A Practical Application of Upon Lite for the Development of a Semi-Informal Application Ontology.

Nina De Lille[1] and Ben Roelens[1,2][0000-0002-2443-8678]

[1] Open University of the Netherlands, Postbus 2960,
6401 DL Heerlen, The Netherlands
[2] Gent University, Tweekerkenstraat 2, 9000 Ghent, Belgium
[1]ncedelille@gmail.com, [2]ben.roelens@ou.nl

Abstract. The UPON Lite methodology is developed as a lightweight approach for Ontology Engineering. In contrast to more rigorous engineering methods, UPON Lite is oriented towards a reduced dependence on ontology engineers, which ensures its ease of use for the development of application ontologies. The existing scientific literature reports on six practical applications of the approach in several domains, but they lack a detailed elaboration of the development process that was followed. Therefore, this paper investigates how the UPON Lite development process is reproducible in an actual business context. This is achieved by a case study analysis of UPON Lite in the context of a public organization. For each step of the methodology, the analysis lists the guidelines of the seminal work and describes the case study implementation. This is the starting point for a further operationalization of UPON Lite to increase its adoption in academia and practice.

Keywords: Ontology Engineering, Application Ontology, Upon Lite Methodology, Case Study

1 Introduction

The UPON Lite methodology [1] is developed as a lightweight approach for Ontology Engineering, which helps domain experts better understand and communicate about their business environment. Ontology Engineering describes the set of activities that need to be performed in the development of an ontology [2]. In this respect, an ontology is defined as "*an explicit specification of a shared conceptualization*" [3]. Depending on the level of generality, a distinction can be made between top-level (or foundational), domain, task and application (i.e. depending on a task in a particular domain) ontologies [4]. In contrast to more rigorous and systematic methodologies, UPON Lite explicitly aims at ease of use and a reduced dependence on ontology engineers [1]. Given the focus of Upon Lite on the collective input of domain experts, it is particularly useful for the design of domain, task and application ontologies.

The UPON Lite development process consists of six interdependent steps: (i) domain terminology, (ii) domain glossary, (iii) taxonomy, (iv) predication, (v) parthood and (vi) ontology [1]. For each step, De Nicola & Missikoff not only describe what the outcome should be, but also how this can be achieved and which challenges should be overcome. Since it came into existence in 2016, the seminal paper of UPON Lite is cited 72 times (i.e. reported by Google Scholar on December 22, 2020). However, only few researchers report on a practical application of this method to develop ontologies. This is confirmed by a literature search, in which "Upon Lite" was used as a search string to search for relevant papers that are published since 2016. This yielded six practical applications (i.e. two papers in the Web of Science Core Collection [5, 6] and four additional works in Google Scholar [7–10]). These applications span different domains, including safety regulations [6], data science [8], intellectual property [9, 10], smart building [7] and social networks [5].

Furthermore, these papers primarily focus on the end-result of the development process, which is the shared conceptualization and standardized vocabulary about the particular domain. However, this limits actual insights for other researchers whether UPON Lite can be replicated as proposed in the seminal work. Therefore, the following research question is put forward:

> *How can the UPON Lite development process be reproduced in an actual business context?*

To tackle this research problem, the UPON Lite development process was applied in the practical context of a public organization. This case study enables us to explore the extent to which the current development process is applicable in a real-life organizational context. This is a first step that is needed for the further adoption of the UPON Lite methodology in both academia and practice.

The paper is structured as follows. While Sect. 2 describes the set-up of the case study methodology, Sect. 3 reports on the guidelines of the seminal work and the description of the case study application. This enables us to conclude the paper and identify opportunities for future research, which are described in Sect. 4.

