# Case Study on Requirements Engineering in Information Mining Project: Metallurgical Enterprise

Pollo-Cattaneo, M.F.
PhD Program on Computer Science. Computer Science School. National University of La Plata. Information System Methodologies Research Group. Technological National University. Buenos Aires, Argentina.
flo.pollo@gmail.com

Pesado, P.
School of Computer Science. National University of La Plata. Buenos Aires, Argentina.
ppesado@lidi.info.unlp.edu.ar

Britos, P.
Information Mining Research Group. National. Lab ID + i in Visualization, Computer Graphics and Creative Code. University of Rio Negro at El Bolsón, Río Negro, Argentina.
paobritos@gmail.com

García-Martínez, R.
Information Systems Research Group. National University of Lanus, Buenos Aires, Argentina.
rgarcia@unla.edu.ar

## ABSTRACT
There can be no debate that Information Mining Projects cause processing tools to arise with the sole purpose of converting available organization data into useful knowledge on account of decision-making. Considering the aforementioned, this type of projects demands due diligence in the requirement specification process, as the latter needs to be thorough and traceable throughout the entire project, and therefore, process associated with requirements engineering are not to be reused in any future projects among the like. Similarly, latest methodologies within the field of Information Mining fail to take practices associated to stakeholders and costumers' requirements management into account. With this aim in mind, a solution model to the Information Mining Project Management necessities is proposed.

## CCS Concepts
• **Computing methodologies→Modeling and simulation**
• **Information systems→Information systems applications**

## Keywords
Information Mining, Requirement Management, Process, Methodology, Requirement Engineering.

## 1. INTRODUCTION
The basis of Information Mining is centered on data processing for nontrivial knowledge collection, an organizational task in which analysis and synthesis tools are indispensable [1]. The aforementioned knowledge being unknown, it can be further exploited by organizations on account of the decision-making process [2]. Information Systems specialists claim data relationships, fluctuations and dependencies to be the core of the process, rather than the data itself [3]; whether these relationships reflect the reality and are then regarded as valid for doing so, the fact that they revolutionize the criteria for decision-making cannot be dismissed [4]. Taking the CRISP-DM [5], P3TQ [6], and SEMMA [7] standout Information Mining Methodologies as examples, it can be seen that these all fail to take into account aspects of those related to both project management and the organizational context where the project is taken on, thus failing to engineer the key concepts in business knowledge [8].
All things considered, this paper employs a proposed model which contributes to the development of a thorough requirements management in the context of an Information Mining project. In order to do so, the detected problem is first described (Section 2), and then a solution proposal is suggested (Section 3). Afterwards, a case study in which the proposed process is implemented is then shown (Section 4), and lastly, conclusions and future work lines are presented (Section 5).

## 2. PROBLEM DESCRIPTION
What a proper Information Mining project needs is a due - consistent and traceable the project throughout - requirements specification, which should, at first, allow for an orderly project management, and also leave out any possible requirements misunderstandings [9]. Nonetheless, there is one glaring difference between this type of projects and traditional software projects, as an Information Mining Project demands no software product construction, but none other than the transformation of data into knowledge: a mere process. For this reason, it is clear that requirements for this type of projects do not abide by any definitions of restrictions and/or functionalities of those which the software product does have to fulfill in the field of Software Engineering [10].
At the origination of an Information Mining project, its objectives describing customer's general necessities - what he wants to get as a final result of the project, generally linked to strategic and tactical business goals [11] - must be identified. This is because by applying this field's algorithms to available data, the latter is transformed into knowledge in order to accomplish all sorts of objectives. What is more, guessing out the organization's real expectations is key to obtaining the desired final project [12].
So as to attain a clearer overview of the project, all parties taking part in it need to manage the same vocabulary [13]. Once the project's aims have been identified, an initial survey on available sources of information has to be carried out. Based upon an analysis of the project objective and information sources, the scope of the project can be defined, thereby obtaining a group of particular objectives. These may be achieved with the aid of Information Mining-based algorithmic processes [14]. Thus, the business problem - the mere motive of the entire project - can then be solved with methodologies other than those originated in Software Engineering, in that they overlook practical aspects of the characteristically Information Mining requirements specification [15].
Latest methodologies - CRISP-D, P3TQ and SEMMA - are centered on Knowledge Discovery in Databases Process (KDD) and emphasize available data detection, together with a simplified

overall vision of where the project develops [16], as stakeholder and customer requirements-associated activities are left apart [17]. The model applied in this paper elaborates on the Process Model for Information Mining projects from [18], which is based on the CRISP-DM methodology [19], considering Small and Medium Enterprises (SME) aspects. Despite this methodology including phase-spread activities of CRISP-DM, especially "Business Comprehension" and "Data Comprehension", COMPETISOFT details none by only indicating what are the techniques to be applied in any activity [20]. Similarly, Requirements Engineering documenting templates defined in [13] are implemented in this paper.

## 3. PROPOSED SOLUTION MODEL

A proposed process model is split into five orderly phases: "Project Definition", "Business Process Engineering", "Business Process Data Engineering", "Business Conceptualization" and "Information Mining Process Specification".

"Project Definition" phase aims to define the project scope, stakeholders, and objectives to accomplish. The "Business Process Engineering" phase seeks to identify and survey the most relevant business processes in the project. The purpose of "Business Process Engineering" is to locate data repositories where information of the various business processes is stored and to survey their contents. The phase of "Business Conceptualization", attempts to define business in terms of concepts developed and vocabulary managed in order to understand the business jargon, therefore revealing business technical words' meanings coined in the business context. Finally, "Information Mining Process Specification" phase intends to identify Information Mining Processes available for use in solving business processes problems, prior to developing the planification of the remaining project activities. A description of the proposed model phases and activities is displayed at [21].

## 4. CASE STUDY

In this section, the results obtained applying the proposed process model to a case study, are offered. Such case study uncovers behavior patterns allowing the description of trailers used as resources in the production area of a metallurgical enterprise. These patterns will be taken under consideration for decision making when planning the assembly line of units. Following, the products obtained in the five phases of the process are shown (for clarity generated graphics and application models of each phase they are also represented in [21]).

