A Socio-Technical Planning in Ontology Engineering

Mesnan Silalahi, Lukman, Dana Indra Sensuse, Indra Budi
Laboratorium e-Government
Faculty of Computer Sciences
University of Indonesia
Depok, Indonesia

Abstract -- Constructing ontologies is more and more a collaborative and a social process, and is usually a costly and time consuming activity. That is why this process must be prepared systematically and supported by tools and methods that will smoothen the path to consensus. The socio-technical approach has been taken to address the social aspects appropriately, in balance with the technical aspect. The socio-technical planning of the the ontology development process have been seen as the most laborious and critical phase. This phase will take up the most effort of the ontology development, and that is why this planning has to be carried out effectively. The lack of approaches in the ontology development planning has called for the inclusion of a socio-technical planning in the ontology engineering methodology with the needed tools to support the methods. In this paper the authors describes the results from the activities and tasks conducted in the planning phase of the ontology engineering process by a case study in the development of the Genetics Resources and Traditional Knowledge (GRTK) ontology. The investigation will be part of the research on a socio-technical approach in the methodology of ontology engineering.

Keywords — Ontology Engineering Socio-Technical, Planning

I. INTRODUCTION

Constructing ontologies is more and more a collaborative and a social process. In ontology engineering there are some important social aspects like a consensus building, the distributed setting, and the participatory and collaborative ontology design. These must be supported by a workbench in order to effectively support the ontology development process. Building an ontology is usually a costly and time consuming task because the ontology should be agreed widely by all the stakeholders and of the fact that knowledge evolves over time and this should be reflected in revisions. Reaching a consensual agreement is not an easy task and it requires a lot of efforts. That is why this activity must be supported by tools and methods that allow participation by expressing views and opinions freely. By this way the initiation of the smooth process to participation of the related parties will open path to consensus. Following the work in [8],the planning phase can adapt the guideline on preparing the ontology development project by answering the question such as: How can the participation processes be planned and implemented in an effective way? How can participant's needs be incorporated in the design of the tool's features and platform structures? How to choose the appropriate tools, and how to develop the participation and the collaborative platform to best fit the participant's needs? How to handle the preparation of the datas, information and knowledge resources related to the domain of interest?. This paper describes the results of the investigation from the literatures and in planning a project of developing an ontology in the domain of genetic resources and traditional knowledge in Indonesia. From the result of the investigation then we propose the methods and techniques as well as the tools in conductiong the activities and tasks identified in the planning phase of a socio-technical approach in ontology engineering methodology.

Every ontology project should begin with a planning phase to define the goal, intended use and scope of the ontology as well as design guidelines. Particularly in collaborative ontology engineering this phase is crucial to receive a shared common view on the domain of the ontology. Inadequate planning will risk the results of getting adequate knowledge resources needed to be included in the ontology. During the planning phase, thematic discussions with the stakeholders or user community are extremely helpful to capture the system requirements and the characteristics of the domain. The planning phase aims to direct people into a more content-centred conceptualization in the domain of interest by preparing the structure of the development processes. The construction of an ontology will depend on iterative elicitation and formalization of the domain knowledge. This will be reflected during the interactions with domain experts, who bring in their knowledge into the concepts whereby then formalized by the ontology engineer. This interaction is the key to successful collaborative ontology engineering. Thus, the planning phase should prepare the needed methods and tools to ensure these interactions happen during the development. Current ontology tools do not explicitly support planning and conceptualization of ontologies, but mainly on the formal editing process.

Support for planning will be vital, especially in a broader and heterogeneous communities. Domain experts will need specific tools to enter informal or semi-formalized knowledge and to discuss it with each other. Oftenly in a case of building large scale ontology they use a broad range of tools and editorial workflow in order to achieve consensus [12]. The proto-ontological resources like the vocabularies and concept terms will be needed mostly by the domain expert during the ontology capture process. The iterative character of ontology conceptualization and formalization also implies that the ontology will evolve and subject to changing during the entire process. One guideline to this challenge may be to establish a basic foundation of the ontology, and that later can be filled with detailed information. To that end, it is useful to provide an Ontology Requirement Specification Document (ORSD)., e.g. as a standard form which lists the basic planning points. This planning phase will be part of the methodological socio-technical approach in building an ontology. The methodology will consist of several phases that forms the ontology life cycle. This methodology approaches the ontology building process by taking the socio-technical aspects into accounts, alongside the whole ontology life-cycle. This lifecyle starting from the planning phase to the evaluation phase and the evolution of the ontology.

