ELEMENTOS PARA UN MAPA DE ACTIVIDADES PARA PROYECTOS DE EXPLOTACIÓN DE INFORMACIÓN

Alumno

Lic. Federico Carlos Peralta

Directores

Dra. Paola Britos (UNRN) y Mg. Darío Rodríguez (UNLa)

TRABAJO PRESENTADO PARA OBTENER EL GRADO
DE
ESPECIALISTA EN INGENIERÍA DE SISTEMAS DE INFORMACIÓN

ESCUELA DE POSGRADO

FACULTAD REGIONAL BUENOS AIRES

UNIVERSIDAD TECNOLÓGICA NACIONAL

JUNIO, 2013

RESUMEN

La confección del mapa de actividades para un proyecto en particular, permite formalizar aquellas actividades que se van a ejecutar durante el desarrollo del mismo, teniendo en cuenta las características específicas del proyecto. El proceso de construcción del mapa de actividades de un proyecto es una tarea compleja. En base a la información contenida en el mapa, el responsable del proyecto podrá gestionar cuidadosamente cada una de las actividades seleccionadas. En este contexto, se propone la construcción de un marco teórico para avanzar en la desarrollo de un mapa de actividades, tomando como base un Modelo de Procesos para Proyectos de Explotación de Información para PyMEs.

Palabras claves: Explotación de Información – Mapa de Actividades – Modelo de Procesos – Proyecto

ABSTRACT

The preparation of the activity map for a particular project allows to formalize the activities to be executed during its development, according with the specific characteristics of it. The construction of activity map is a complex task. Based on this information, the project manager will be able to manage carefully each of the selected activities. In this context, we propose the construction of a theoretical framework in order to advance in the development of the activity map, based on a Process Model for Data Mining Projects for SEMs.

Key words: Data Mining – Activity Map – Process Model – Project

ÍNDICE

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

NOMENCLATURA

Catalyst	Metodología para el desarrollo de proyectos de Explotación de Información; conocida como P^3TQ: Producto (Product), Lugar (Place), Precio (Price), Tiempo (Time) y Cantidad (Quantity).
CRISP_DM	Metodología para el desarrollo de proyectos de Explotación de Información (Cross Industry Standard Process for Data Mining).
Competisoft	Modelo de Procesos para el desarrollo de Software basado en MoProSoft.
Daimler Chrysler	Empresa creadora de la metodología CRISP-DM.
Datawarehouse	Almacén o bodegas de datos.
KDnuggets	Portal de información sobre la Explotación de Datos y descubrimiento de conocimiento.
MII	Modelo de Negocio definido en la metodología P^3TQ.
MIII	Modelo de Explotación de Información definido en la metodología P^3TQ.
MoProSoft	Modelo de Procesos para la Industria de Software desarrollado en México.
NCR	Empresa creadora de la metodología CRISP-DM.
P^3TQ	Metodología para el desarrollo de proyectos de Explotación de Información: Producto (Product), Lugar (Place), Precio (Price), Tiempo (Time) y Cantidad (Quantity); conocida como Catalyst.
PIN	Problemas de Inteligencia de Negocio.
RRHH	Recursos Humanos.
SAS	Fabricante de Software de Inteligencia Empresarial, creadora de la metodología SEMMA (inglés: SAS Institute).
SPSS	Empresa creadora de la metodología CRISP-DM.
TEI	Técnicas de Explotación de Información.
SEMMA	Metodología para el desarrollo de proyectos de Explotación de Información: Muestreo (Sample), Exploración (Explore), Modificación (Modify), Modelado (Model) y Valoración (Assess).

1. INTRODUCCION

En este Capítulo se plantea el contexto del Trabajo de Especialidad (sección 1.1), se establecen sus objetivos (sección 1.2), y se resume la estructura del mismo (sección 1.3).

1.1. CONTEXTO DEL TRABAJO DE ESPECIALIDAD

Las organizaciones actuales toman sus decisiones, cada vez más, basándose en el conocimiento procedente de los datos almacenados en sus bases o almacenes de datos. Conceptos conocidos como "Explotación de Datos", "Inteligencia de Negocio" y "Gestión del Conocimiento", se están desarrollando a un ritmo muy acelerado [Amón Uribe y Jiménez Ramírez, 2009].

La Explotación de Información es una disciplina que ha mostrado una gran evolución en los últimos años. Las organizaciones han comenzado a analizar y explotar grandes masas de información, residentes en sus sistemas informáticos, con el propósito de obtener nuevos conocimientos a partir de las mismas. Su principal objetivo es descubrir información oculta o implícita, que no es posible conseguir mediante la utilización de métodos estadísticos convencionales [Moine et *al*., 2011].

Existe tanta información en una organización que es indispensable identificar cuál es la información que tiene un mayor impacto en las operaciones del negocio. Año tras año, las empresas pierden muchísimo dinero, su credibilidad falla, la insatisfacción de los clientes se acrecienta y todo ello como consecuencia de la mala calidad de los datos con que toman sus decisiones, o que simplemente, utilizan para llevar a cabo sus operaciones [Vilalta y Espinosa, 2008].