### 4.1 Applying the First Phase of Project Definition

The following activities are described: "Identifying Project Objectives" (Figure 1), "Identifying Project Stakeholders" (Figure 2), and "Identifying the Project Scope" (Figure 3).

| Activity: Identifying Project Objectives. |
|---|
| Input: Stakeholder meetings information. |
| Process:<br>1. Documentation analysis to detect project objectives.<br>2. Generation of a functional decomposition tree (figure 4).<br>3. Correspondence between function decomposition tree and objective documentation templates (figure 5). |
| Output: Project and Requirement Objectives Templates (figure 6). |

**Figure 1. Identifying Project Objectives Activity**

| Activity: Identifying Project Stakeholders. |
|---|
| Input: Stakeholder meetings information. |
| Process:<br>1. Documentation analysis for project stakeholders detection and registration in template.<br>2. Documentation analysis to detect success criteria and project expectations.<br>3. Success criteria "AND/OR" graph development (figure 7).<br>4. Expectation knowledge map development (figure 8).<br>5. and Success Criteria Documentation Template (figure 9) and correspondence between expectations knowledge map and Expectations Documentation Template (figure 10). |
| Output: Project Stakeholders, Project Success Criteria and Project Expectations Templates (figure 11). |

**Figure 2. Identifying Project Stakeholders Activity**

| Activity: Identifying the Project Scope. |
|---|
| Input: Stakeholder meetings information. |
| Process:<br>1. Documentation analysis to define work boundaries in the project, detection of business problems and generation of the Project Scope Definition Template.<br>2. Documentation analysis to define project assumptions and information restrictions.<br>3. Development of a Project and Requirement assumptions conceptual map (figure 12 shows an example of conceptual map of project assumptions).<br>4. Develop to knowledge map of project restrictions and requirement restrictions (by means of example, figure 13 shows a knowledge map for project restrictions).<br>5. Correspondence establishment between the assumptions conceptual map and the assumptions documentation template (figure 14) and between a restrictions knowledge map and the restrictions documentation template (figure 15).<br>6. Documentation analysis to define project risks and contingency plans associated with each of these risks and generation of risks and contingency plans documentation templates. |
| Output: Project Scope Definition Template, Project Restrictions Template, Project Assumptions Template, Requirement Restrictions Template, Requirement Assumptions Template, Project Risks Template, Project Contingency Plans Template, Requirement Risks Template and Requirement Contingency Plans Template (figure 16 shows a project objective template example). |

**Figure 3. Identifying Project Scope Activity**

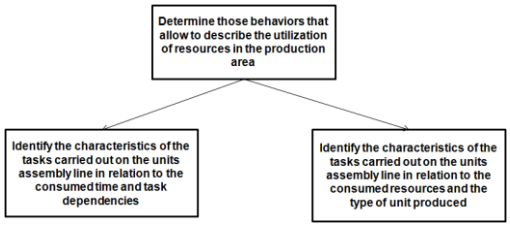



**Figure 4. Functional Decomposition Tree**

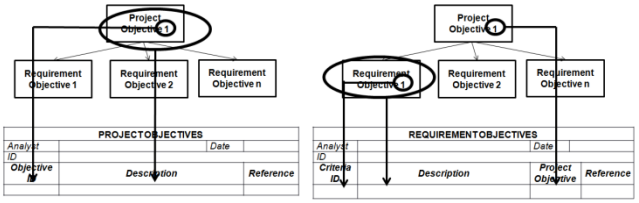


**Figure 5. Correspondence between a Functional Decomposition Tree and Objective Documentation Templates**

| PROJECT OBJECTIVES | | | |
|---|---|---|---|
| *Analyst* | María Florencia Pollo | *Date* | 06/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | |
| ***Objective ID*** | ***Description*** | | ***Reference*** |
| 1 | Determine those behaviors that allow to describe the utilization of resources in the production area. | | Minute of Meeting 1 |

| REQUIREMENT OBJECTIVES | | | | |
|---|---|---|---|---|
| *Analyst* | María Florencia Pollo | *Date* | 06/10/2014 | |
| *ID* | RESOURCES-PRODUCTION | | | |
| ***Criteria ID*** | ***Description*** | ***Project Objective*** | ***Reference*** | |
| 1 | Identify the characteristics of the tasks carried out on the units assembly line in relation to the consumed time and task dependencies. | 1 | Minute of Meeting 1 | |
| 2 | Identify the characteristics of the tasks carried out on the units assembly line in relation to the consumed resources and the type of unit produced. | 1 | Minute of Meeting 1 | |

**Figure 6. Output Products of the "Identifying Project Objectives" Activity**

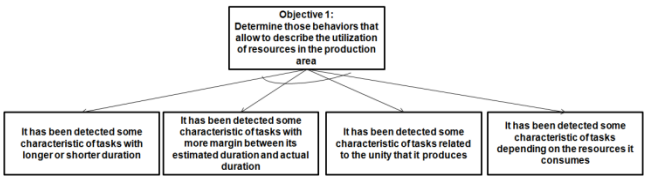



**Figure 7. Success Criteria "AND/OR" Graphs**

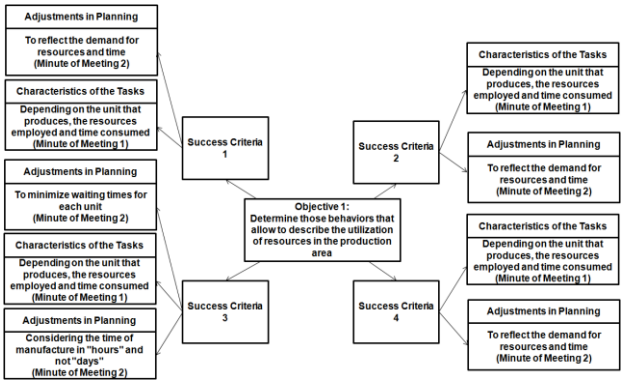



**Figure 8. Expectation Knowledge Map**

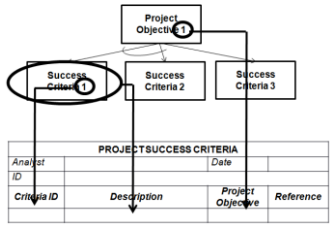



**Figure 9. Correspondence between Success Criteria AND/OR Graph and Success Criteria Documentation Template**

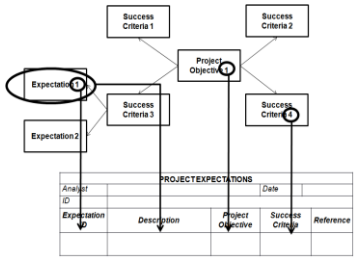



**Figure 10. Correspondence between an Expectations Knowledge Map and the Expectations Documentation Template**

| REQUIREMENT OBJECTIVES | | | |
|---|---|---|---|
| *Analyst* | María Florencia Pollo | *Date* | 06/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | |
| ***Criteria ID*** | ***Description*** | ***Project Objective*** | ***Reference*** |
| 1 | Identify the characteristics of the tasks carried out on the units assembly line in relation to the consumed time and task dependencies. | 1 | Minute of Meeting 1 |
| 2 | Identify the characteristics of the tasks carried out on the units assembly line in relation to the consumed resources and the type of unit produced. | 1 | Minute of Meeting 1 |

