Organization-Ontology Based Framework for Implementing the Business Understanding Phase of Data Mining Projects

Sumana Sharma and Kweku-Muata Osei-Bryson
Virginia Commonwealth University
sharmas5@vcu.edu, kmuata@isy.vcu.edu

Abstract

CRISP-DM is a detailed and widely used data mining methodology that aims to provide explicit guidance regarding how the various phases of a data mining project could be executed. The 'business understanding' phase marks the beginning of a data mining project and forms the foundation for the execution of the remaining phases. Unfortunately, the real-world implementation of this pivotal phase is performed in a rather unstructured and ad-hoc manner. We argue that the reason for this lies in the lack of support in form of appropriate tools and techniques that can be used to execute the large number of activities (=67) prescribed within this phase. This paper presents an organization-ontology based framework that not only incorporates the applicable tools and techniques, but also provides the ability to present the output of activities in a form that allows for at least their semi-automated integration with activities of this phase and succeeding phases.

1. Introduction

Enterprise decision making is continuously transforming in the wake of ever increasing amounts of data and information. A recent KD nuggets poll (June 2007), http://www.kdnuggets.com/polls/ based on the largest data size data-mined found that nearly 22% of the respondents reported mining databases of 1 terabyte or more which is double the 11.5% of respondents who mined terabyte size databases in 2006. Such findings affirm the surge of interest in the field of Data Mining (DM) that offers a plethora of models and techniques to quickly analyze huge volumes of data, thereby increasing an organization's ability to make sound decisions. However, implementation of DM techniques requires a significant commitment of time and resources, and must therefore be a well-coordinated and disciplined effort. A Data Mining methodology provides the much needed structure to the data mining efforts of an organization. CRISP-DM, an acronym for CRoss Industry Standard Process for Data Mining [1], is a leading DM methodology which describes the life cycle of a data mining project in form of six different phases, namely, business understanding, data understanding, data preparation, modeling, evaluation and deployment (see Figure 1). Per CRISP-DM methodology, the business understanding phase forms the foundation of the DM project.

The Business Understanding (BU) Phase is however not specific to the CRISP DM methodology. Other DM methodologies (for e.g., [2],[3],[4],[5],[6]) also acknowledge the importance of the development of business understanding, before dwelling into later stages such as data preparation and modeling. Being successful in data mining requires planning and understanding the business problem [7], a focus of the BU phase. Decisions made during this phase affect the choices and outcomes of succeeding phases. However, unlike phases such as data preparation and modeling, the BU phase is constrained by a lack of support in form of applicable tools and techniques which specify "how" an activity can be carried out.

We must clarify that there is no dearth of research related to activities that could be used to develop business understanding. Many DM models and methodologies (see examples above) provide prescriptive lists of activities to execute this phase. However, there is relatively little research that explicitly specifies the tools or techniques that could be used to actually carry out the prescribed activities. In fact, [8] believe this to be a broader issue plaguing all phases of DM methodologies and state that "DM methodologies provide very little detailed advice to the novice miner on how to actually carry out a given step". We believe that this issue is more dominant in case of the BU phase and argue that the general lack of support is responsible for the unstructured and ad-hoc manner in which this

1530-1605/08 $25.00 © 2008 IEEE

phase is often carried out in real-world data mining projects. The description of many published DM case studies seems to indicate that a business understanding of the data mining problem domain is either presumed or only minimal efforts are made to execute the BU phase in a formal manner. This issue has been highlighted and somewhat addressed by [9] who describes how real world business problems (to be addressed through data mining) can be modeled. While the author has not based his approach on any particular DM methodology, he discusses various tools to carry out many (though not all) of the activities prescribed under the BU phase of the CRISP-DM methodology. However, these are only presented in a linear fashion, with the description of each activity followed by a brief description of a proposed tool. The overall framework which consists of nested sequences of action boxes, discovery boxes, technique boxes and example boxes is complicated to navigate, and may appear to be cumbersome or even cost prohibitive to actors involved in carrying out the critical BU phase.

The user guide portion of CRISP DM methodology [1] also purports to provide detailed advice about "how" to execute DM activities. With respect to the BU phase, the user guide does not meet its intended objective and only proposes a checklist of activities to be performed to accomplish the tasks associated with this phase. The only applicable tool mentioned in this phase is the use of an organizational chart, to "identify divisions, manager's names and responsibilities etc". Clearly, organizations also need support for the diverse array of other activities associated with this important phase. Besides, the usefulness of organizational charts, a primarily static entity, to identify organizational actors and their interrelationships can be also be debated.