II. THE ACTIVITIES AND METHODS

The identified activities which will be carried out in the planning phase of a socio-technical approach including: the problems analysis of the domain, then defining the goals, intended use and the scope of the ontology, after that the identification of knowledge resources, information and datas, the stakeholders analysis and the identification of team members in the development project as well as the development of a workbench for the ontology itself. These activities and the related tasks will be subjects to investigation. In the initiation phase of the ontology development, the project leader will build the knowledge in the domain of interest in order to have a deeper understanding of the problem. This will include to have a comprehensive understanding of the social dynamics within the communities that have interest in the ontology. This comprises the stakeholders and the data holders. The tasks for carring out this activity will include browsing the internet for relevant informations, building rapport with the community and interviewing the stakeholders of the current work environment on the limitations and problems. By getting a deeper understanding of the problems of the domain, the team leader will be able to have the initial identification the goal, intended use and the scope of the ontology. Later in the development phase, specific communication channels should be established for fundamental discussions on the purpose, scope, and on the structure of the ontology.

The initial determination of the scope of the ontology can be carried out by interviewing the stakeholders and the identification of the availability and readiness of the related parties. The scope of the ontology can be made very broad that incorporates many parties, but can also be started from a small ones. The scenario taken must consider the availability of the resouces. The availability of knowledge resources will determine the configuration of the selected tool in supporting the ontology development processes. After scoping the ontology and determination of the intended use of the ontology, the team leader and key members can create a draft of the list of the competency questions that is needed for many uses later in the development process. All these informations will be documented for the purpose of conducting a kick-off meeting in starting the project. Although this information will build incrementally, it is important to make it explicit by documenting all the necessary requirements of the ontology in the list of competency questions and putting that all in the ORSD The use of competency questions showed to be very useful to guide ontology capturing, formalization and evaluation [2].

A. Stakeholders Analysis and Team Members

A diverse stakeholder basis is necessary to a balanced mix of views and sustainability of ontologies, especially their use and long term maintenance. It is becoming increasingly important to broaden the stakeholder base and to make this process accessible to as many participants as possible, but not at the expense of validity and ‘ontological rigour’, although even validity and rigor depend on where certain boundaries are set [3]. In our case, collaborative ontology engineering is envisioned in a setting of starting the project with a closed and rather small groups which later, on the second round will be broadened to suit the availability of research funding. In the second round, the scope will be broadened which includes external institution and possibly the open public. The scenario will require a collaborative work environment particularly in regard of access rights to data, user identification and authentification. The team comprises knowledge engineer, ontology engineer, domain experts and ontology users. This scenario requires extendable and configurable tools that has to be considered from the planning phase. Current development in the collaborative ontologies building, perceived the contribution by a community of domain experts as a critical part in the workflow. That is why we focus on building a good rapport with the domain experts, the work environment, the supporting tools and related incentives. While doing the stakeholder analysis the communication channel start to build-up in a face to face mode. This will give opportunity to build a personal rapport with the stakeholders which aims to the optimal work environment, while building a common ground and

shared concern and goal. This type of communication is needed at the startup of the projects. Later in the design phase the supporting tools will take control of the communication channel.

B. Identification of Knowledge Resources

A lot of knowledge is scattered in various non structured forms. An ontology aims to map, synthesize and resolve the conflicts that exist within the knowledge sources that constitute the body of knowledge of any given domain. When developing an ontology, all of the above knowledge sources should be considered, including information systems, databases and other existing ontologies. A few of the identified sources of knowledge and the related information systems and databases are such as:. Database of Indonesian Natural Medicines (The National Agency of Drug and Food Control), Database of Protection of Medicinal Plants and Traditional Medicine (Centre for Scientific Documentation and Information – Indonesian Intitute of Sciences), Database Plant Resources of South East Asia (Prosea), Database of National Biodiversity Information, Information Database of Infectious Diseases in Indonesia [10], The management of the GRTK had been conducted in Indonesia by research institutions and the people of Indonesia, and the datas can be categorized in:

- Resources which are already managed into a database and can be accessed;
- Resources which are already managed into a simple database but not accessible by publict;
- Resources which are not yet managed and hold by researchers or individuals (tacit).

All these information are collected and the sample datas are pulled into the data respository which is accessible via the web portal through an indexing and searching server. The scope of the ontology in the first round was the development of ontology for the medicinal plants whereby species of the sample datas are such as: Zingiber officinale, Morinda citrifolia, and Tamarindus indica. This web portal will also includes the knowledge resources needed and the proto-ontological datas. This configuration will need a comprehensive tools such as the semantic CMS like Drupal.