| PROJECT STAKEHOLDERS | | | | | | |
|---|---|---|---|---|---|---|
| *Analyst* | María Florencia Pollo | | | *Date* | 06/10/2014 | |
| *ID* | RESOURCES-PRODUCTION | | | | | |
| ***Posición*** | ***Organization/ Area*** | ***Role in the Project*** | ***Associated Business Process*** | ***Contact Information*** | | |
| | | | | ***Name*** | ***Email*** | ***Phone*** |
| General Manager | Metal S.R.L. / General Management | Sponsor | --- | Santiago Lambs | slambs@ metal.com | 4234-0056 |
| Production Manager | Metal S.R.L. / Production | Stakeholder | Units Production | Lucas Martis | lmartis@ metal.com | 4880-8890 |
| Production Planning Responsible | Metal S.R.L. / Production | Stakeholder | Production Planning | José Martis | jmartis@ metal.com | 4880-5467 |
| Informatic System Responsible | Swinter S.R.L. / Systems | Stakeholder | --- | Diego Perez | dperez@ swinter.com | 4999-7766 |

| PROJECT SUCCESS CRITERIA | | | |
|---|---|---|---|
| *Analyst* | María Florencia Pollo | *Date* | 09/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | |
| ***Criteria ID*** | ***Description*** | ***Project Objective*** | ***Reference*** |
| 1 | It has been detected some characteristic of tasks with longer or shorter duration. | 1 | Minute of Meeting 2 |
| 2 | It has been detected some characteristic of tasks with more margin between its estimated duration and actual duration. | 1 | Minute of Meeting 2 |
| 3 | It has been detected some characteristic of tasks related to the unity that it produces. | 1 | Minute of Meeting 2 |
| 4 | It has been detected some characteristic of tasks depending on the resources it consumes. | 1 | Minute of Meeting 2 |

| PROJECT EXPECTATIONS | | | | |
|---|---|---|---|---|
| *Analyst* | María Florencia Pollo | | *Date* | 09/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | | |
| ***Expectation ID*** | ***Description*** | ***Project Objective*** | ***Success Criteria*** | ***Reference*** |
| 1 | Find the characteristics of the tasks carried out on the assembly line depending on the unit that produces, the resources employed and time consumed. | 1 | 1, 2, 3, 4 | Minute of Meeting 1 |
| 2 | Propose adjustments in planning to reflect more accurately the demand for resources and time each. | 1 | 1, 2, 4 | Minute of Meeting 2 |
| 3 | Propose adjustments in planning to minimize waiting times for each unit to move from one sector to another. | 1 | 3 | Minute of Meeting 2 |
| 4 | Propose adjustments in planning considering the time of manufacture of the units in "hours" and not "days" as it is currently done. | 1 | 3 | Minute of Meeting 2 |

**Figure 11. Output Products of the " Identifying Project Stakeholders" Activity**

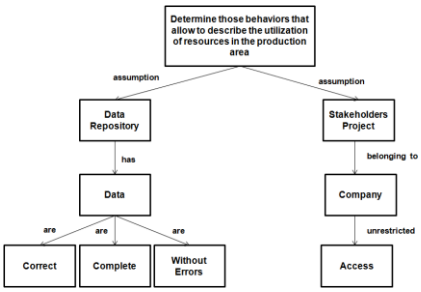



**Figure 12. Conceptual Map of Project Assumptions**

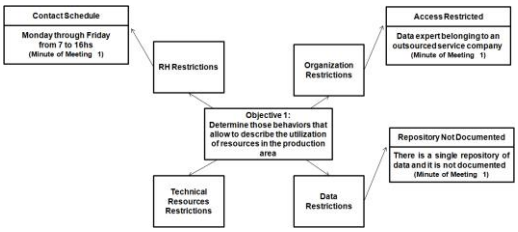



**Figure 13. Knowledge Map of Project Restrictions**

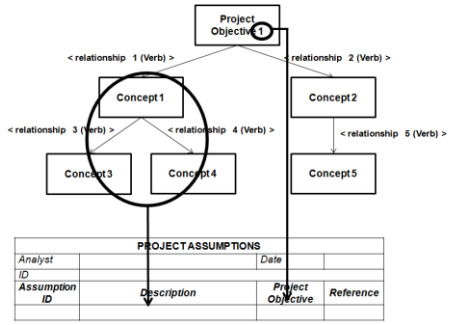



**Figure 14. Correspondence between Conceptual Map of Project Assumptions and the Assumptions Documentation Template**

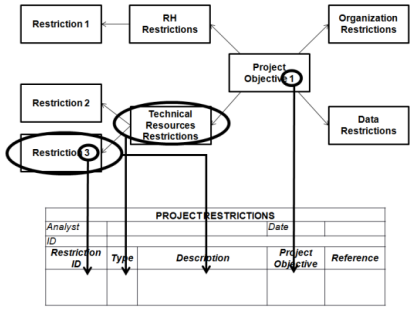



**Figure 15. Correspondence between a Knowledge Map of Project Restrictions and the Restrictions Documentation Template**

| PROJECT SCOPE DEFINITION | | | |
|---|---|---|---|
| *Analyst* | María Florencia Pollo | *Date* | 10/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | |
| ***ID*** | ***Business Objectives Included in the Project*** | | ***Project Objective*** |
| 1 | Improve assembly line production units in the area, making it more effective and efficient, taking into account the characteristics of the tasks carried out in relation to the times consuming and dependencies between tasks (Minute of Meeting 1). | | 1 |
| 2 | Improve assembly line production units in the area, making it more effective and efficient, taking into account the characteristics of the tasks carried out in relation to the resources-consuming and type of unit that occurs (Minute of Meeting 1). | | 1 |
| ***ID*** | ***Business Objectives Excluded of the Project*** | | |
| 1 | Project is excluded from the tasks of the production area not belonging to the assembly line units this area (Minute of Meeting 1). | | |