On the other hand, the strengths of the CRISP-DM methodology lie in its detailed guidance related to the various phases of a typical DM project, related tasks, their desired outputs and prescribed activities. This paper leverages its strengths by providing a framework which aids development of business understanding and incorporates relevant tools and techniques proposed in the broader Information Systems Literature to facilitate the same.

Anchoring our effort is the observation, that the different phases utilize the output of the preceding phase(s) as the input. While such a dependency is shown in the CRISP-DM reference model, it is not incorporated in the design of the various phases of the methodology, and is not necessarily implemented in practice. It appears to us that the dependencies between the various phases provide an opportunity for linking at least some of their outputs in a semi-automatic way. Such semi-automation is likely to provide an increase in the efficiency and reliability with which DM projects are currently implemented.

Our analysis of the methodology has also revealed opportunities for semi-automating (by linking the outputs of activities) the execution of at least some of the activities of the BU phase itself. CRISP-DM user guide briefly mentions some of these dependencies. While we concur with [7] who state that it is impossible to "completely" automate the understanding of a business problem, we also believe that a lot of activities that aid the development of understanding a business problem can be incorporated in a suitable framework that allows for at least their semi-automation wherever possible, thereby paving the way for more efficient DM projects.

By semi-automation we mean that the dependencies between the various activities can be leveraged by linking their outputs, wherever possible. The outputs of several activities of the BU phase can be used as an input to some activities within the BU phase and some beyond the BU phase. An example of the former is that the output of activities 'specification of benefits' and 'estimation of costs' can be used to at least semi-automate (if not completely automate) the execution of activity 'conduct a cost benefit analysis'. An example of the latter could be that the output of activity 'data mining success criteria for model assessment' could be used as an input to activity 'selection of data mining models' in the modeling phase, whereby the selection of the best model could be at least semi-automated by having the DM software select the model that best fits the evaluation criteria. This saves the user from manually assessing the large number of models churned out by DM software to see which model(s) meets all the specified success criteria. Accordingly, our concurrent objective is to ensure, wherever possible, that (1) the desired outputs of the activities within the BU phase are generated in a

form that will allow for at least their semi-automated integration with activities within the succeeding phases of the DM project; and (2) that the desired outputs of the various activities of the BU phase are generated in a form that will allow for at least their semi-automated integration with other activities within this phase itself. Section 2 describes the BU phase of the CRISP-DM methodology. Section 3 presents our proposed solution. Section 4 describes the summary and presents directions for future research.

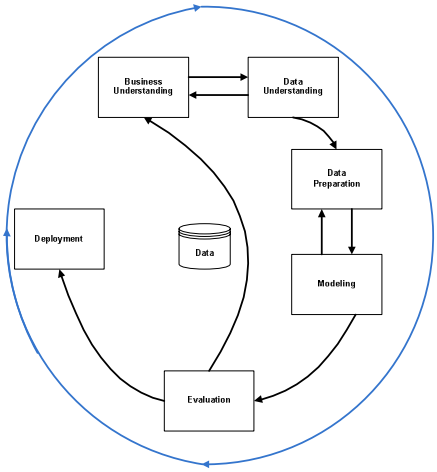

Figure 1: Phases of CRISP DM reference model [Source: CRISP DM v.10 guide]

2. BU Phase of CRISP-DM methodology

The BU Phase of the CRISP-DM methodology is centered on understanding the objectives and requirements of the data mining project from a business perspective, and converting this knowledge into a data mining problem definition and a preliminary plan to achieve the set objectives. This phase consists of four main tasks: (1) Determination of Business Objectives, (2) Assessment of situation (3) Determination of Data Mining Goals and (4) Production of Project Plan.

The first task, 'determining of business objectives' aims at developing a thorough understanding, from a business perspective of what the customer really wants to accomplish. The second task, 'assessment of situation' is a fact finding exercise about the resources, constraints, assumptions and other factors related to the project. The third task, 'determining of data mining goals' is related to translating the business goal(s) into a data mining goal(s). The final task, 'production of a project plan' is intended to describe the intended plan for achieving the data mining goals(s) and business goal(s). CRISP-DM user guide describes the desired outputs expected to be generated after accomplishing each of these tasks, and a checklist of activities to help achieve these desired outputs (see Section 3.3)

3. Solution

Our solution approach is guided by our research objectives, (1) creation of a framework that incorporates tools and techniques for executing the various activities of the BU Phase; and (2) ensuring, wherever possible, that the desired outputs of the activities of the BU phase are generated in a form that will allow for at least their semi-automated integration with activities of the succeeding phases of the DM project, and with activities of the BU phase itself.