En los últimos años, la diversidad, el número y la complejidad de los proyectos de Explotación de Datos ha aumentado ligeramente, lo que hace que los procesos para el desarrollo de este tipo de proyectos tengan que estandarizarse para lograr resultados que puedan ser integrados, reutilizados e intercambiados en un futuro [Mariscal et *al*., 2007].

Se han ido desarrollando diversas metodologías que posibilitan gestionar la complejidad de proyectos de Explotación de Información de una manera uniforme. Actualmente, la comunidad científica considera ya probadas las metodologías SEMMA, CRISP-DM y P^3TQ [Britos, 2008].

A través del mapa de actividades, se puede adaptar una metodología estándar a un proyecto en particular, considerando las características específicas del mismo [Diez et *al*., 2003].

1.2. OBJETIVOS DEL TRABAJO DE ESPECIALIDAD

1.2.1. Objetivo General

- Construir un estado de la cuestión sobre mapas de actividades para proyectos de Explotación de Información.

1.2.2. Objetivos Específicos

- Identificar y contextualizar la temática objeto de estudio.
- Describir y comparar distintas metodologías existentes que aplican a proyectos de Explotación de Información.
- Establecer los elementos teóricos que respaldan la construcción de un estado de la cuestión sobre mapas de actividades.
- Describir la estructura del Modelo de Procesos desarrollado por Vanrell.

1.3. VISIÓN GENERAL DEL TRABAJO DE ESPECIALIDAD

En el Capítulo 1, se plantea el contexto del Trabajo de Especialidad, se establece el objetivo general y los objetivos específicos del mismo, y se resume la estructura del presente trabajo.

En el Capítulo 2, se realiza una descripción de los mapas de actividades y se presentan una serie de ventajas asociadas a la construcción y aplicación de los mismos para proyectos de Explotación de Información; luego, se presenta una descripción detallada de cada una de las tres metodologías principales que se utilizan para este tipo de proyectos: CRISP-DM, P^3TQ y SEMMA, y se realiza una comparación entre todas ellas; a continuación, se introduce el Modelo de Procesos para proyectos de Explotación de Información desarrollado por Vanrell, efectuando un estudio descriptivo del mismo; y finalmente, se realiza un análisis comparativo entre las tres metodologías expuestas y el modelo de procesos en cuestión.

En el Capítulo 3, se describe de manera general la problemática identificada, razón por la cual se lleva a cabo este trabajo de investigación. Se destaca la utilidad del mapa de actividades para el responsable del proyecto y otros miembros del equipo, lo cual permite inferir la importancia de resolver el problema planteado.

Por último, en el Capítulo 4, se presentan las conclusiones más importantes de este trabajo que constituyen el balance final de esta investigación.

2. ESTADO DEL ARTE

En este Capítulo, se presenta una descripción de los mapas de actividades (sección 2.1) y se mencionan las ventajas asociadas a la construcción y aplicación de los mismos (sección 2.1.1). Luego, se realiza una breve introducción al área de Explotación de Información (sección 2.2). A continuación, se describen las principales metodologías que se utilizan para este tipo de proyectos: CRISP-DM (sección 2.2.1), P³TQ (sección 2.2.2) y SEMMA (sección 2.2.3), realizando un estudio comparativo entre todas ellas (sección 2.2.4). Luego, se efectúa un análisis descriptivo del Modelo de Procesos para proyectos de Explotación de Información desarrollado por Vanrell (sección 2.2.5). Por último, se realiza una comparación entre el modelo en cuestión y las tres metodologías presentadas en este trabajo (2.2.6).

2.1. MAPA DE ACTIVIDADES DE UN PROYECTO

Al comienzo de un proyecto resulta crítica la decisión sobre qué ciclo de vida se elegirá para el proyecto en cuestión. Una vez que el responsable del proyecto ha realizado tal selección, y orientado en cierta medida por ella, debe adaptar el proceso software genérico al modelo de ciclo de vida elegido mediante el establecimiento del mapa de actividades [Britos et *al*., 2006 ; Juristo, 2003].

El éxito de un proyecto dependerá del ciclo de vida seleccionado para llevar a cabo el desarrollo del proyecto en cuestión, ya que puede ayudar a garantizar que se ejecuten los pasos necesarios para alcanzar el objetivo planteado [Mariscal et *al*., 2007].

Según [Juristo, 2003], el mapa de actividades es una tabla donde se detallan todas las actividades que se van a ejecutar para un determinado proyecto. Por tanto, como se puede observar en la *Tabla 2.1*, dicha tabla consta de dos entradas, una de ellas corresponde al proceso software con sus actividades correspondientes, mientras que la entrada restante corresponde al ciclo de vida seleccionado y sus propias etapas o fases.

Existen diversos tipos de proyectos: grandes, medianos, pequeños, etc. Existe una estrecha relación entre el tipo de proyecto y el ciclo de vida. Por ejemplo, un proyecto medio se suele corresponder con un ciclo de vida prototipado, un proyecto pequeño casi siempre se corresponde con un ciclo de vida en cascada, etc. Cualquiera sea el tipo de proyecto a desarrollar, el mapa debe siempre adaptarse a las características concretas del mismo. La confección de mapas de actividades pretende en cierta medida facilitar la labor del responsable del proyecto [Juristo, 2003].