C. Preparing the source of Vocabulary and Concept Terms and the Proto-Ontological Data in Biodiversity Domain

We want to reuse, map and relate terms from basic vocabularies with concept definitions. Reuse terms and share a common definitions and understanding of biodiversity concepts such as: Darwin Core (DwC), Dublin Core (DCMI), Plant Ontology (PO), Gene Ontology (GO), TDWG Ontology, Disease Ontology (DO). These proto-ontological data may be textual, e.g. glossaries, single concepts enriched with explanations or definitions, unstructured collections of concepts or other notes and references to external sources. All this collected domain knowledge resources is needed as a basis to set up a formal ontology. This should be done by experienced ontology engineers. Optimal results will be achieved if ontology engineers pose concrete questions regarding the domain directly to the domain experts..

A. Preparing a workbench for Ontology Engineering

The selection of the software tools is based on the scenario of an e-participation which demands the introduction of new participation facilities into the traditional processes and the availability of the free open source software that can be downloaded from internet. Developers of ontologies use a variety of tools to smoothen the achievement of consensus and to ensure quality [9]. We first review comprehensive tools that is available whereby the tools is a web based tools that is configurable and extendable with support for webservice integration with other tools. The basic functionalities that we are trying to achieve are such as functionality for discussion to reaching consensus by availability of discussion and argumentation and voting components, functionality for record changes, annotation and associated comments, functionality for provenance links in concept histories, functionality for personalized views of an ontology based on user's role and level of expertise, functionality to control user roles and access rights, and the support for different levels of ontology expressiveness

Most of the tools that have been developed are to support individuals locally. Support for distributed team members to collaborate on building ontologies has been somewhat lacking. Importantly, the existing tools mostly only focus on the ontology editing. Reference [6] reported on new tools that are developed with the focus to support the collaborative ontology building process such as BibSonomy, Collaborative Protege, DBin, Hozo, Ontowiki, and Soboleo. The consideration which was raised was the support of ontology visualization. Users need to understand and modify ontology elements developed by others, so having visual aids and other features to increase ontology comprehension is critical. Furthermore she identified that people doubt any nontrivial ontology can be built using only asynchronous communication such as the discussion tool.

The success of a participation scenario depends heavily on the organizational planning and the incorporation of such initiatives into the processes along the different stages in the ontology building life-cycle. That is why this activity must be supported by tools and methodologies that allow participation by expressing views and opinions freely. By this support, the participation of the related parties will smoothen the wide acceptability of the developed ontology. The

Web 2.0 applications has showed us how communities can interactively contribute and share ideas through blogging, comments, and voting. Applying these functionalities to the semantic web means giving more support to users in participatory ontology building. Several researches showed how semantic web users can successfully collaborate to negotiate and build ontologies when provided with a tool that supports such activities.