| PROJECT ASSUMPTIONS | | | |
|---|---|---|---|
| *Analyst* | María Florencia Pollo | *Date* | 10/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | |
| ***Assumption ID*** | ***Description*** | ***Project Objective*** | ***Reference*** |
| 1 | The data stored in the data repository are correct, complete, without errors. | 1 | Minute of Meeting 1 |
| 2 | You have unrestricted access to project stakeholders belonging to the company. | 1 | Minute of Meeting 1 |

| PROJECT RESTRICTIONS | | | | |
|---|---|---|---|---|
| *Analyst* | María Florencia Pollo | | *Date* | 10/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | | |
| ***Restriction ID*** | ***Type*** | ***Description*** | ***Project Objective*** | ***Reference*** |
| 1 | Data | There is a single repository of data and it is not documented. | 1 | Minute of Meeting 1 |
| 2 | Organization | It has a data expert responsible for the computer system but is an outsourced service company, so access to it is restricted. | 1 | Minute of Meeting 1 |
| 3 | HR | Only possible contact persons related to the production area in working hours (ie, Monday through Friday from 7 to 16hs). | 1 | Minute of Meeting 1 |

**Figure 16.a. Output Products of "Identifying the Project Scope" Activity**

| PROJECT RISKS | | | |
|---|---|---|---|
| *Analyst* | María Florencia Pollo | *Date* | 13/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | |
| ***Risk ID*** | ***Description*** | ***Project Objective*** | ***Reference*** |
| 1 | Difficulties and/or delays due to lack of knowledge and experience in Information Mining team. | 1 | Minute of Meeting 3 |
| 2 | It is not possible to determine behaviors that help describe the use of resources because the available data are not sufficiently current or representative as there are units that are manufactured very sporadically. | 1 | Minute of Meeting 3 |
| 3 | It is not possible to determine behaviors that help describe the use of resources because there are tasks whose implementation changes depending on the form of work on each unit. | 1 | Minute of Meeting 3 |
| 4 | It is not possible to determine behaviors that help describe the use of resources because data on plant employees are not current enough and involve people who are no longer working at the company. | 1 | Minute of Meeting 3 |

| PROJECT RISKS CONTINGENCY PLAN | | | |
|---|---|---|---|
| *Analyst* | María Florencia Pollo | *Date* | 13/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | |
| ***Action ID*** | ***Description*** | ***Risk*** | ***Reference*** |
| 1 | Conduct training staff on projects Operational Information. | 1 | Minute of Meeting 3 |
| 2 | Conduct interviews with project stakeholders to identify current business data. | 2,3,4 | Minute of Meeting 3 |
| 3 | Filter data to work only with the current business data. | 2,3,4 | Minute of Meeting 3 |
| 4 | Consider for each task the unit that is occurring. | 2,3,4 | Minute of Meeting 3 |
| 5 | Conduct an analysis involving production data standards that are not performed frequently but are considered important. | 2,3,4 | Minute of Meeting 3 |

**Figure 16.b. Output Products of "Identifying the Project Scope" Activity**

## 4.2 Applying the Second Phase of Business Process Engineering

The following activities are described: "Identifying Business Processes" (figure 17) and "Surveying Business Processes" (figure 18).

| |
|---|
| Activity: Identifying Business Processes. |
| Input: Project Scope Definition Template, Project Objective Template, Requirement Objective Template, Project Success Criteria Template and Project Expectations Template. |
| Process:<br>1. Documentation analysis to define business activities related to the project objectives and generation of the business processes diagram template.<br>2. Generation of the column of business process, associated with each stakeholder in the project stakeholder template (defined in the first phase of the process). |
| Output: Business Process Diagram Template (figure 19). |

**Figure 17. Identifying Business Processes Activity**

| |
|---|
| Activity: Surveying Business Processes. |
| Input: Business Process Diagram Template. |
| Process:<br>1. Conduction of interviews with stakeholders to understand each business process and to generate a business process template per process. |
| Output: Business Process Templates. The business processes defined are two: "Units Production" and "Production Planning" (figure 20). |

**Figure 18. Surveying Business Processes Activity**

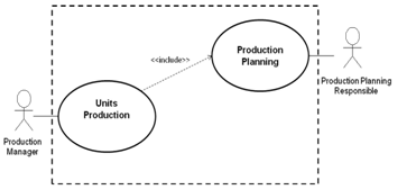
| BUSINESS PROCESSES DIAGRAM | | | | |
|---|---|---|---|---|
| *Analyst* | Maria Florencia Pollo | | *Date* | 16/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | | |
| ***Businesses Project List*** | | | | |
| ***Process Identifier*** | ***Process Name*** | ***Short Description*** | ***Responsible*** | ***Related Project Objective*** |
| PU001 | Production Planning | The aim of the process is to perform planning activities related to the production units. | Production Planning Responsible | 1 |
| PU002 | Units Production | The aim of the process is to run all the activities planned to produce units and store data feed for each action. | Production Manager | 1 |
| ***Main Diagram*** | | | | |
|  | | | | |

**Figure 19. Output Products of "Identifying Business Processes" Activity**

| BUSINESS PROCESSES | | | |
|---|---|---|---|
| **Production Planning** | | | |
| *Analyst* | Maria Florencia Pollo | *Date* | 21/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | |
| *Process Identifier* | PU001 | *Process Short Name* | PRODUCT-UNID001 |
| **Short Process Description** | | | |
| The aim of the process is to perform planning activities related to the production units. | | | |
| **Users Involved** | | | |
| ***ID*** | ***Role Involved*** | ***Responsibility in the Process*** | |
| 1 | Production Manager | It is responsible for monitoring the planning and assign personnel to the planned tasks. | |
| 2 | Production Planning Responsible | It is responsible for conducting the planning of each unit is applied to the production area. | |
| ***Preconditions*** | | | |
| ***ID*** | ***Precondition*** | | |
| 1 | The seller has a "purchase specification", with detailed customer data and the order placed (including Delivery Date agreed). | | |
| ***Step*** | ***Description*** | | |
| 1 | Responsible for planning assigned a lot and performs the "Registration Production Order", where you enter the model specifications and version, technical specifications, doors, floor, accessories, paintings and observations. | | |
| 2 | The system suggests different standards from the input data. | | |
| 3 | Responsible for planning selects the unit according to the required standard. | | |
| 4 | The system displays the sectors and the tasks for the selected standard. | | |
| 5 | Responsible for planning and materials Dates assigned to each task, some task can remove or add new tasks own of the requested unit. | | |
| 6 | The system asks for confirmation of the planning done and sends such planning the production manager. | | |
| 7 | The head of production controls planning tasks and assigns to different plant employees specialized in these activities. | | |
| 8 | The system asks for confirmation of your input and generates a production order. | | |
| ***Post conditions*** | | | |
| ***ID*** | ***Post condition*** | | |
| 1 | The planning was done and can start with production plant. | | |