Proceso Software		**Requisitos**	**Diseño**	……
	Actividad 1			
	Actividad 2			
	.			
	.			
	.			
	.			
	Actividad n			

Tabla 2.1. *Mapa de actividades.* Extraído de [Juristo, 2003].

En el trabajo citado en el párrafo anterior, se destaca que cuando se confecciona un mapa de actividades, el responsable de proyecto debe marcar con una cruz las actividades que se llevarán a cabo. Además, se puede volcar en el contenido del mapa información adicional acerca de cada una de ellas. Por ejemplo, se puede hacer referencia a la importancia de la actividad, utilizando el símbolo (+) para las actividades de mucha importancia, y el símbolo (-) para aquellas que tienen una importancia normal. También podemos distinguir el tipo de actividad, utilizando la letra (O) para distinguir las obligatorias, y la letra (C) para las condicionales. Muchas otras características específicas pueden ser incluidas en el mapa con el propósito de ampliar la información que se tiene de cada una de las actividades detalladas en el mismo.

Según [Diez et *al.*, 2003], el responsable del proyecto puede utilizar el mapa de actividades como una guía durante el desarrollo del mismo. Sin embargo, la obligatoriedad en el cumplimiento de cada una de las actividades detalladas en dicho mapa no constituye un requisito excluyente a la hora de llevar a cabo el proyecto.

En el mismo artículo, el autor menciona que es importante destacar que la elección y ejecución de las actividades detalladas en el mapa dependen de la experiencia del responsable del proyecto, de su sentido común o percepción de la realidad por parte del mismo; por tal motivo las actividades pueden ser completadas parcial o totalmente y en muchos casos se puede llegar a prescindir de alguna de ellas.

Si bien el mapa permite orientar al responsable del proyecto, en él no se brindan detalles acerca de cómo se deben poner en práctica las actividades seleccionadas y otras particularidades inherentes a ellas; en consecuencia, la información relacionada a las formas de abordar y ejecutar las mismas es sumamente limitada [Diez, 2003].

El mismo autor señala que a medida que transcurre el proyecto, el mapa de actividades va tomando su forma definitiva. Es importante que se reflejen todos los cambios relevantes que se dan durante el desarrollo del mismo; por tal motivo el mapa puede sufrir la adición de nuevas actividades o la remoción de alguna de ellas. Por consiguiente, el mapa de actividades es un documento que no es estático; por el contrario, el mismo tiene un cierto grado de dinamismo que está sujeto a las variaciones que experimentan las características específicas del proyecto en cuestión a lo largo de su desarrollo.

2.1.1. Ventajas de los Mapas de Actividades

Una de las formas de adaptar una metodología estándar, a un proyecto determinado, es a través de la confección del mapa de actividades para ese proyecto en particular, teniendo en cuenta las características del mismo. Cabe señalar, que el proceso de construcción del mapa de actividades de un proyecto, requiere de una considerable experiencia en la aplicación de metodologías estándares de desarrollo, capacidad analítica y una serie de conocimientos afines en la materia, ya que no es una tarea simple de efectuar, ni tampoco puede ser realizada de forma automática [Diez, 2003].

La confección adecuada del mapa de actividades aporta una serie de ventajas a la hora de llevar adelante el desarrollo de un proyecto, facilitando el cumplimiento efectivo de los objetivos propuestos en el mismo. La elección y realización de actividades que no se ajustan a las características específicas de un proyecto, pueden ocasionar una serie de trastornos durante el ejercicio del mismo [Diez et *al.*, 2003].

De acuerdo al análisis realizado por estos autores, es imprescindible que la construcción del mapa sea realizada al inicio del proyecto. En tal caso, el responsable del proyecto deberá conocer las particularidades del mismo para evitar de este modo volcar información obsoleta o incompleta en el mapa. Disponer de ésta información de manera temprana, permite realizar estimaciones válidas para el proyecto en cuestión. Si la confección del mapa es realizada en forma tardía, la utilidad del mismo se ve reducida, ya que muchas de las actividades requeridas necesarias van a estar en curso e inclusive ya completadas en su totalidad.

Por tanto, siguiendo con el análisis de dichos autores, ellos manifiestan que las estimaciones del proyecto se realizan en base a los datos recolectados hasta el momento de proyectos anteriores, considerando las actividades realizadas en ellos y comparando las mismas con aquellas que se definen en el mapa del proyecto en cuestión. Entre las estimaciones que se pueden realizar por parte del responsable del proyecto, podemos señalar:

- Tiempo de ejecución del mismo.

- Costo económico.
- Cantidad de recursos necesarios.
- Perfiles de RRHH.
- Habilidades para realizar determinadas tareas.
- Esfuerzo demandado.

