proposed yet. In the following we briefly describe direct application of the ontology lifecycle vision (recently developed within Semantic Web initiative) in translational medicine:

- **Bench** — Working ontology O_W is maintained by domain experts using the tools for *collaborative ontology creation* presented in Section 2. It reflects knowledge specific for given sub–domain of medicine research — this can be for instance relevant part of RDF dump of Gene Ontology (available at `http://www.godatabase.org/dev/`), extended with a respective taxonomy of experimental methods, processes and general concepts. Ontology is regularly evaluated using and the agreed changes are managed by *SemVersion* system [14] then. *OntoManager* [5] tool is suitable for the evaluation in this phase. Research documents and reports on experimental results are processed by *ontology learning* compoment and candidate ontology subset is selected after quality assesment using NLP–oriented evaluation methods [5] with O_W ontology serving as a gold standard. These ontologies are then used for guided population and extension of the "master" ontology in the context of *collaborative ontology creation* phase.
- **Bedside** — The *collaborative creation* phase is supressed here — appropriate part of ontology developed in the previous part of the use case is fully sufficient as a working ontology here. Optionally, it only needs to be initially extended by a very basic taxonomy of clinically–specific concepts and relations. *Ontology learning* component plays a major role on the bedside, processing the vast amounts of data in the form of electronic medical records and proprietary warehouses. It is relatively easy to mine these resources, because they usually

present already pre–structured data. The working ontology is then populated with this "practical" knowledge by the very same means of *evaluation*, *negotiation* and *versioning* components as on the bench part.

- **Transition** — This phase of the use case fully utilises the argumentation–based method of meaning *negotiation* [9]. Ontologies representing the *bench* and *bedside* parts are associated (by human agents) with ordering that defines pereferred mappings of concepts. This can reflect for example preference of lexical mappings at the *bench* side, as researchers may be interested in studying concrete drug effects at the *bedside*. On the other hand, practitioners at the *bedside* may be interested in augmenting their ontology with new methods proposed in basic research and thus prefer rather structural mappings. After generation of possible matches by ontology alignment service [3], optimal mapping according to these preferences is sought by the argumentation–based algorithm described in [9].

4 Conclusion and Future Work

We have presented an ontology lifecycle scenario that integrates so far diverse techniques employed in knowledge management process. Clearly put use case of application of this scenario in translational medicine forms the core contribution of the paper.

Our current and future work rests in continuous smoothing of borders between different lifecycle's phases by development and implementation of respective communication protocol and universal reasoning services. In order to show and deploy the tangible benefits for translational medicine research, the whole scenario needs to be accordingly prototyped and implemented then, preferably in cooperation with an appropriate healthcare industry partner.

Acknowledgments
The work presented here has been supported by the EU IST 6th framework's Network of Excellence 'Knowledge Web' (FP6-507482) and partially by Academy of Sciences of the Czech Republic, 'Information Society' national research program, the grant AV 1ET100300419.

References

[1] P. Cimiano and J. Völker. Text2onto - a framework for ontology learning and data-driven change discovery. In A. Montoyo, R. Munoz, and E. Metais, editors, *Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB)*, volume 3513 of *Lecture Notes in Computer Science*, pages 227–238, Alicante, Spain, JUN 2005. Springer.

[2] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. GATE: A framework and graphical development environment for robust NLP tools and applications. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics*, 2002.

[3] M. Ehrig and S. Staab. QOM - quick ontology mapping. In *ISWC 2004: Third International Semantic Web Conference. Proceedings*, pages 683–697, 2004.

[4] M. Fernandez-Lopez, A. Gomez-Perez, and N. Juristo. Methontology: from ontological art towards ontological engineering. In *Proceedings of the AAAI97 Spring Symposium Series on Ontological Engineering*, pages 33–40, Stanford, USA, March 1997.

[5] J. Hartmann, P. Spyns, A. Giboin, D. Maynard, R. Cuel, M. C. Suarez-Figueroa, and Y. Sure. Methods for ontology evaluation (D1.2.3). Deliverable 123, Knowledge Web, 2005.

[6] V. Kashyap, T. Hongsermeier, and S. Aronson. Can semantic web technologies enable translational medicine? (or can translational medicine help enrich the semantic web?). Technical Report No. CIRD-20041027-01, W3C & Partners Healthcare System, 2005.

[7] V. Kashyap, A. Morales, T. Hongsermeier, and Q. Li. Definitions management: A semantics-based approach for clinical documentation in healthcare delivery. In *Proceedings of ISWC 2005*, pages 887–901, Berlin, 2005. Springer–Verlag.

[8] S. Kruk, J. Breslin, and S. Decker. MarcOnt initiative. Líon Deliverable 3.01, DERI, Galway, 2005.

[9] L. Laera, V. Tamma, J. Euzenat, T. Bench-Capon, and T. R. Payne. Reaching agreement over ontology alignments. In *Proceedings of 5th International Semantic Web Conference (ISWC 2006)*. Springer–Verlag, 2006. (in press).

[10] E. K. Neumann and D. Quan. Biodash: A semantic web dashboard for drug development. In *Proccedings of Pacific Symposium on Biocomputing*, pages 176–187, 2006.

[11] N. Noy and M. Musen. The prompt suite: Interactive tools for ontology merging and mapping, 2002.

[12] T. Polajnar, H. Cunningham, V. Tablan, and K. Bontcheva. Report: Controlled language ie component version 1 (d2.2.1). Deliverable 221, SEKT, 2006.

[13] V. Tablan, T. Polajnar, H. Cunningham, and K. Bontcheva. User–friendly ontology authoring using a controlled language. In *Proceedings of LREC 2006 - 5th International Conference on Language Resources and Evaluation*. ELRA/ELDA Paris, 2006.

[14] M. Völkel, C. F. Enguix, S. R. Kruk, A. V. Zhdanova, R. Stevens, and Y. Sure. SemVersion – versioning RDF and ontologies (D2.3.3). Deliverable 233, Knowledge Web, 2005.

[15] A. V. Zhdanova, R. Krummenacher, J. Henke, and D. Fensel. Community–driven ontology management: Deri case study. In *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence*, pages 73–79. IEEE Computer Society Press, 2005.