| BUSINESS PROCESSES | | | |
|---|---|---|---|
| **Units Production** | | | |
| *Analyst* | Maria Florencia Pollo | *Date* | 21/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | |
| *Process Identifier* | PU002 | *Process Short Name* | PRODUCT-UNID002 |
| **Short Process Description** | | | |
| The aim of the process is to run the activities planned for the production of units requested by customers and make a record of the progress of activities undertaken to control manufacturing times of each unit. | | | |
| **Users Involved** | | | |
| ***ID*** | ***Role Involved*** | ***Responsibility in the Process*** | |
| 1 | Production Manager | It is responsible for monitoring the progress of activities. | |
| 2 | Employee | It is responsible for performing the assigned activities (according to production planning) and load date of commencement and completion thereof. | |
| ***Preconditions*** | | | |
| ***ID*** | ***Precondition*** | | |
| 1 | Receiving the unit in progress. | | |
| ***Step*** | ***Description*** | | |
| 1 | The plant employee enters the order number given to the unit to be built. | | |
| 2 | The system displays the planning of each task. | | |
| 3 | The plant employee Data enters the start of the assignment. | | |
| 4 | The plant employee performs the task and enter the Termination Date. In the case of having to perform rework involving running longer the same, the employee must addit as a new task. You can add any comments if required. | | |
| 5 | The system asks for confirmation by the employee plant. In the case of not being correct, it returns to the previous step. | | |
| ***Post conditions*** | | | |
| ***ID*** | ***Post condition*** | | |
| 1 | The task assigned was done, so the unit can continue the production process. | | |

**Figure 20. Output Product of "Surveying Business Processes" Activity**

## 4.3 Applying the Third Phase of Business Process Data Engineering

The following activities are described: "Identifying Data Repositories" (figure 21) and "Surveying Data Repositories" (figure 22).

| |
|---|
| Activity: Identifying Data Repositories. |
| Input: Business Process Diagram Template, templates corresponding to every business process and Stakeholders interviews information. |
| Process:<br>1. Documentation analysis to define the data repositories used or consulted by each business process and generation of the data repositories template. |
| Output: Data Repositories Template (figure 23). |

**Figure 21. Identifying Data Repositories Activity**

| |
|---|
| Activity: Surveying Data Repositories. |
| Input: Data Repositories Template. |
| Process:<br>1. Documentation analysis to define the data repositories structure. |
| Output: Data Structure Template (figure 24). |

**Figure 22. Surveying Data Repositories Activity**

| DATA REPOSITORIES | | | | | |
|---|---|---|---|---|---|
| *Analyst* | Maria Florencia Pollo | | | *Date* | 24/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | | | |
| ***Rep. ID*** | ***Repository Name*** | ***Repository Type*** | ***Description*** | ***Business Processes Affected*** | ***Responsible*** |
| BDPU001 | UNITS_ PRODUCTION | Data Base | Data task tracking, production sectors, customer requests and standards unit. | PU001 PU002 | Metal S.R.L. |

**Figure 23. Output Product of "Surveying Data Repositories" Activity**

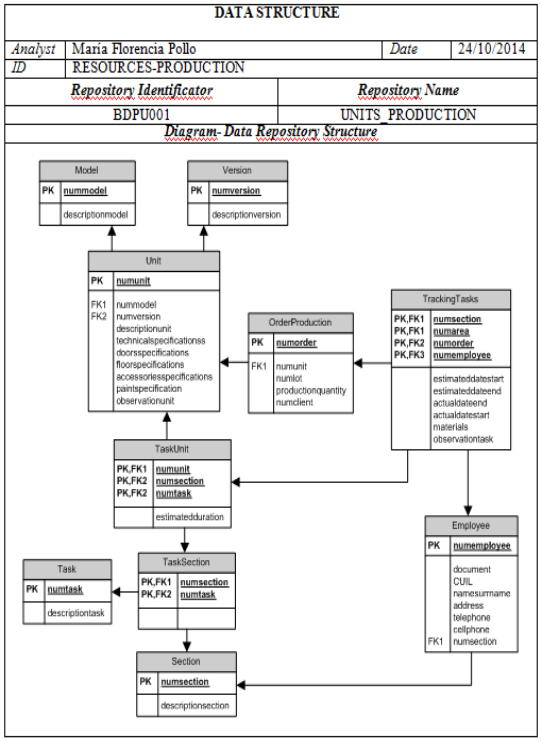



**Figure 24. Output Product of "Surveying Data Repositories" Activity**

## 4.4 Applying the Fourth Phase of Business Conceptualization

The following activities are described: "Developing a Business Dictionary" (figure 25) and "Developing a Business Model" (figure 26).

| |
|---|
| Activity: Developing a Business Dictionary. |
| Input: Business Process Diagram Template, templates corresponding to every business process, Data Repositories Template and Data Structure Template. |
| Process:<br>1. Documentation analysis to define the business keywords. Upon identification, words are validated by stakeholders and thus the business dictionary is built up.<br>2. Wording glossary build-up, detailing keywords. By means of example, in figure 27 two keywords of the utmost importance to the business domain.<br>3. Building of the Concept-Relation and Concept-Attributes-Value chart (figures 28 and 29, respectively) upon the wording glossary, and generation of the Concept relations graph from the two latter charts (figure 30).<br>4. Correspondence establishment between wording glossary and definitions, acronyms and abbreviations template (figure 31).<br>5. Correspondence establishment among concept-attribute-value chart, wording glossary, and requirement-related attributes template (figure 32). |
| Output: Business Dictionary Template, Template of Abbreviations, Acronyms, and Definitions and Template of Attributes related to Requirements. Figure 33 shows an extract of each template. |

**Figure 25. Developing a Business Dictionary Activity**

| |
|---|
| Activity: Developing a Business Model. |
| Input: Data Repositories Template, Data Structure Template and Business Dictionary Template. |
| Process:<br>1. Documentation analysis to establish the relationships among business processes, data repositories and business words and business model diagram build-up. |
| Output: Business Model Diagram (figure 34). |