En ciertas ocasiones, las metodologías estándares no son aceptadas fácilmente por todos los miembros de los equipos. Una correcta confección de un mapa de actividades del proyecto, permite demostrar en cierto modo la flexibilidad que la metodología estándar posee, obteniendo confianza y aceptación de la misma por parte de los miembros de la organización [Diez et *al*., 2003].

A su vez, los autores citados en el párrafo precedente, señalan que el mapa de actividades constituye una guía para cada uno de los miembros del equipo de proyecto, en el cual se detallan qué actividades deberán realizarse durante el avance del mismo y por ende, ellos pueden llegar a conocer qué habilidades deberán poseer para cumplir satisfactoriamente con las mismas.

2.2. EXPLOTACIÓN DE INFORMACION

Un proceso de Explotación de Información permite la extracción de conocimiento no-trivial que reside de manera implícita en los datos disponibles en diversas fuentes de información a partir de la ejecución de un conjunto de tareas que se relacionan lógicamente [Rodríguez et *al*., 2010]. La entrada a dicho proceso está constituida generalmente por registros procedentes de bases de datos operacionales o bien almacenes de datos (*Datawarehouse*) [Moine et *al*., 2011].

Normalmente, no son los datos en sí lo más importante para el responsable de un sistema, o para un especialista, sino el conocimiento que se encuentra contenido en sus fluctuaciones, dependencias y relaciones [Britos, 2008].

La Explotación de Información se basa en la búsqueda de patrones relevantes en un conjunto de datos mediante la utilización de técnicas informáticas (Redes neuronales, Inteligencia Artificial, Algoritmos Genéricos, Sistemas Expertos, etc.), como estadísticas (Regresión, Análisis de Varianza, Análisis de Agrupamiento o *Clustering*, Prueba Chi-Cuadrado, Series de Tiempo, Análisis Discriminante, etc.) [Mendez y Rodriguez, 2009].

En [Britos, 2008], se señala que se han ido desarrollando diversas metodologías que permiten gestionar de una manera uniforme la complejidad de proyectos de Explotación de Información. La comunidad científica considera validadas las metodologías SEMMA, CRISP-DM y P^3TQ.

2.2.1. Metodología CRISP-DM

La metodología CRISP-DM fue creada por el grupo de empresas SPSS, NCR y Daimler Chrysler en el año 2000. Esta metodología constituye la guía de referencia más utilizada en el desarrollo de proyectos de Explotación de Datos [Moine *et al*., 2011].

La metodología CRISP-DM según [Chapman et *al.*, 2000] consiste en un modelo jerárquico de procesos, constituido por un conjunto de tareas organizadas en cuatro niveles de abstracción, que van desde el nivel más general hasta los casos más específicos (*Figura 2.1*).

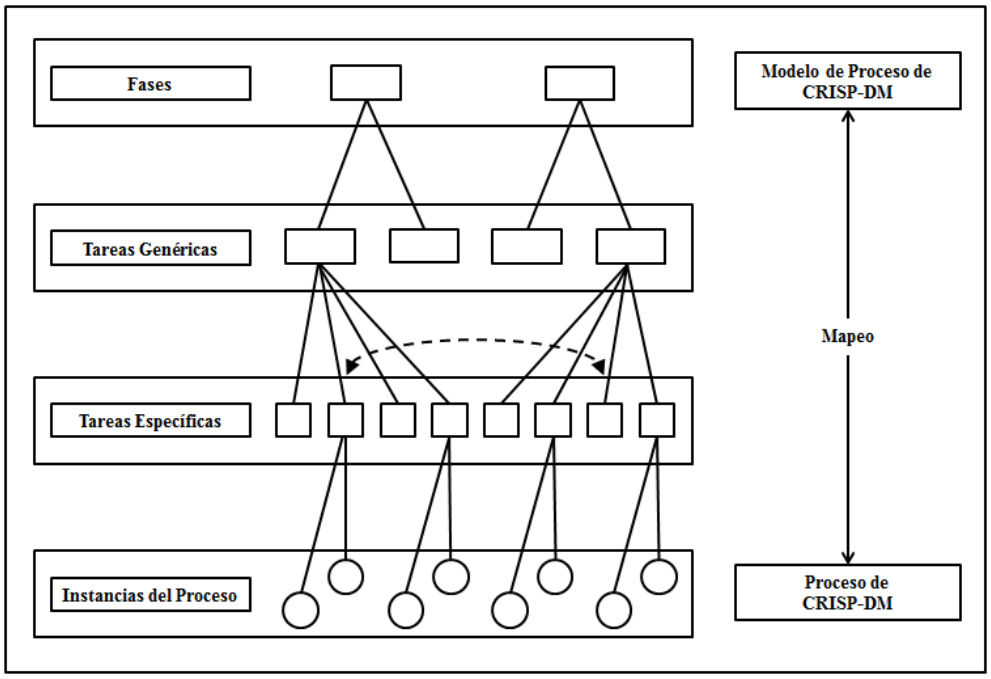

Figura 2.1. *Niveles de abstracción de la metodología CRISP-DM.* Extraído de [Chapman et *al*., 2000].