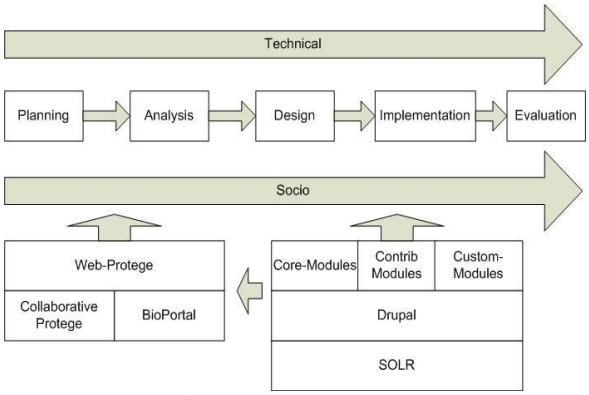

Fig. 1. A workbench for Socio-technical Ontology Engineering

The tools that are integrated will depend on the scenario of the ontology building. The scenario will be determined during the initial phase of the ontology building process. A scenario will require a special workflow to support the process. A special purpose tools will be different than a tools which will be used for general purpose one with public user as a user target. A sample of special purpose tools could be the software tool based on Scratchpads (based on Drupal) which primary purpose is as a biodiversity data publishing tool in the domain of biodiversity. For a socio-technical approach, there should be several tools that can support various needs and circumstances in a collaborative scenario. In some cases there is a need for tools that everybody can make changes, and on the other there is a need for functionality to do voting, argumentation aggregation and for making changes whereby only authorised user can make changes. So there is a need for tools that is capable to support various mechanism in consensus building, that depends on wether the environment is controlled or open. Based on analysis of workflows in several large scale collaborative ontology development projects, [12] showed that the workflow for each project is different and involves different steps. Some of the methodologies in ontology engineering which addressed the collaborative process used the wiki technology as a tool which is in most cases integrated with an ontology editor and visualization tool to help user easily comprehends the ontology. The use of a graphical language for expressing ontologies proved to be essential for ontology capture. Reference [2] stated that it is very hard to communicate with domain experts without it. Further he stated that hypertext proved to be an adequate format for documenting ontologies. Using hypertexts, ontologies can be easily browsed, and people can use them to learn about the domain. This result was confirmed by the result in which stated that users pointed out explicitly that a Web interface was ideal for a collaborative tools.

The framework is based on a CMS software called Drupal. Drupal has enjoyed the support of thousand of developers and has proven tracks in supporting the community driven web content publishing and social networking. Recent advancements in the direction of semantic web has made this software promising to use as a platform in the development of a workbench for ontology engineering. For that purpose some tools or modules will be integrated within Drupal to support the various methods that is identified to carry out the various tasks in the ontology building processes. The selection of the Drupal as a workbench to ontology building was carried out with the usefulness in minds that can be shown to the stakeholders. The extendability of the Drupal modules and the support from the community will ease the development of application of semantic web search. The developed application aims to show the usefulness of the built ontology for the solution of the current problems. In our case study, we develop a semantic web portal for biodiversity and traditional knowledge. The web portal is a collaborative platform whereby the domain ontology is build in parallel with the development of semantic web application. The Semantic application with Drupal will be configured in two main activities namely the content management and the ontology management. The content management will focus on the publishing of the taxon pages which panels consist of datas from different sources integrated within Drupal. This can be achieved by configuring the pages with a semantic views module that is now the core module of the Drupal. The ontology management will mainly focus on the functionalities for ontology creating, editing, deleting, annotating and storing. This could be partly supported either with tools from the Drupal modules such as Neologism for light ontology building or delivered from integrated tool such as Collaborative Protege and the Web Protege. The choice of integrating Semantic Drupal with WebProtege will support more capabilities and functionalities to socio-technical ontology development process. For the purpose of building a more heavyweight ontology, account will be taken to using the more comprehensive tools such as Collaborative Protege plus the web based tool such as WebProtege. All the functionalities for supporting the social process can be supported within Drupal.

III. RESULTS AND DISCUSSION

The socio-technical approach has been taken to address the social aspects in balance with the technical aspects in planning an GRTK ontology. This approach has seen the importance of the human aspect in the social construction of a technical engineering.

Planning phase in the ontology development process is a crucial step because of the high workload of the setup activities for methods and tools used before the actual bulding of the ontology started.

In the planning phase of the building a GRTK ontology, we identify some activites, tasks, and the needed methods and tools. These results had been described in the preceding sections. The results will be incorported in the proposed methodology in ontology engineering. Tabel I summarize the preliminary results of the investigation carried out on the planning phase of a socio-technical methodology in ontology engineering. The socio-technical approach of the proposed methodology incorporates the concepts of meta-design which requires the availability of a socio-technical system. This system was developed and setup during the planning phase.

The ontology development project is carefully planned in order to get the development started smoothly and then proceeds to open a path of reaching consensus.

Tabel I. The Activities, Tasks, Methods and Techniques in The Planning Phase

Activities	Tasks	Methods	Techniques
Domain Problems Identification	Identify the current issues, problems and limitations within the domain	Problems Domain Analysis	Brain-storming,
	Identify Solutions to the Problems		Brain-storming,
	Identify and document the Intentions and Goal of the Ontology		Brain-storming, ORSD
Stakeholder Identification	Identify Stakeholders	Stakeholders Analysis	Survey and Browsing
	Build Communication with Identified Stakeholders		Interviews
	Build a common ground and shared concerns		Interviews
Participants Identification	Identify Domain Experts		Interviews
	Identify Data Holders		Survey
	Identify Users		Interviews
	Identify Knowledge Engineers		
Resources Identification	Identify The Proto-Ontological Datas	Data Analysis	Survey and Browsing
	Identify The Standards		Survey and Browsing
	Identify the Existing and Related Ontology		Survey and Browsing
	Identify existing applications information and datas in the Domain		Survey and Data Acquisitions
Tools Selection	Review Existing Tools	Software Engineering	Software Evaluation
	Review Relevant Tools		Software Evaluation
	Test the Tools		Software Testing
Work Environment Preparation	Develop, Configure and Customize Software Tools		Web-Portal Development
	Uploading Knowledge Resources to the Repository		Web-Portal Development
	Prepare Manuals, Training Materials and Introduction Presentation		Knowledge Repository Development
Manage the Initiation		Kickoff Meeting	Survey
	Identify the Accomodation and Budget	Project Management	
	Build the agreed Mechanisms and Procedures	Requirements Analysis	FGD, ORSD
	Determine the scope of the Ontology		FGD, Competency Questions
	Build the Commitments		