There are a total of 67 different activities prescribed under the BU phase (see section 3.3). Even a cursory analysis of these activities is enough to reveal the diverse array of concepts and entities involved in the BU phase. These span organizational actors, divisions, business and data mining goals of the project, necessary resources and a plan for implementing the project. There are several dependencies between these concepts and entities. For instance, identification of relevant organizational actors and their roles and responsibilities is indispensable for the execution of the project. These actors are in turn responsible for determination of business and data mining goals which they accomplish using resources allocated to the project.

Note that these concepts and entities actually belong to the larger organization and are typically represented in an organization ontology. We believe that an organization ontology can provide the framework within which the activities of the BU phase can be executed. The organization-ontology framework can also be used to incorporate the tools and techniques applicable to the various activities of the BU phase. In Section 3.1, we discuss what we mean by an 'organization ontology', provide our rationale about 'why' it should be used a framework, and 'how' it can lead to generation of outputs of activities in a form that allows for

at least their semi-automated integration with other activities. In section 3.2, we propose the design of an organization-ontology based framework. In section 3.3 we propose how tools and techniques necessary for implementation of various activities can be incorporated within the proposed framework. It also lists the various tasks involved in the BU phase, their desired outputs and the activities suggested in the CRISP-DM methodology's user guide to accomplish each of these tasks.

3.1 Rationale for organization-ontology based framework to implement the BU phase

One of the most cited definition of ontology is the one provided by Gruber [10]. He defines ontology as "an explicit specification of a conceptualization". In essence, an ontology is the formal specification of concepts belonging to a certain domain, and their interrelationships. An 'organization ontology' models an organization in form of an information system [11]. By "formalizing" the presence and relationships between various concepts and entities, it is able to facilitate their fast and easy retrieval. As we pointed out above, the activities of the BU phase span across concepts and their relationships described in an organization ontology. When used as a framework, the organization ontology affords us the ability to execute the various activities of the BU phase, by specifying the concepts and relations relevant to each activity in form of simple queries (see [11] for examples). The tools and techniques used by activities of the BU phase can be modeled as instances of *Resources,* a concept (or class) within the organization-ontology framework. Our research reveals that no other framework has the ability to account for as many of the activities involved in BU phase. Since each activity of the BU phase can be specified through relevant concepts and their relationships, the output of each activity can also be linked to appropriate concepts of the organization ontology. For instance, an activity specification of 'primary and secondary business objectives' of the BU phase, leads to an output in form of primary and secondary business objectives. These objectives can be regarded as sub-goals in an organization ontology and are supposed to be decomposition of a broader organizational goal(s). These can be modeled, as such within the organization ontology. Such modeling also allows for the presentation of output in a form that allows it to be linked to another activity such as 'production of a project plan', which among other things, incorporates the primary and secondary business objectives of the project. Such linking between activities can be semi-automated or automated depending on the context and the nature of the activity. The use of ontologies for semi automating or automating activities ranging from Business Process Management [12], to use of Intelligent Discovery Assistants to aid the selection and ranking of valid DM processes [13], inspection and quarantine processes at airports and even achievement and dissemination of strategic intelligence [14] have been proposed in the literature. [13] state that by using ontology-based Intelligent Discovery Assistants they have been able to "automate parts of the data mining process that are better understood leaving research to concentrate on areas that are not well understood". Clearly, the same opportunity is also available in the context of the BU phase. Sections 3.2 and 3.3 provide more detail about how such an opportunity can be leveraged.

3.2 Proposed Ontological framework for implementing the BU phase

Besides providing an ability to formalize the existence of concepts and relations, ontologies also allow for their easy extension (specialization) to suit a particular domain. In this section we describe an example of an organization ontology and how we have extended it to develop an organization-ontology based framework suited to the context of the BU phase of a DM project.

An example of organization ontology is presented by [11] who consider an organization to be a set of constraints on the activities performed by agents. Their proposed ontology refers to numerous concepts, such as *Organizational Actors, Organizational Goals, Activities, Resources,* etc, all of which are linked with each other through relevant relationships. We have extended upon the organization ontology proposed by [11] in developing an organization-ontology based framework to aid the BU phase of a DM project (see Figure 2). Note that the proposed organization ontology has a high degree of generality. This was intended as we wanted to represent all the necessary concepts and entities that are considered by an organization before it embarks on any projects that require significant commitment of time and resources.