A nivel más general, CRISP-DM organiza el desarrollo de un proyecto de Explotación de Datos, en una serie de 6 fases, que constituyen el primer nivel de abstracción (ver *Figura 2.3*). Cada una de las fases se encuentra organizada en varias tareas genéricas de segundo nivel o sub-fases. A partir de estas tareas genéricas, se desarrolla el tercer nivel, en el cual se proyectan las tareas especializadas, donde se describen cómo las acciones de las tareas genéricas deben llevarse a cabo en situaciones específicas. El cuarto y último nivel, agrupa el conjunto de acciones, decisiones y resultados sobre el proyecto de Explotación de Información en cuestión [Chapman et *al.*, 2000].

De acuerdo a lo expresado por estos autores, la metodología CRISP-DM aporta al usuario documentación adicional que será utilizada como herramienta de ayuda en el desarrollo del proyecto de Explotación de Información: el modelo de referencia y la guía del usuario.

Continuando con el análisis de [Chapman et *al.*, 2000], el documento del modelo de referencia proporciona información acerca de las fases, tareas generales y salidas de un proyecto de Explotación de Información en general. Por tanto, la guía del usuario brinda información detallada sobre cada fase, proporcionando consejos y listas de comprobación sobre las tareas correspondientes a cada fase. De hecho, describe cómo realizar un proyecto de Explotación de Datos.

Como se puede observar en la *Figura 2.2*, la metodología CRISP-DM estructura el ciclo de vida de un proyecto en seis fases; las flechas indican las relaciones más usuales e importantes entre ellas, aunque se pueden establecer distintas relaciones entre las distintas fases componentes. El círculo exterior simboliza la naturaleza cíclica del modelo de proceso de Explotación de Datos propiamente dicho. La secuenciación de fases no es rígida. Las fases definidas para un proyecto de desarrollo de software clásico (inicio, requerimientos, análisis y diseño, construcción, integración y pruebas y cierre) claramente difieren de las fases propias de esta metodología [Chapman et *al.*, 2000].

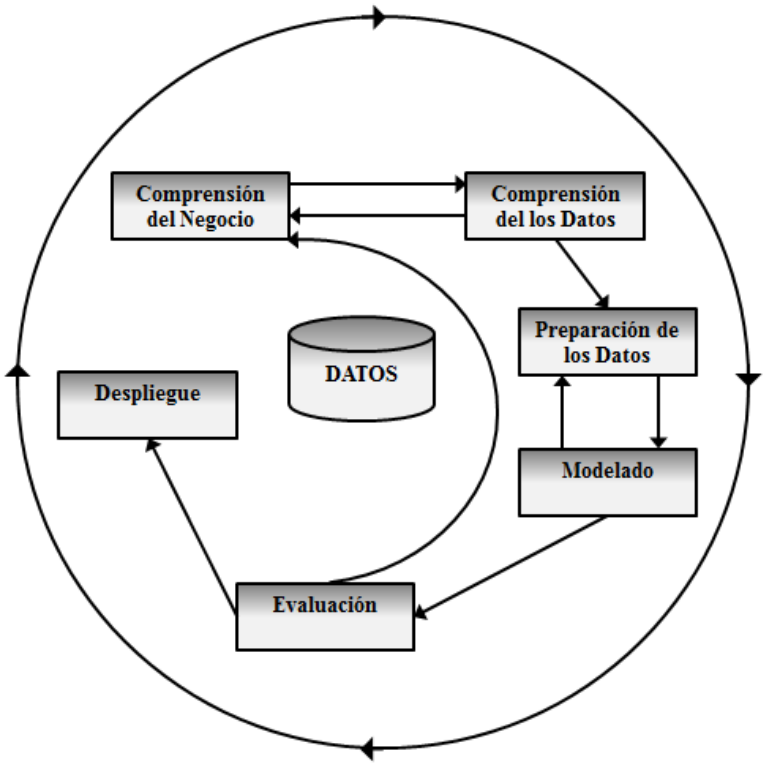

Figura 2.2. *Fases del modelo de proceso de la metodología CRISP-DM.* Extraído de [Chapman et *al.*, 2000].

En [Chapman et *al.*, 2000] se hace referencia a las fases de la metodología CRISP-DM:

- Fase I – "Comprensión del negocio"

Esta primera fase inicial, se basa en el entendimiento de los objetivos del proyecto y la comprensión de los requerimientos del mismo desde el punto de vista del negocio, a fin de definir el problema a

resolver y diseñar una planificación preliminar para el cumplimiento efectivo de los objetivos en cuestión.

- Fase II – "Comprensión de los datos"

La segunda fase de análisis, comienza con la recolección inicial de los datos con el propósito de familiarizarse con los mismos, identificando problemas de calidad asociados a ellos e información adicional relevante para la formulación de las primeras hipótesis.

- Fase III – "Preparación de los datos"

La fase de preparación de los datos, abarca todas aquellas actividades destinadas a la construcción del conjunto de datos finales. Las tareas de esta fase pueden ser ejecutadas varias veces, sin un orden predefinido. Las mismas incluyen la selección de tablas, registros y atributos, así como también la transformación y limpieza de datos para que puedan ser tratados por las herramientas de modelado.