**Figure 26. Developing a Business Model Activity**

| Term: | **TASK** |
|---|---|
| Type: | Concept. |
| Description: | Activity carried out by one or more operators in a particular sector. The task has a final length which can span more than one working day. |
| Function: | It is the object of study of the project carried out. |
| Value Type | --- |
| Values Range | --- |
| Synonyms/Acronyms: | Activity. |
| Data Source: | Data base (BDPU001 – UNITS_PRODUCTION). |
| Reference: | Minute of Meeting 4; Minute of Meeting 5; Minute of Meeting 6. |

| Term: | **PRODUCTION ORDER** |
|---|---|
| Type: | Concept. |
| Description: | Order production of a unit (trailer, body or semitrailer) for a given customer. |
| Function: | The order will pass through different areas where different tasks are performed, each carried out by different employees. Each unit produced is associated with an order. However, each standard unit can be produced many times, by different orders. |
| Value Type | --- |
| Values Range | --- |
| Synonyms/Acronyms: | Order. |
| Data Source: | Data base (BDPU001 – UNITS_PRODUCTION). |
| Reference: | Minute of Meeting 4; Minute of Meeting 5; Minute of Meeting 6. |

**Figure 27. Wording Glossary Extract**

| Concepts | Associated Concepts | Relation | Description |
|---|---|---|---|
| Unit | Task | involves | The standard unit includes a set of tasks. |
| Task | Production Order | developed in | The tasks are performed in a specific order production. |
| Unit | Production Order | made in | The standard unit made in a given production order. |

**Figure 28. "Concept-Relationship" Chart**

| Concepts | Attributes | Values |
|---|---|---|
| Production Order | Production Quantity | Numeric (5 digit) |
| | Client Number | Numeric (3 digit) |
| | Lot Number | Numeric (5 digit) |
| Task | Estimated duration | Estimated Date End – Estimated Date Start |
| | Actual duration | Actual Date End – Actual Date Start |
| | Employee | Numeric (2 digit) |
| | Materials | Alphanumeric (maximum 900 characters) |
| | Task Observation | Alphanumeric (maximum 900 characters) |
| | Area | Numeric (2 digit) |
| | Type of Task | Numeric (3 digit) |
| Unit | Accessories Specifications | Alphanumeric (maximum 900 characters) |
| | Paint Specifications | Alphanumeric (maximum 900 characters) |
| | Floor Specifications | Alphanumeric (maximum 900 characters) |
| | Doors Specifications | Alphanumeric (maximum 900 characters) |
| | Technical Specifications | Alphanumeric (maximum 900 characters) |
| | Model | Numeric (2 digit) |
| | Observation Unit | Alphanumeric (maximum 900 characters) |
| | Type of Unit | Numeric (2 digit) |
| | Version | Numeric (2 digit) |

**Figure 29. "Concept-Attribute-Value" Chart**

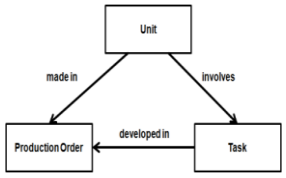



**Figure 30. Graph of Relationships between Concepts**

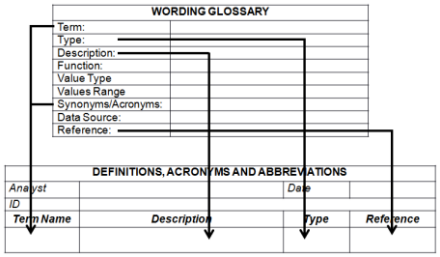



**Figure 31. Correspondence between Wording Glossary and Template of Definitions, Acronyms, and Abbreviations**

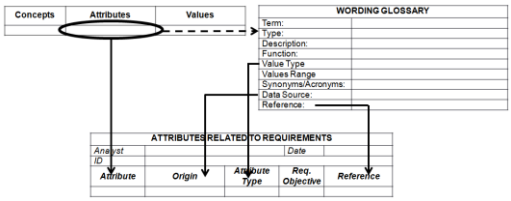



**Figure 32. Correspondence among Concept-Attribute-Value Chart, Wording Glossary, and Template of Attributes related to Requirements**

| BUSINESS DICTIONARY | | | |
|---|---|---|---|
| *Analyst* | María Florencia Pollo | *Date* | 24/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | |
| ***Term Name*** | ***Traceability*** | | |
| | ***Business Process*** | ***Term Utilization*** | |
| Production Order | Units Production | Information display | |
| | Production Planning | Information recording | |
| Task | Units Production | Information display | |
| | Production Planning | Information recording | |
| Unit | Units Production | Information display | |
| | Production Planning | Information recording | |

| DEFINITIONS, ACRONYMS AND ABBREVIATIONS | | | |
|---|---|---|---|
| *Analyst* | María Florencia Pollo | *Date* | 24/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | |
| ***Term Name*** | ***Description*** | ***Type*** | ***Reference*** |
| Production Order | Production order of a unit (trailer, body or semitrailer) for a given customer. | Concept | Minute of Meeting 4<br>Minute of Meeting 5<br>Minute of Meeting 6 |
| Task | Activity carried out by one or more operators in a particular sector. The task has a final length which can span more than one working day. | Concept | Minute of Meeting 4<br>Minute of Meeting 5<br>Minute of Meeting 6 |
| Unit | Product whose characteristics are as close to the requested product. | Concept | Minute of Meeting 5<br>Minute of Meeting 6 |

**Figure 33.a. Output Products of "Developing a Business Dictionary" Activity**

| ATTRIBUTES RELATED TO REQUIREMENTS | | | | |
|---|---|---|---|---|
| *Analyst* | María Florencia Pollo | | *Date* | 24/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | | |
| ***Attribute*** | ***Origin*** | ***Attribute Type*** | ***Req. Objective*** | ***Reference*** |
| Production Quantity | Database (BDPU001 – UNITS_PRODUCTION) | Numeric (5 digit) | 1,2 | Minute of Meeting 4<br>Minute of Meeting 5<br>Minute of Meeting 6 |
| Client Number | Database (BDPU001 – UNITS_PRODUCTION) | Numeric (3 digit) | 1,2 | Minute of Meeting 4<br>Minute of Meeting 5<br>Minute of Meeting 6 |
| Lot Number | Database (BDPU001 – UNITS_PRODUCTION) | Numeric (5 digit) | 1,2 | Minute of Meeting 4<br>Minute of Meeting 5<br>Minute of Meeting 6 |