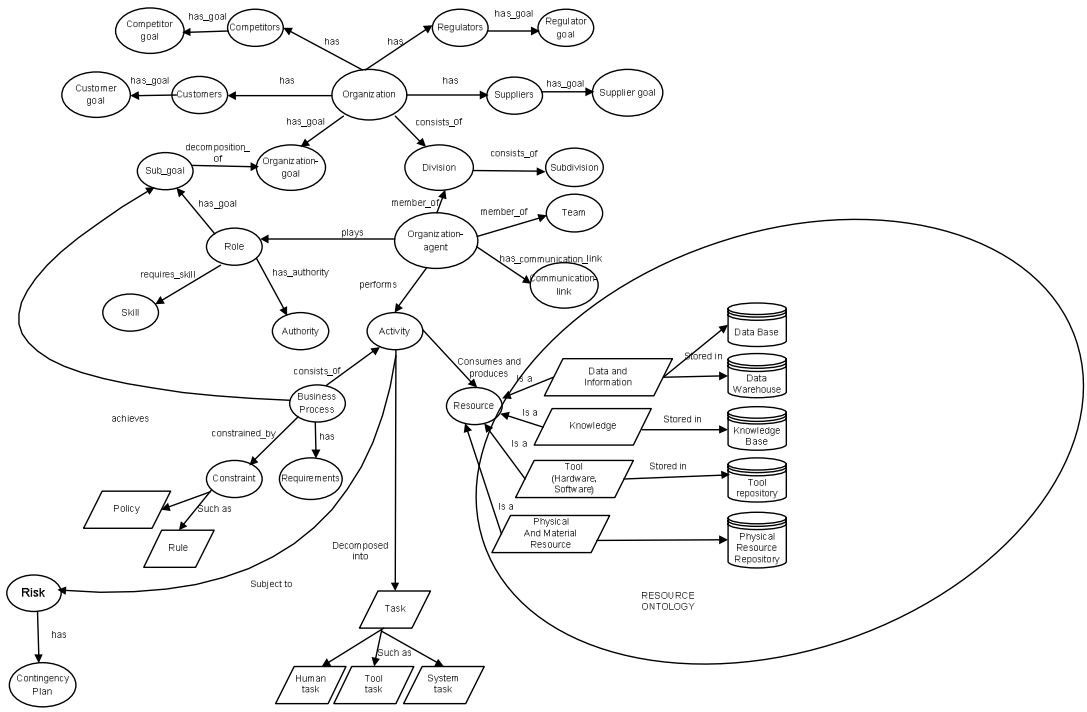

Figure 2: Extended Organization Ontology

We have followed the 'second generation knowledge engineering approach' of [11] in developing our proposed enterprise ontology. Their approach to 'engineering' ontologies (in contrast to the method of extracting rules from experts) starts with defining an ontology's requirements in form of questions that an ontology must be able to answer. This is referred to as the competency of the ontology. The second step consists of defining the terminology of the ontology – its objects, attributes and relations. The third step is to specify the definitions and constraints on the terminology, where possible and the last step is to test the competency of the ontology by "proving" the competency questions. The scope of this paper has been limited to proposing the competency questions and relevant objects (concepts) and their relationships that can be used to prove the competency questions.

We define the enterprise ontology's requirements or competency as the various activities described in the CRISP-DM methodology. In other words, we started with the aim to develop an enterprise ontology which could be used to execute all the suggested activities. These activities can be envisioned as questions. For instance, the activity 'identification of key persons and their roles' can be posed as a competency question: what are the key persons and their roles? This is an example of activities (questions) which can be executed (answered) by simple navigation of the ontology. As we pointed out above, here we are referring to the use of the enterprise ontology as a tool itself. There are some other activities (questions), however, which can be executed (answered) using certain other tools; the tools being modeled as instances of *Resources* under the ontology. For instance, the activity 'estimation of operating costs' can be posed as a competency question: what are the operating costs of this DM project? The ontology can model the operating cost of a project as an attribute of '*Activity'* (a class or concept in the proposed ontology). However the actual response in form of a value will have to be sourced from a project management type tool (instance of *Tool Resource*) that allows for calculation of operating costs associated with the various activities.