- Fase IV – "Modelado"

En esta fase, se seleccionan y aplican las técnicas de modelado más apropiadas para el proyecto en cuestión, calibrando sus parámetros a valores óptimos. Básicamente, existen varias técnicas para un mismo tipo de problemas en proyectos de Explotación de Datos. Algunas de ellas, demandan requerimientos específicos sobre los datos que se van a procesar, por tal motivo muchas veces es necesario volver a la fase de preparación de los datos antes de avanzar con el modelado de los mismos.

- Fase V – "Evaluación"

Esta fase, involucra la evaluación del modelo y revisión de los pasos ejecutados en relación a los objetivos del negocio, y busca determinar si hay alguna razón de negocio para el cual el modelo es deficiente, asegurándonos de esta forma, alcanzar los objetivos inicialmente propuestos. Al final de esta fase, se debe tener una decisión sobre el uso de los resultados alcanzados.

- Fase VI – "Implementación"

La fase de despliegue o implementación, dependiendo de los requisitos del proyecto, puede ser tan sencilla como la generación de un simple reporte o tan compleja como la implementación de un proceso de Explotación de Datos repetible en toda la empresa.

En determinadas ocasiones, el propio cliente es quién lleva a cabo los pasos concretos de la implementación y no el propio analista de datos; lo cual permite al cliente conocer de manera anticipada qué acciones son requeridas con el fin de hacer uso del modelo creado.
En la *Figura 2.3*, se detallan las fases componentes de la metodología CRISP-DM [Britos, 2008]:

Figura 2.3. *Fases componentes de la metodología CRISP-DM*. Extraído de [Britos, 2008].

En la *Tabla 2.2*, se especifican las tareas componentes y las actividades de cada fase de la metodología CRISP-DM [Britos, 2008]:

FASE	**TAREAS COMPONENTES**	**ACTIVIDADES ASOCIADAS**
Comprensión del negocio	Determinar los objetivos del negocio.	□ *Background*. □ Objetivos del negocio. □ Criterios de éxito del negocio.
	Evaluar la situación.	□ Inventarios de recursos. □ Requisitos, supuestos y requerimientos. □ Riesgos y contingencias. □ Terminología. □ Costos y beneficios.
	Determinar objetivos del Proyecto de Explotación de Información.	□ Las metas del Proyecto de Explotación de Información. □ Criterios de éxito del Proyecto de Explotación de Información.
	Realizar el Plan del Proyecto.	□ Plan de Proyecto. □ Valoración inicial de herramientas.
Comprensión de los datos	Recolectar los datos iniciales.	□ Reporte de recolección de datos iniciales.
	Descubrir datos.	□ Reporte de descripción de los datos.
	Explorar los datos.	□ Reporte de exploración de datos.
	Verificar la calidad de datos.	□ Reporte de calidad de datos.

Tabla 2.2. *Fases, tareas y actividades de la metodología CRISP-DM*. Extraído de [Britos, 2008].

Preparación de los datos	Caracterizar el conjunto de datos.	□ Conjunto de datos. □ Descripción del conjunto de datos.

	Seleccionar los datos.	□ Inclusión/Exclusión de datos.
	Limpiar los datos.	□ Reporte de calidad de datos limpios.
	Estructurar los datos.	□ Derivación de atributos. □ Generación de registros.
	Integrar los datos.	□ Unificación de datos.
	Caracterizar el formato de los datos.	□ Reporte de calidad de los datos.
Modelado	Seleccionar una técnica de modelado.	□ La técnica modelada. □ Supuestos del modelo.
	Generar el plan de pruebas.	□ Plan de pruebas.
	Construir el modelo.	□ Configuración de parámetros. □ Modelo. □ Descripción del modelo.
	Evaluar el modelo.	□ Evaluar el modelo. □ Revisación de la configuración de parámetros.
Evaluación	Evaluar Resultado.	□ Valoración de resultados mineros con respecto al éxito del negocio. □ Modelos aprobados.
	Revisar.	□ Revisión del proceso.
	Determinar próximos pasos.	□ Listar posibles acciones.
Implementación	Realizar el plan de implementación.	□ Plan de Implementación.
	Realizar el plan de monitoreo y mantenimiento.	□ Plan de monitoreo y mantenimiento.
	Realizar el informe final.	□ Informe final. □ Presentación final.
	Realizar la revisión del proyecto.	□ Documentación de la experiencia.

Tabla 2.2. (Continuación). *Fases, tareas y actividades de la metodología CRISP-DM.* Extraído de [Britos, 2008].

2.2.2. Metodología P^{3}TQ (Catalyst)

Según [Moine et *al*., 2011], la metodología Catalyst [Pyle, 2003], conocida como P^{3}TQ: Producto (*Product*), Lugar (*Place*), Precio (*Price*), Tiempo (*Time*) y Cantidad (*Quantity*), fue propuesta por Dorian Pyle en el año 2003. Esta metodología básicamente propone dos modelos: el "Modelo de Negocio (MII)" y el "Modelo de Explotación de Información (MIII)".