2 Methodology

A case study is the appropriate methodology for this research as we want to answer *"how" the UPON Lite development process can be reproduced in an actual business context.* Furthermore, the development process is a contemporary set of events executed in the real-life context of a public organization [11]. Finally, UPON Lite makes the development process less dependent on ontology engineers, which means that the researcher has little control about the actual outcome of the process [11]. As we investigate a single unit of analysis (i.e. the application of the UPON Lite development process) in a one-time application, the research has a holistic single-case design [11].

The case study organization is a public agency, which is responsible to perform internal audits with authorities at different administrative levels with the aim of managing financial, legal and organization risks. In total, 30 employees participated in one or more steps of the development process, including three managers (i.e. 60% participation), five senior auditors (i.e. 71% participation), 20 junior auditors (i.e. 80% participation) and two audit support officers (i.e. 100% participation). It is important to note that a different constellation of domain experts was involved in the different steps, which is important for a broad validation of the results. The organization was chosen as a convenience sample as it is in need for a better communication between their staff members. For new or externally hired employees, it takes quite some time to master the organization-specific jargon. Moreover, there appear to be ambiguous terms between different divisions. As such, this context could benefit from the development of a semi-informal ontology, which is *"expressed in a restricted and structured form of natural language to increase clarity and reduce ambiguity"* [12]. Given this low level of formality, the UPON Lite methodology is suited for the development of an appropriate ontology.

3 Results

3.1 Domain Terminology

Seminal Guidelines [1]**.** The first step is oriented towards the identification of a lexicon, which lists the relevant terms that characterize the targeted domain. The analysis of domain corpora (e.g. textual documents, directories, dictionaries, taxonomies, standards and ontologies) is a suitable starting point, but postprocessing of the extracted terms requires intervention of domain experts. To keep this intervention effective, this social validation can be implemented by a simple voting system.

Case Study Implementation. To identify characteristic terms in the case study domain, it was decided to use four public textual documents including a management plan, a charter and management guidelines. The selection of the terms is based on the identification of the common terms, with an upper limit of 50 concepts. This choice was made to keep the scope of the next steps manageable. To this end, a self-written python program was used to retrieve the nouns from the documents. These were manually refined to identify the most common concepts. As the objective of Upon Lite is to iteratively adapt the ontology as the domain evolves, the ontology can be extended in a next iteration. Furthermore, the experts could add extra domain terms if needed. The 50 terms were socially validated by a working group of five employees (i.e. one senior auditor, three junior auditors and one audit support officer), which could decide to accept or reject their relevance. Each term that received acceptance by the majority of the employees (i.e. > 50%), was retained. This social validation resulted in the acceptance of 40 terms, while no extra terms were added.

3.2 Domain Glossary

Seminal Guidelines [1]**.** The aim is to set-up a domain glossary, in which the terms of the lexicon (from step 1) are textually defined, while also indicating possible synonyms. If possible, definitions should be based on authoritative sources. In case of various contradictory descriptions, different points of view can be resolved by social validation.

For the identification of the synonyms, it should be decided which is the 'preferred term' by a vote of the domain experts. Finally, the terms are structured in three categories (i.e. object, process and actor) and a distinction is made between complex, atomic and reference properties [13]. The resulting list of entries could be sorted alphabetically.

Case Study Implementation. Four employees (i.e. one senior auditor, two junior auditors and one audit support officer) supplemented the lexicon of 40 terms with a description and the assignment of a category (i.e. object, process, or actor). In addition, it was asked to indicate synonyms. Given the amount of work, each expert was presented a different set of 20 terms, making sure that each term of the lexicon was covered. The answers were consolidated by the ontology engineer and contradicting issues were identified.

The results were presented to the others employees for social validation, with a specific focus on open issues with respect to definitions, the assigned category and the preferred term in case of synonyms. In total, 13 experts participated in this validation round (i.e. two managers, two senior auditors, eight junior auditors and one audit support officer), which led to an alphabetically ordered glossary. Specific results per category can be found in Table 1.

Table 1. Results of the domain glossary.