The organization ontology proposed by [11] consists of the following classes: *Organization, Organization Goal, Sub Goal, Division, Sub Division, Organization Agent, Team,*

Communication Link, Role, Skill Authority, Activity, Resource and Constraint. Upon navigating their organization ontology, we find that an organization consists of divisions, and divisions consist of sub-divisions. Organizational actors or agents are members of division(s) and also part of teams set up to pursue specific tasks. In contrast to divisions, teams are temporary in nature and set up when needed. Note that the concept of a team is especially important in the context of DM projects, where a variety of organizational actors come together to implement a DM project. Organizational agents play one or more roles within the organization. Each role is associated with one or more sub-goal(s) which are decomposition of the organizational goals. Each role requires certain skills and is allocated proper authority at the level that the role can achieve its goals. Organization agent performs activities which use, produce, consume or release resources. The activities are constrained by constraints in form of policies and rules.

We have extended the organization ontology of [11] by including certain other critical concepts which must be considered if the organization ontology can be leveraged by an organization when it performs knowledge intensive projects. We describe such classes below. We also note any changes proposed by us to the original organization ontology of [11].

First, we believe that the organization ontology must be extended to include relevant entities, such as *Customers, Suppliers, Regulators and Competitors*. As [15] noted, "Decision problems are thrust upon us by the actions of others: competitors, customers, government, and stakeholders; or by circumstances: recessions and natural disasters". By including customers, suppliers and regulators and their specific goals, within our organization ontology we can ensure that the organization goals are in sync with the goals of these important stakeholders. By including competitors and competitor's goals, we can ensure that the competitor is not gaining a competitive edge by aligning its goals to that of the organization's customers. We believe that knowledge intensive organizations should make efforts to steadily populate the ontologies of their customers, suppliers, regulators and competitors. By formalizing vital information and knowledge in form of an ontology and linking it to the organization ontology, an organization can (1) position itself to maintain a strong competitive edge over its competitors; (2) maintain a healthy relationship with its customers and suppliers; and (2) ensure efficiency by tying its processes to the most current objectives of the government regulators.

Second, we believe that activities are a decomposition of organizational business processes. A *Business Process* (an ordered execution of activities) helps to achieve organizational sub-goal(s). It is a critical construct in any organization and should be explicitly modeled within an organization ontology. This relation (process-has-sub goal) is in addition to the relation that all roles have certain sub-goals, and when an organization agent is assigned to the role, a commitment is created on part of the agent to accomplish the goals associated with the given role. Our view is different from that of [11] who state that since plans and processes are constructed by combining activities, the aggregate activities or processes should also be regarded as an activity. We believe that by separating representing processes and activities, an organization may be able to better leverage its knowledge assets, and assess whether or not processes are helping to accomplish organizational sub goals.

Third, we believe that it is an organizational business process, and not activities, that is constrained by constraints in form of rules, policies etc. This view is different from that of [11] who model activities to be constrained by constraints. By explicitly modeling a business process as a separate class we have been able to represent the relation (process-has-constraint) and remove the ambiguity that may arise by directly associating activities with constraints.

Fourth, organizational business processes are based on requirements. *Requirements* can be regarded as a specification of what the organization expects from the implementation of the business process. By incorporating requirements in the organization ontology, an organization can easily assess whether or not specific business processes are fulfilling their associated requirements.

Fifth, an activity can be described as composed of a series of *Tasks* which can in turn be categorized into *System Task, Tool Task, And Human Task*. Such increase in the granularity of the organization ontology is necessary to

implement critical projects, which often span multiple nested activities.

Sixth, the various types of resources, i.e. sub classes of resources should also be included in the organization ontology. Organizational resources can be classified into sub classes such as *Data Resources* (such as Databases, Data Warehouses etc), *Information Resources* (such as Metadata repository, documents), *Knowledge Resources* (such as Organizational Actors, Knowledge Base, Case Base, Group Memory), *Tool Resources* (such as Hardware, Software), and *Physical or Material resources* (such as infrastructure). By organizing resources in form of the proposed sub classes we are able to create a resource ontology. Tools and techniques used for execution of a particular activity can be researched by simply by navigating the organization ontology link between activity and resources and lead to efficient and easy retrieval of resources.

Seventh, activities are often exposed to risk which may degrade the activity's execution or performance. When organizations launch critical projects they need to thoroughly analyze the risk possibilities and develop contingency plans that help to avoid or minimize the occurrence of the risks. Since *Risk* is an important construct within an organization, it should be explicitly modeled in the organization ontology. A *Contingency Plan* class representing the plans to avert or minimize the risks should also be modeled explicitly in the organization ontology.