- ***Modelo de Negocio (MII)***

El Modelo de Negocio (MII) aporta una guía de pasos para el desarrollo y la realización de un modelo que permite la identificación de un problema en particular de negocio (o la oportunidad de llevar a cabo la realización del mismo), y los requerimientos reales de la organización en cuestión [Moine et *al*., 2011].

Según los mismos autores, este modelo tiene en cuenta diferentes circunstancias para el proyecto de Explotación de Datos, proponiendo acciones concretas según el contexto desde el cual se parte. En el caso de aquellos proyectos donde no existe una definición real del problema u oportunidad de negocio, se recomienda iniciar analizando las relaciones P^3TQ que existen en la cadena de valor organizacional (precio/lugar/producto/tiempo/cantidad) y que son significativas para la empresa.

Según lo expresado por [Britos, 2008], el modelado en MII depende del contexto en el cual está inmerso el negocio, lo que promueve el planteamiento de distintos escenarios. Ellos son: dato, oportunidad, prospectiva, definido y estratégico:

- 1er ESCENARIO – "DATO":

El proyecto se inicia con un conjunto de datos y la premisa es explorar este conjunto para encontrar relaciones interesantes. En este caso, se debe:

- "Determinar la procedencia y los datos a recolectar".
- "Identificar los recursos humanos para el proyecto".
- "Discutir el proyecto con los recursos humanos".
- "Caracterizar el conjunto de datos en término de las relaciones P³TQ (*Product*, *Place*, *Price*, *Time*, *Quantity*)".
- "Caracterizar la motivación del negocio para recolectar y almacenar los datos".
- "Descubrir quiénes o qué departamento originó el proyecto y qué se espera de él".

- 2do ESCENARIO – "OPORTUNIDAD":

El proyecto se inicia con una situación de negocio (problema u oportunidad) que debe ser explorada. En este caso, se debe:

- "Identificar las características de los recursos humanos relevantes".
- "Explotar las situaciones de negocio con los recursos humanos".
- "Determinar el marco de situación del negocio".
- "Definir los objetivos de negocio relevantes".
- "Buscar los datos a utilizar".
- "Presentar el caso de negocio a los recursos humanos".

- 3er ESCENARIO – "PROSPECTIVA":

El proyecto es diseñado con el fin de descubrir dónde la Explotación de Información puede brindar un valor en el entorno de la organización. En este caso, se debe:

- "Caracterizar las claves de la organización en relación a P³TQ".
- "Identificar los principales procesos de flujo de información de la organización".
- "Identificar los potenciales recursos humanos".
- "Hablar con los potenciales recursos humanos".
- "Descubrir cuáles de los 26 niveles de gestión son los más involucrados para cada uno de los recursos humanos".
- "Caracterizar los modelos más aplicables al negocio".
- "Explorar las fuentes de datos".
- "Preparar los casos de negocio para cada oportunidad significante".
- "Presentar el caso de negocio a los recursos humanos".

- ESCENARIO 4 – "DEFINIDO":

El proyecto comienza con la premisa de crear la especificación del modelo de Explotación de Datos con un propósito específico. En este caso, se debe:

- "Identificar los recursos humanos".
- "Discutir los requerimientos con los recursos humanos".
- "Enmarcar la situación de negocio".
- "Buscar los datos necesarios".
- "Definir los requerimientos a desarrollar".

- ESCENARIO 5 – "ESTRATEGICO":

El proyecto comienza con una estrategia de análisis para brindar soporte a un escenario planeado por la organización. En este caso, se debe:

- "Identificar los recursos humanos potenciales".
- "Hablar con los recursos humanos potenciales".
- "Enmarcar la situación de negocio".
- "Si es necesario, trabajar interactivamente con los recursos humanos para crear un mapa de los escenarios estratégicos".
- "A partir del mapa, crear un modelo sistémico de la situación estratégica".
- "Caracterizar las claves de la organización en relación a P³TQ".
- "Relacionar el mapa con las claves de la organización en relación a P³TQ".

- "Si es necesario, simular una situación estratégica para descubrir incertidumbres, ambigüedades, errores en la compresión intuitiva y descubrir relaciones cruciales (aquellas no entendidas, bien caracterizadas, o para las cuales los pequeños cambios tienen grandes efectos)".
- "Caracterizar las claves de las relaciones de los sistemas en término de los 26 niveles de gestión".
- "Descubrir cuáles de los 26 niveles de gestión son los más involucrados para cada recurso humano".
- "Caracterizar los modelos de los niveles de negocio más aplicados".
- "Explorar las fuentes de datos".
- "Enmarcar cada problema de negocio u oportunidad en un modelo estratégico con particular atención en las estrategias, interacciones estratégicas y los riesgos incluidos en el banco de pruebas de riesgo y las expectativas".
- "Explotar los datos para caracterizar las relaciones actuales con el sistema modelado y la simulación, tratando de que concuerden con la situación real".
- "Mostrar las relaciones descubiertas dentro de los mapas del sistema y la simulación y realizar la simulación a través del rango de escenarios requeridos".