Category	Unanimous assignment	No unanimous assignment	Validated Synonym
Object	9	7	0
Process	14	3	2
Actor	6	1	0

Besides the lack of a unanimous decision about the relevant category for 11 terms, the ontology engineering identified five outstanding issues with respect to a consistent naming and description of terms. These issues were taken to the taxonomy step, in which further validation is provided (see Sect. 3.3).

3.3 Taxonomy

Seminal Guidelines [1]**.** The objective is to implement a social approach to organize the domain terms in a generalization/specialization hierarchy in each of the term categories (i.e. object, process and actor). This step also includes the identification of structural concepts, which are abstract terms that are rarely used in practice but enable to structure the domain knowledge. The results can be represented in a tabular form, in which each of the columns represents a certain level of specialization. In this step, the results of the previous activities (i.e. terminology and glossary) can be further revised.

Case Study Implementation. The generalization/specialization hierarchy per category was prepared by the ontology engineer. This was particularly useful to establish a first proposal of possibly relevant structural concepts. The hierarchies were represented in a tabular and visual form and subsequently discussed by five staff members (i.e. four junior auditors and one audit support officer) during a collaborative meeting. Based on their remarks, the ontology engineer was able to refine these hierarchies. The results show that four specialization levels were used for the Object category, three for Actor, while two levels were sufficient for the Process category. The specific number of terms for each specialization level can be found in Table 2. The bold figures concern structural terms. Besides this, the open issues from the domain glossary step were discussed between the participants of the meeting. This discussion enabled to resolve the issues by obtaining an agreement on a certain alternative by the majority of the attendees (i.e. > 50%).

Table 2. Numeric results of the domain glossary.

	Top-level concepts	**First-level specialization**	**Second-level specialization**	**Third-level specialization**
Object	**2**	**5**	19	2
Process	**2**	13	-	-
Actor	**2**	6	1	-

3.4 Predication

Seminal Guidelines [1]**.** In this step, terms representing atomic, complex, or reference properties are identified and connected to the domain entities they characterize. Furthermore, the type and cardinality of the properties should be determined. If the latter aspects are too technical for the domain experts, its specification might be taken over by the ontology engineer. All these predication aspects can be represented in a tabular structure.

Case Study Implementation. The predication was prepared by the ontology engineer, which was presented for discussion and social validation in a collaborative meeting with five domain experts (i.e. one senior auditor and four junior auditors). The results of the predication were visualized in a tabular form [1] and via diagrams. Individual contact with two staff members was established afterwards to resolve outstanding issues, for which no majority was obtained during the collaborative meeting. In total, 48 reference properties with cardinality constraints were identified.

3.5 Parthood

Seminal Guidelines [1]**.** The parthood step is oriented towards connecting (material and non-material) entities to their components in a part-whole hierarchy. Specific challenges are the differentiation between parthood and specialization and between parthood and properties. Therefore, social validation should be combined with the intervention of the ontology engineer.

Case Study Implementation. Possibly relevant parthood relations were identified by the ontology engineer. This enables to complete partial decompositions by introducing additional terms (e.g. pre-investigation in Fig. 1). This preparation was validated based on a discussion by four employees (i.e. one manager, one senior auditor and two junior auditors) during a collaborative meeting. Agreement was obtained between the majority of the employees, which resulted in five decompositions visualized both in a tabular form and via diagrams. The visual representation of the decomposition of 'Audit'1 can be found in Fig. 1.