3.3 Relevant tools and techniques to implement the activities of the BU Phase

In this section we describe the tools (hardware or software) and techniques to perform the various activities associated with the BU phase of the CRISP-DM methodology. These tools and techniques are instances of particular *Resource* sub classes in the organization-ontology framework proposed by us in the above section. Note that certain activities such as 'identification of organization agents and their roles' can be executed by simple navigation of the ontology. Alternatively, the user could pose simple queries to look up the same information. The classes involved in such a query would be *Organizational Agent* and *Role*. This is an instance of activity where the organization ontology is being used as a tool itself, and no other tool is required for its execution. However, some other activities such 'description of business problem' may require particular tools such as Causal Maps and Cognitive Maps for their execution. These are instances of the *Tool Resource* Sub class of the Resource Ontology (parent class being *Resource)*. Due to space constraints, these particular instances have been described (in quotes) along side the appropriate classes or sub classes relevant to execution of each of the activities (the actual query would only include the relevant classes, sub classes and the relations). For each of those activities that cannot be executed by simple navigation of the ontology, we have proposed specific tools (instances of particular *Resource* sub classes) that can be used to execute the activity. Organizations can use the prescribed tools or techniques to accomplish the given activity. Activities have been classified by grouping them under the desired outputs they should generate and the broader task that they are expected to accomplish. Note that the sequence of some desired outputs and some of the prescribed activities may seem counterintuitive. For instance, the activity 'identification of project sponsor', the organizational agent who approves the necessary funding for the DM project) may not be completed before 'at least' some preliminary business objectives have been formulated. However, addressing the sequencing of activities of the CRISP-DM methodology is outside the scope of this paper. Assuming that the proposed organization ontology is populated with necessary instances, the activities can be executed by posing simple queries. We have highlighted which classes would be relevant in executing the queries for all the 67 activities prescribed under the BU phase.

Task 1: Determine Business Objectives

Desired Output 1.1: Collect relevant background information

- Develop organization charts identifying divisions, sub divisions and project groups: *Division, Sub division and Team*
- Identify key persons and their roles: *Organizational agent, Role*
- Identify a project sponsor: *Organizational agent, Role*
- Identify a steering committee and its members: *Organizational agent, Role*
- Identify business units impacted by the DM project: *Division*
- Describe the problem in general terms: *Organizational agent, Role, Resource (Tool*

Resource – "Causal Maps and Cognitive Maps")

- Check the current status of the project: *Resource (Knowledge Resource – "DM project Base")*
- Clarify prerequisites of the project: *Organizational agent, Role, Resource (Tool Resource – "Requirements Analysis Tool, Causal Maps")*
- Present DM to business: *Organizational agent, Role*
- Identify target groups for the project report *Organizational agents, Role, Resource (Tool Resource – "Requirements Analysis Tool")*
- Identify the users' needs and expectations: *Organizational agent, Role, Resource (Tool Resource – "Requirements Analysis Tool")*
- Describing any current solution in use: *Resource (Knowledge Resource – "DM project Base")*

Desired Output 1.2: Determine business objectives

- Informal description of problem *Organizational Agent, Resource (Tool Resource – "Causal Mapping Tools, GSS Type tools")*
- Specification of primary and secondary business objectives(s): *Organizational Agent, Role, Resource (Tool Resource – "AHP, Goal Map")*
- Specification of any other business requirements: *Organizational Agent, Role, Resource (Tool Resource – "Requirements Analysis Tool")*
- Specification of expected benefits: *Organizational Agent, Role, Resource (Tool Resource – "GSS")*

Desired Output 1.3: Determine business success criteria

- Specification of Business Success Criteria: *Organizational Agent, Role, Resource (Tool Resource – "Requirements Analysis Tool, GSS Type Tools")*
- Identification of personnel who will assess the criteria: *Organizational agent, Role*

Task 2: Assess situation

Desired Output 2.1: Compile inventory of resources

- Identify the base hardware: *Resource (Tool Resource)*
- Establish the availability of hardware: *Organizational Agent, Role*, *Resource (Tool Resource)*
- Determine if schedule conflicts are present: *Organizational Agent, Role, Resource (Tool)*
- Identify hardware for DM tool: *Resource (Tool Resource)*
- Identify data sources: *Resource (Data Resource)*
- Documentation of types of data sources: *Organizational Agent, Role, Resource (Data Resource)*
- Identify knowledge sources and their type: *Resource (Knowledge Resource)*
- Check available tools and techniques: *Activity, Resource (Tool Resource)*
- Describe relevant background knowledge: *Resource (Knowledge Resource)*
- Identification of system administrator, database administrator etc: *Organizational Agent, Role*
- Identification of market analysts, DM experts etc: *Organizational Agent, Role*
- Checking availability of domain experts for later phases: *Organizational Agent, Role*