**Figure 33.b. Output Products of "Developing a Business Dictionary" Activity**

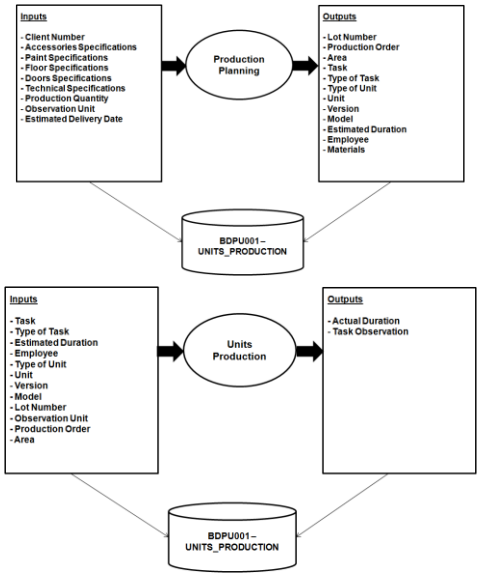



**Figure 34. Output Product of "Developing a Business Model" Activity**

## 4.5 Applying the Fifth Phase of Information Mining Process Specification

The following activities are described: "Formalizing Business Problems" (figure 35), "Identifying Information Mining Processes" (figure 36) and "Developing a Project Plan" (figure 37).

Activity: Standardizing Business Problems.

Input: Project Scope Definition Template, Project Objective Template, Requirement Objective Template and Business Process Diagram.

Process:

1. Analysis of the project objectives, project scope definition, business processes and data repositories and development of a business problems list.
2. If necessary, project objectives and requirement templates are verified and updated in parallel to the former process.

Output: Business Problem Template (figure 38).

**Figure 35. Formalizing Business Problems Activity**

Activity: Identifying Information Mining Processes.

Input: Business Problem Template.

Process:

1. Identification of the Information Mining processes to be applied to solve each of the business problems.
2. If necessary, assumptions, restrictions and requirement-related attributes templates are validated and updated in parallel to the former process.

Output: Information Mining Process Template (figure 39).

**Figure 36. Identifying Information Mining Processes Activity**

Activity: Developing the Project Plan.

Input: Project Scope Definition Template, Project Objective Template, Requirement Objective Template, Project Stakeholder Template, Project Risks Template, Requirement Risks Template and Information Mining Process Template.

Process:

1. Development of the project plan upon analysis of all prefetched documentation.
2. Announcement of the specified plan to business stakeholders.

Output: Project Plan Template (figure 40). Under no circumstances does this template aim to replace or overlap project administration activities, but it does aim to enclose the pivotal information for ulterior stages concretion.

**Figure 37. Developing the Project Plan Activity**

| BUSINESS PROBLEMS | | | | |
|---|---|---|---|---|
| *Analyst* | María Florencia Pollo | | *Date* | 28/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | | |
| ***Problem ID*** | ***Description*** | ***Priority*** | ***Business Processes Affected*** | ***Project Objective*** |
| 1 | Set the improvements to be made on the assembly line units of the production area by identifying dependencies between tasks carried out and the relationship between estimated and actual times that consuming. | High | PU001<br>PU002 | 1 |
| 2 | Set the improvements to be made in assembly line production area units by identifying the main features of the tasks carried out by consuming the resources and the drive type that occurs. | High | PU002 | 1 |

**Figure 38. Output Products of "Standardizing Business Problems" Activity**

| INFORMATION MINING PROCESSES | | | |
|---|---|---|---|
| *Analyst* | María Florencia Pollo | *Date* | 28/10/2014 |
| *ID* | RESOURCES-PRODUCTION | | |
| ***Problem ID*** | ***Description*** | ***Information Mining Processes to Use*** | |
| 1 | Set the improvements to be made on the assembly line units of the production area by identifying dependencies between tasks carried out and the relationship between estimated and actual times that consuming. | Discovery of group behavior rules using "Task Type" as objective attribute.<br><br>Discovery of group behavior rules using "Life Expectancy" as objective attribute.<br><br>Discovery of group behavior rules using "Actual Duration" as objective attribute. | |
| 2 | Set the improvements to be made in assembly line production area units by identifying the main features of the tasks carried out by consuming the resources and the drive type that occurs. | Weighting of behavior rules using "Unit Type" as objective.<br><br>Weighting group membership rules. | |

**Figure 39. Output Product of "Identifying Information Mining Processes" Activity**

| PROJECT PLAN | | | |
|---|---|---|---|
| **Discovery of Behavioral Patterns that Help describe the Use of Resources in the Coupled Production Area** | | | |
| *Project Leader* | María Florencia Pollo | *Date* | 28/10/2014 |
| *ID* | RESOURCES-PRODUCTION | *Project Short Name* | Production Area Resource Utilization |

| PROJECT OBJECTIVES | |
|---|---|
| ***ID*** | ***Objective*** |
| 1 | Determine behaviors that help describe the use of resources in the production area. |

| Business Stakeholders | | | | |
|---|---|---|---|---|
| ***Project Role*** | ***Stakeholder Position*** | ***Name and Surname*** | ***Phone*** | ***Email*** |
| Sponsor | General Manager | Santiago Lambs | 4234-0056 | slambs@metal.com |
| Stakeholder | Production Manager | Lucas Martis | 4880-8890 | lmartis@metal.com |
| Stakeholder | Production Planning Responsible | José Martis | 4880-5467 | jmartis@metal.com |
| Stakeholder | Informatic System Responsible | Diego Perez | 4999-7766 | dperez@swinter.com |

| Scope | |
|---|---|
| ***ID*** | ***Tasks included within the Project*** |
| 1 | Improve assembly line production units in the area, making it more effective and efficient, taking into account the characteristics of the tasks carried out in relation to the times consuming and dependencies between tasks. |
| 2 | Improve assembly line production units in the area, making it more effective and efficient, taking into account the characteristics of the tasks carried out in relation to the resources-consuming and type of unit produced. |