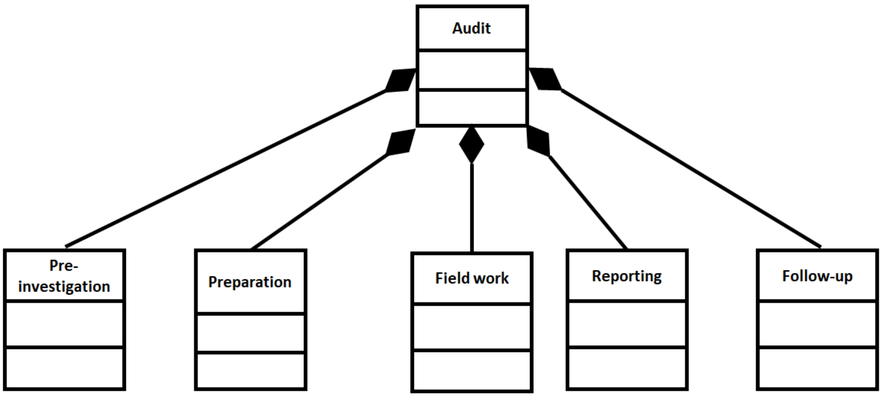

Fig. 1. Example of the decomposition of 'Audit'.

3.6 Ontology

Seminal Guidelines. The last step in the development process aims at the implementation of the final ontology based on the results of the previous steps. This includes a representation of the relations, type constraints and cardinality constraints. Furthermore, the conceptual knowledge could be encoded through a formal language, such as OWL [14]. A last important aspect is the evaluation of the ontology based on syntactic, semantic, pragmatic and social quality [15].

Case Study Implementation. All documents, tables and diagrams from the previous steps were collected and made available to all employees. To ensure that the semi-informal ontology can be easily accessed, it is implemented as a Word document with internal links to the tables and diagrams. This choice was motivated by Wikipedia, as this online encyclopedia integrates a huge amount of information through internal and clickable links. After the presentation of the complete ontology, its quality was evaluated by a questionnaire. Table 3 gives an overview of how the semantic, pragmatic and social quality is operationalized in this questionnaire. As we did not use a formal language to describe the ontology, the syntactic quality is not relevant in this case. As the

[1] Case-specific details are omitted to guarantee the anonymity of the organization.

original questionnaire [16] was used for the evaluation of conceptual models, the items were rephrased to the use of an ontology. Each item was measured on a 7-point Likert scale with response options ranging from strongly disagree to strongly agree. Additionally, the domain experts were given the opportunity to give qualitative feedback about the quality of the ontology.

Table 3. Operationalization of the quality of the ontology

Quality dimension	Operationalization
Semantic quality	- Check for inconsistencies by domain experts [1] - Perceived semantic quality [16]
Pragmatic quality	- Perceived ease of understanding [16]
Social quality	- Perceived usefulness [16] - User Satisfaction [16]

The questionnaire was completed by 22 employees, including two managers, three senior auditors, 16 junior auditors and one audit support officer. With respect to semantic quality, the check of the ontology by the domain experts did not reveal any inconsistencies in the presented models. This qualitative evaluation is supplemented by a median score of 5.75 for the perceived semantic quality [16]. This shows that the respondents tend to agree with the items underlying the construct. A similar conclusion can be drawn with respect to the perceived ease of understanding, which has a median score of 5.5. Social quality was measured by both perceived usefulness and user satisfaction with respective median scores of 5 and 5.25. This shows a slight agreement with the items. Twelve respondents also provided qualitative feedback, of which ten reactions primarily concern positive sentiments with respect to the usefulness of the ontology to improve the internal communication. The negative reactions question the motivation for the development of the ontology and the use of structural terms.

4 Conclusion & Future Research

This research investigated *how the UPON Lite development process can be reproduced in an actual business context*. The case study analysis shows that the main guidelines, as proposed by de Nicola and Missikoff [1], are applicable in the case study context. The success of the application is also supported by the positive evaluation of the semi-informal ontology by the domain experts. However, the seminal work could be extended by more detailed guidelines about keeping the scope of the development steps manageable (see Sect. 3.1 and 3.2), the use of visual diagrams to represent results (see Sect. 3.3 – 3.5), alternative forms of the implementation of the ontology (see Sect. 3.6) and an operationalization of the evaluation (see Table 3). Besides this, a further specification of the role of the ontology engineer in the development process is needed. Although the seminal work claims that its role should be reduced to the implementation of the final ontology (i.e. step 6), the case study learnt that the ontology engineer is also indispensable to prepare and take ownership of the other steps in the development process. To improve the generalizability of these findings, future work is needed to show their applicability by executing other case studies in private and public organizations.