Desired Output 2.2: List requirements, assumptions and constraints

- Specify target profile: *Customer, Resource Ontology (Data)*
- Capture requirements on scheduling: *Organization Agent Role, Resource Ontology (Tools – Requirements Analysis Tool)*
- Capture requirements on comprehensibility, maintainability etc: *Organization Agent, Role, Resource Ontology (Tools – Requirements Analysis Tool)*
- Capture requirements on security, legal restrictions, etc: *Organization Agent, Role, Resource (Tool Resource – "Requirements Analysis Tool")*
- Clarify all assumptions: *Organization Agent, Role, Resource (Tool Resource - "Cognitive Maps, GSS Type Tools")*
- List assumptions on data quality: *Organization Agent, Role, Resource (Data), Resource (Information Resource – "Metadata Repository")*
- List assumptions on external factors: *Organization Agent, Role, Resource (Tool*

Resource – "Porter's Five Forces and SWOT")

- Clarify assumptions leading to estimates: *Organization Agent, Role, Resource (Tool Resource – "GSS Type Tools")*
- Assumptions related to presentation of results: *Organization Agent, Role, Resource (Tool Resource – "Cognitive Maps"*
- Check general constraints: *Organization Agent, Role, Rules, Resource (Tool Resource – "Requirements Analysis Tool")*
- Check access right to data sources: *Organization Agent, Role, Authority*
- Check technical accessibility of data: *Resource (Data Resource)*
- Check if relevant knowledge is accessible: *Organization Agent, Role, Authority, Resource (Knowledge Resource)*
- Check budget constraints: *Organization Agent, Role, Resource (Tool Resource – "Requirements Analysis Tool")*

Desired Output 2.3: Identify risks and contingencies

- Identify risks to the project: *Activity, Risk*
- Develop contingency plans for each risk: *Risk, Contingency Plan*

Desired Output 2.4: Create a glossary of business and DM terminology

- Creation of glossaries: *Resource (Knowledge Resource – "Business Metadata Base")*
- Understand terminology of domain experts: *Organization Agent, Role, Resource (Knowledge Resource – "Business Metadata Base), Resource (Tool Resource – "Concept Map")*
- Understand business terminology: *Organization Agent, Role, Authority, Resource (Knowledge Resource – "Business Metadata Base")*

Desired Output 2.5: Conduct a cost benefit analysis

- Estimate costs of data collection: *Division, Organizational Agent. Role, Resource (Tool Resource – "Project Management Cost Estimation Type Tool")*
- Estimate cost of developing and implementing a solution: *Division, Organizational Agent, Role, Resource (Tool Resource – "Project Management Cost Estimation Type Tool")*
- Identify benefits of implementing the DM project: *Organizational Agent. Role, Resource (Tool Resource – "Project Management Benefits Estimation Type Tool")*
- Estimate operating costs: *Division, Organizational agent, Role, Resource (Tool Resource – Project Management Cost Estimation Type Tools)*
- Identify hidden costs: *Resource Ontology (Tool Resource – "Project Management Cost Estimation Type Tool")*

Task 3: Determine DM Goals

Desired Output 3.1: Determine DM goals

- Translate the business question(s) to DM goal(s): *Organization agent, Role. Resource (Tool Resource –" GSS Type Tool")*
- Specify DM problem type: *Organization agent, Role. Resource (Tool Resource – "GSS Type Tool")*

Desired Output 3.2: Determine DM success criteria

- Specify criteria for model assessment: *Organization agent, Role, Resource (Tool Resource – "Requirements Analysis Tool, GSS Type Tool")*
- Define benchmarks for evaluation criteria: *Organization agent, Role, Resource (Tool Resource – "Requirements Analysis Tool, AHP, GSS Type Tool")*
- Specify subjective assessment criteria: *Organization agent, Role, Resource (Tool Resource – "Requirements Analysis, GSS Type Tool"))*

Task 4: Produce a Project Plan

Desired Output 4.1: Formulation of project plan

- Define the initial process plan: *Organization agent, Role, Resource (Tool Resource – "Project Management Work Breakdown Structure Tools, Gantt Chart, PERT, Project Charter")*
- Develop coherent procedure to solve business questions: *Organization agent, Role, Resource (Tool Resource –"Work Breakdown Structuring Tool")*
- Estimate effort and resources needed to deploy the solution: *Organization agent, Role, Resource (Tool Resource – "Project*