**Figure 40.a. Output Product of "Developing a Project Plan" Activity**

| PROJECT PLAN | |
|---|---|
| **Project Risks** | |
| ***ID*** | ***Risk*** |
| 1 | Difficulties and / or delays due to lack of knowledge and experience in Information Mining team. |
| 2 | It is not possible to determine behaviors that help describe the use of resources because the available data are not sufficiently current or representative as there are units that are manufactured very sporadically. |
| 3 | It is not possible to determine behaviors that help describe the use of resources because there are tasks whose implementation changes depending on the form of work on each unit. |
| 4 | It is not possible to determine behaviors that help describe the use of resources because data on plant employees are not current enough and involve people who are no longer working at the company. |
| **Calendar** | |
| (Project GANTT must be included) | |
| **Tracking and Control Mechanisms** | |
| Weekly reviews will be conducted with the aim of analyzing the progress planned vs. actual progress of the project to implement the actions that make order to avoid delays in the project and resolve any conflicts that arise. | |

**Figure 40.b. Output Product of "Developing a Project Plan" Activity**

# 5. CONCLUSIONS

This paper's main contribution is to provide the community with tools so as to enable them to carry out a thorough management of requirements in the context of an Information Mining Project. Consequently, this paper seeks to improve current methodologies in which this approach has been severely overlooked, which is why the process model proposed was applied to controversial case-studies, covering various business issues. This being the case, both customer requirements and requirements-related project objectives were successfully and efficiently managed, as well as information regarding business processes and their problems, together with the data repositories employed, were obtained. Hence, the development of a project master plan - guiding the project through its very stages - was successfully developed.

As future lines of work, the in-depth analysis and description of structural procedures is heavily suggested, as it is believed that the statement of every document process can lead to a higher rate of automation in these processes' techniques. Moreover, it is considered that a software tool capable of assessing the condition of templates and minutes advised by the process - additionally distinguishing the various versions of a single document, thereby reporting each version's contents together with a change history - is highly desirable project to be substantiated.

# 6. ACKNOWLEDGMENTS

The research presented in this paper has been partially financed by the 11211 EIU TIBA, research project of the National University of Technology - UTN (for its acronym in Spanish) Buenos Aires.

# 7. REFERENCES

[1] Schiefer, J., Jeng, J., Kapoor, S. & Chowdhary, P. 2004. *Process Information Factory: A Data Management Approach for Enhancing Business Process Intelligence*. Proceedings 2004 IEEE International Conference on E-Commerce Technology, 162-169.

[2] Thomsen, E. *BI's Promised Land*. 2003. Intelligent Enterprise, 6(4), 21-25.

[3] Curtis, B., Kellner, M., Over, J. 1992. *Process Modelling*. Communications of the ACM, 35(9), 75-90.

[4] Kanungo, S. 2005. *Using Process Theory to Analyze Direct and Indirect Value-Drivers of Information Systems*. Proceedings of the 38th Annual Hawaii International Conference on System Sciences, pp. 231-240.

[5] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. 2000. *CRISP-DM 1.0 Step-by-step Data Mining Guide*. http://tinyurl.com/crispdm (accessed 24.06.16).

[6] Pyle, D. 2003. *Business Modeling and Business Intelligence*. Morgan Kaufmann Publishers.

[7] SAS. 2008. *SAS Enterprise Miner: SEMMA*. http://tinyurl.com/semmaSAS (accessed 24.06.16).

[8] Vanrell, J., Bertone, R., & García-Martínez, R. 2010. *Modelo de Proceso de Operación para Proyectos de Explotación de Información*. Anales del XVI Congreso Argentino de Ciencias de la Computación, 674-682. ISBN 978-950-9474-49-9.

[9] Britos, P., Fernandez, E., García-Martínez, R. 2007. *Propuesta Metodológica para la Educción de Requisitos en Proyectos de Explotación de Información.* IX Workshop de Investigadores en Ciencias de la Computación, pp. 497-501. ISBN 978-950-763-075-0.

[10] Wiegers, K. 2003. *Software Requirements. Microsoft Press.*

[11] Cao, L., Luo, D., & Zhang, C. 2007. *Knowledge Actionability: Satisfying Technical and Business Interestingness*. International Journal of Business Intelligence and Data Mining, 2(4), 496-514.

[12] Gupta, G. K. 2014. *Introduction to Data Mining with Case Studies.* PHI Learning Pvt. Ltd.

[13] Britos, P., Dieste, O., García-Martínez, R. 2008. *Requirements Elicitation in Data Mining for Business Intelligence Projects*. In Advances in Information Systems Research, Education and Practice. David Avison, George M. Kasper, Barbara Pernici, Isabel Ramos, Dewald Roode Eds. (Boston: Springer), IFIP International Federation for Information, 274: 139–150.

[14] Garcia–Martinez, R., Britos, P., Pollo–Cattaneo, F., Rodriguez, D., Pytel, P. 2011. *Information Mining Processes*

*Based on Intelligent Systems.* Proceedings of II International Congress on Computer Science and Informatics (INFONOR–CHILE 2011), 87–94. ISBN 978–956–7701–03–2.

[15] Rose, K. H. 2013. *A Guide to the Project Management Body of Knowledge (PMBOK® Guide) - Fifth Edition*. Project Management Journal, 44(3), e1-e1.

[16] Cao, L. 2012. *Actionable Knowledge Discovery and Delivery*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(2), 149-163.

[17] García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P., Vanrell. J. 2011. *Towards an Information Mining Engineering.* Software Engineering, Methods, Modeling and Teaching. Editorial Universidad de Medellín, 83-99. ISBN 978-958-8692-32-6.

[18] Vanrell, J., Bertone, R., García-Martínez, R. 2012. *Un Modelo de Procesos para Proyectos de Explotación de Información*. Proceedings Latin American Congress on Requirements Engineering and Software Testing, 46-52. ISBN 978-958-46-0577-1.

[19] Oktaba, H., Garcia, F., Piattini, M., Ruiz, F., Pino y F.J. y Alquicira, C. 2007. *Software Process Improvement: The COMPETISOFT Project Computer*, 2007, 40(10), 21-28.

[20] Mariscal, G., Marbán, Ó., González, Á. L. y Segovia, J. 2007. *Hacia la Ingeniería de Data Mining: Un modelo de Proceso para el Desarrollo de Proyectos*. II Congreso Español de Informática, 139-148.

[21] Pollo-Cattaneo, M.,Pesado, P., Britos, P. & García-Martínez, R. 2016. *Process Model Application for Requirements Engineering in Metallurgical Enterprise Information Mining Project*. Technical Report GEMIS-TD-2016-01-RT-2016-08. http://grupogemis.com.ar/web/gemis-td-2016-01-rt-2016-08/ (accessed 19.08.16).