References

1. De Nicola, A., Missikoff, M.: A Lightweight Methodology for Rapid Ontology Engineering. Commun. ACM. 59, 79–86 (2016). https://doi.org/10.1145/2818359.
2. Gomez-Perez, A., Fernandez-Lopez, M., Corcho., O.: Ontological Engineering. Springer-Verlag London (2004).
3. Guarino, N., Oberle, D., Staab, S.: What Is an Ontology? In: Staab, S. and Studer, R. (eds.) Handbook on Ontologies. pp. 1–17. Springer, Berlin, Heidelberg (2009).
4. Guarino, N.: Semantic matching: Formal ontological distinctions for information organization, extraction, and integration BT - Information Extraction A Multidisciplinary Approach to an Emerging Information Technology. In: Pazienza, M.T. (ed.) SCIE. pp. 139–170. Springer, Berlin, Heidelberg (1997).
5. Sebei, H., Hadj Taieb, M.A., Ben Aouicha, M.: SNOWL model: social networks unification-based semantic data integration. Knowl. Inf. Syst. 62, 4297–4336 (2020). https://doi.org/10.1007/s10115-020-01498-5.
6. Boualem, S.A., Meryem, B., Fatima, N.: Maintenance & information security ontology. In: 4th International Conference on Control, Decision and Information Technologies (CoDIT). pp. 312–317 (2017). https://doi.org/10.1109/CoDIT.2017.8102610.
7. Koh, J.: Metadata Models and Methods for Smart Buildings, UC San Diego, CA, USA (2020).
8. Boyd, C.: Towards a domain ontology for data assemblages. In: iConference 2019 Proceedings (2019). https://doi.org/https://doi.org/10.21900/iconf.2019.103347.
9. Jiang, P., Atherton, M., Harrison, D., Malizia, A.: Framework of mechanical design knowledge representations for avoiding patent infringement. In: Maier, A., Škec, S., Kim, H., Kokkolaras, M., Josef, O., Fadel, G., Salustri, F., and Loos, M. Van der (eds.) DS 87-6 Proceedings of the 21st International Conference on Engineering Design (ICED 17). pp. 81–90. , Vancouver, Canada (2017).
10. Jiang, P., Atherton, M., Sorce, S., Harrison, D., Malizia, A.: Design for invention: a framework for identifying emerging design–prior art conflict. J. Eng. Des. 29, 596–615 (2018). https://doi.org/10.1080/09544828.2018.1520204.
11. Yin, R.: Case study research design and methods (2th ed.). SAGE Publications Inc., Thousand Oaks, CA (2003).
12. Uschold, M., Gruninger, M.: Ontologies: Principles, Methods and Applications. Knowl. Eng. Rev. 11, 93–136 (1996).
13. D'Antonio, F., Missikoff, M., Taglino, F.: Formalizing the OPAL eBusiness Ontology Design Patterns with OWL. In: Proceedings of the Third International Conference on Interoperability for Enterprise Software and Applications. pp. 345–356. Springer, London, U.K. (2007).
14. w3c: OWL Web Ontology Language, http://www.w3.org/TR/owl-features/.
15. Burton-Jones, A., Storey, V., Sugumaran, V., Ahluwalia, P.: A Semiotic Metrics Suite for Assessing the Quality of Ontologies. Data Knowl. Eng. 55, 84–102 (2005).
16. Maes, A., Poels, G.: Evaluating quality of conceptual modelling scripts based on user perceptions. Data Knowl. Eng. 63, 701–724 (2007). https://doi.org/https://doi.org/10.1016/j.datak.2007.04.008.