Management Planning and Management Tools, Work Breakdown Structuring Tool")

- Identify critical steps: *Organization agent, Role, Resource (Tool Resource – "Critical Path Analysis Tool")*
- Mark decision points and review points: *Organization agent, Role, Resource (Tool Resource – "Project Management Planning and Management Tools, Milestone Charts")*
- Identify major iterations: *Organization agent, Role, Resource (Tool Resource – "GSS Type Tool")*

Desired Output 4.2: Initial assessment of tools and techniques

- List selection criteria for DM tools: *Organization agent, Role, Resource (Tool Resource – "GSS Type Tool")*
- Choose potential tools and techniques: *Organization agent, Role, Resource (Tool Resource – "GSS Type Tool")*
- Evaluate appropriateness of techniques: *Organization agent, Role, Resource (Tool Resource – "GSS Type Tool")*
- Prioritize applicable techniques: *Organization agent, Role, Resource Ontology (Tool Resource – "GSS Type Tools")*

4. Summary

This paper presents an organization-ontology based framework to execute the various activities prescribed in the BU phase of the CRISP-DM methodology. The proposed solution achieves our concurrent objectives of provision of a framework that incorporates tools and techniques for executing the various activities of the BU phase and ensuring that the output of the activities is generated in a form that can allow for at least their semi-automated linkage with activities of this phase and with activities of the succeeding phases of the methodology. Future research could include specifying attributes and relations, applicable constraints and formal axioms to test the competency of the proposed organization-ontology. There is also a need for more research into novel ways in which organization ontologies and enterprise models can be leveraged to implement knowledge intensive, mission critical projects.

References

1. CRISP-DM, *Cross Industry Standard Process for Data Mining 1.0: Step by Step Data Mining Guide*. 2003.
2. Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth, *From Data Mining to Knowledge Discovery: An Overview*, in *Advances in Knowledge Discovery and Data Mining*. 1996b, AAAI Press. p. 1-34.
3. Berry, M. and G. Linoff, *Data Mining Techniques for Marketing, Sales and Customer Support*. 1997: John Wiley and Sons.
4. Cabena, P., et al., *Discovering Data Mining: From Concepts to Implementation.* 1998: Prentice Hall.
5. Anand, S. and A. Buchner, *Decision Support Using Data Mining*. 1998, London: Financial Times Pitman Publishers.
6. Cios, K. and L. Kurgan, *Trends in Data Mining and Knowledge Discovery*, in *Advanced Techniques in Knowledge Discovery and Data Mining*, N. Pal and L. Jain, Editors. 2005, Springer. p. 1-26.
7. Berry, M. and G. Linoff, *Mastering Data Mining: The Art and Relationship of Customer Relationship Management*. 2000: John Wiley and Sons
8. Charest, M., et al. *Intelligent Data Mining Assistance via CBR and Ontologies*. in *Proceedings of the 17th International Conference on Database and Expert Systems Applications (DEXA'06)*. 2006.
9. Pyle, D., *Business Modeling and Data Mining*. 2003: Morgan Kaufmann Publishers.
10. Gruber, T.R., *A Translation Approach to Portable Ontology Specifications.* Knowledge Acquisition, 1993. **5**(2): p. 199-220.
11. Fox, M.S., et al., *An Organization Ontology for Enterprise Modeling*, in *Simulating Organizations: Computational Models of Institutions and Groups*. 1998, AAAI/MIT Press: Menlo Park CA. p. 131-152.
12. Hepp, M. and D. Roman. *An Ontology Framework for Semantic Business Process Management*. in *Proceedings of Wirtschaftsinformatik* 2007. Karlsruhe (forthcoming).
13. Bernstein, A., S. Hill, and F. Provost, *Toward Intelligent Assistance for a Data Mining Process: An Ontology-Based Approach for Cost-Sensitive Classification.* IEEE Transactions on Knowledge and Data Engineering 2005. **17**(4): p. 503-518
14. Tham, K.D. and H.M. Kim. *Towards Strategic Intelligence with Ontology Based Enterprise Modelling and ABC*. in *Proceedings of the IBER Conference*. 2002. Las Vegas, NV.
15. Keeney, R.L., *Value-Focused Thinking: Identifying Decision Opportunities and Creating Alternatives.* European Journal of Operations Research, 1996. **92**: p. 537-549.