The methodology describes the phases in the ontology engineering with socio-technical activities whereby the tools needed to support is mainly within the drupal platform. This platform offers the flexibility and the capability in adding new functionalities apart from the core modules. A semantic CMS based on the Drupal platform for the development of a framework in supporting the ontology engineering will have the

capability to take the socio-technical aspects into accounts whereby the creation and editing of more complex and expressive ontology will be handled by a mature tools such as WebProtege. Effort in integrating this tool with Drupal has been made for a more comprehensive framework.

Preparing the kickoff meeting before the development started is described in activites like building knowledge and the problems analysis in the domain of interest, preparing the knowledge sources and tools within the portal to access. This will support the domain experts in the building process. But before all that development can proceeds smootly, the team leader has to build good rapport with all the stakeholders and build a shared concern and goal. For this purpose, the team has attended several events of the community activities such as FGDs, socialization and training. The training was about the integration of the biodiversity data across the R&D Centres within the Indonesian Institutes of Sciences. The tools used for that was GBIF toolkit. The shared goal is then put explicitly in the ORSD documentation. The goal, intended use, and scope of the ontology defined in this phase is not final but will evolve iteratively in the life-cycle of the ontology building process. To sum-up the planning phase will try to take all possibilities into consideration by viewing everything as an open-system and in order to make the right decision at the right time and to tackle every problems or challenge systemically.

REFERENCES

[1] Correndo, Gianluca, and Harith Alani. "Survey of tools for collaborative knowledge construction and sharing." Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology-Workshops. IEEE Computer Society, 2007.

[2] de Almeida Falbo, Ricardo. "Experiences in using a method for building domain ontologies." Proceedings of the 16th International Conference on Software Engineering and Knowledge Engineering, SEKE'2004. International Workshop on Ontology in Action, OIA'2004. 2004.

[3] Di Maio, Paola. "'Just enough' ontology engineering." Proceedings of the International Conference on Web Intelligence, Mining and Semantics. ACM, 2011.

[4] KOZAKI, Kouji, Osamu SAITO, and Riichiro MIZOGUCHI. "A Consensus-Building Support System based on Ontology Exploration." (2011).

[5] Mainz, Indra, et al. "Ontoverse: collaborative ontology engineering for the life sciences." Information Wissenschaft und Praxis 59.2 (2008): 91.

[6] Noy, Natalya F., Abhita Chugh, and Harith Alani. "The CKC challenge: Exploring tools for collaborative knowledge construction." Intelligent Systems, IEEE 23.1 (2008): 64-68.

[7] Paulsen, Ingo, et al. "Ontoverse: Collaborative knowledge management in the life sciences network." Proceedings of the German eScience Conference. 2007.

[8] Scherer, Sabrina, Maria A. Wimmer, and Stefan Ventzke. "Hands-on guideline for e-participation initiatives." E-Government, E-Services and Global Processes. Springer Berlin Heidelberg, 2010. 49-61.

[9] Sebastian, Abraham, et al. "A generic ontology for collaborative ontology-development workflows." Knowledge Engineering: Practice and Patterns. Springer Berlin Heidelberg, 2008. 318-328.

[10] Sensuse, Dana Indra, and Lukman Lukman. "Knowledge Management Model and Strategy of Genetic Resources and Traditional Knowledge in Indonesia." International Journal of Computer Science Issues(IJCSI) 9.2 (2012).

[11] Siorpaes, Katharina, and Martin Hepp. "myOntology: The marriage of ontology engineering and collective intelligence." Bridging the Gep between Semantic Web and Web 2 (2007): 127-138.

[12] Tudorache, Tania, et al. "Ontology Development for the Masses: Creating ICD-11 in WebProtégé." Knowledge Engineering and Management by the Masses (2010): 74-89.