- ***Modelo de Explotación de Información (MIII)***

Según [Britos, 2008], el Modelo de Explotación de Información (MIII) brinda una guía de pasos para la realización y ejecución de modelos de Explotación de Información a partir del Modelo de Negocio desarrollado (MII). Los pasos a seguir en MIII son:

"Preparación de los datos". En este caso, se debe:

- "Comprobar las variables de la matriz de característica".
- "Comprobar las variables básicas para el problema".
- "Comprobar los datos básicos para el problema".
- "Comprobar las variables anacrónicas".
- "Comprobar la suficiencia de los datos".
- "Comprobar la representación de los resultados".
- "Comprobar la representación de rasgos básica".

"Selección de herramientas y modelado inicial". En este caso, se debe:

- "Definir la estructura de datos para llevar adelante la Explotación de Información".
- "Caracterizar los datos de entradas y salidas".

- "Seleccionar las herramientas de Explotación de Información".
- "Construir los valores que comprueben el modelo".
- "Si los datos no se comprenden: Crear el modelo exploratorio inicial".
- "Si se van a clasificar los datos: Descubrir el tipo apropiado de modelo de clasificación inicial".
- "Si se van a predecir los datos: Descubrir el tipo apropiado de modelo predictivo".

"Ejecución". En este caso, se debe:

- "Si es un modelo deductivo: Especificar la explicación del mismo".
- "Si el modelo de clasificación o predicción es binario: Especificar una matriz de confusión".
- "Si el modelo de clasificación o predicción es un valor continuo: Especificar una matriz de confusión, comparar la predicción con un gráfico residual, comparar la predicción con la situación actual".
- "Si el modelo de clasificación o predicción es una clase: Especificar una matriz de confusión, comparar la predicción con un gráfico residual, comparar la predicción con un argumento actual, especificar pruebas del modelo residual".
- "Si el modelo de clasificación o predicción es un valor categórico: Especificar la predicción con un gráfico residual, comparar la predicción con situaciones actuales, especificar pruebas del modelo residual, realizar histogramas residuales, comparar situaciones actuales con gráficos residuales XY, comparar la situación actual con una predicción de rangos, comparar la situación actual con curvas de predicción, comparar la situación actual con la predicción apta, especificar la varianza residual, especificar el modelo perfecto".

"Evaluación de resultados". En este caso, se debe:

- "Si es un modelo deductivo: Revisar los requerimientos descubiertos durante la ejecución, explicar en forma narrativa: a) los descubrimientos, el informe debe incluir: patrones, descubrimiento de explicaciones plausibles, *clustering*, conteos, contrastes y comparaciones, variables de particionamiento, generalidades de particularidades, proponer factores plausibles explícitos e implícitos latentes, identificar y explicar las relaciones entre variables (o variables grupales), crear explicaciones de cambios lógicos, creando coherencias conceptuales; y b) la verificación, el informe debe incluir: comprobación de la representatividad, comprobación de la tendencia, triangulación (usando fuentes de datos diferentes, usando métodos de modelado diferentes, utilizando teorías diferentes),

considerando los límites, incorporando pruebas negativas, incorporando pruebas externas empíricas".

- "Si es un modelo de clasificación: Revisar las exigencias de la entrega desarrollada antes de la ejecución del proceso de Explotación de Información, repasar los descubrimientos realizados durante la formación, preparar una explicación de soporte, crear la calibración de los modelos, revisar los modelos requeridos para entregar".
- "Si el modelo es en tiempo real: Identificar las novedades".

"Comunicación de resultados". En este caso, se debe:

- "Dar a las partes restantes del proyecto los resultados y sugerir como implementarlos".

La metodología P^3TQ, en sus dos modelos, está constituida por un conjunto de pasos denominados "cajas" *(boxes).* Conceptualmente, dicha metodología determina que luego de ejecutar una acción se deben evaluar los resultados obtenidos y determinar cuál es el paso que se debe ejecutar posteriormente [Moine et *al*., 2011]. Los *boxes* del modelo explican en forma detallada los conceptos y/o acciones que se realizan [Mendez y Rodriguez, 2009]. Esto es expuesto en el trabajo de [Britos, 2008], en el cual señala que cada uno de los modelos está estructurado en base a:

- Caja de actividades: señalan el conjunto de pasos a cumplir.
- Caja de descubrimientos: informan las acciones de Exploración de Datos necesarias para poder decidir qué hacer en el paso subsiguiente.
- Caja de técnicas: brindan información complementaria acerca de los pasos recomendados en la caja de descubrimientos o de acción.
- Caja de ejemplos: proporcionan una descripción detallada acerca de cómo usar una determinada técnica. Estas cajas son aplicables en el Modelo de Explotación de Información (MIII).

En la *Figura 2.4*, se visualiza la interacción de los diferentes modelos de la metodología P^3TQ y sus componentes [Britos, 2008]:

2.2.3. Metodología SEMMA

La metodología SEMMA, creada por el SAS *Institute*, se define como el proceso de selección, exploración y modelado de grandes cantidades de datos para revelar patrones de negocio desconocidos [Moine et *al*., 2011].

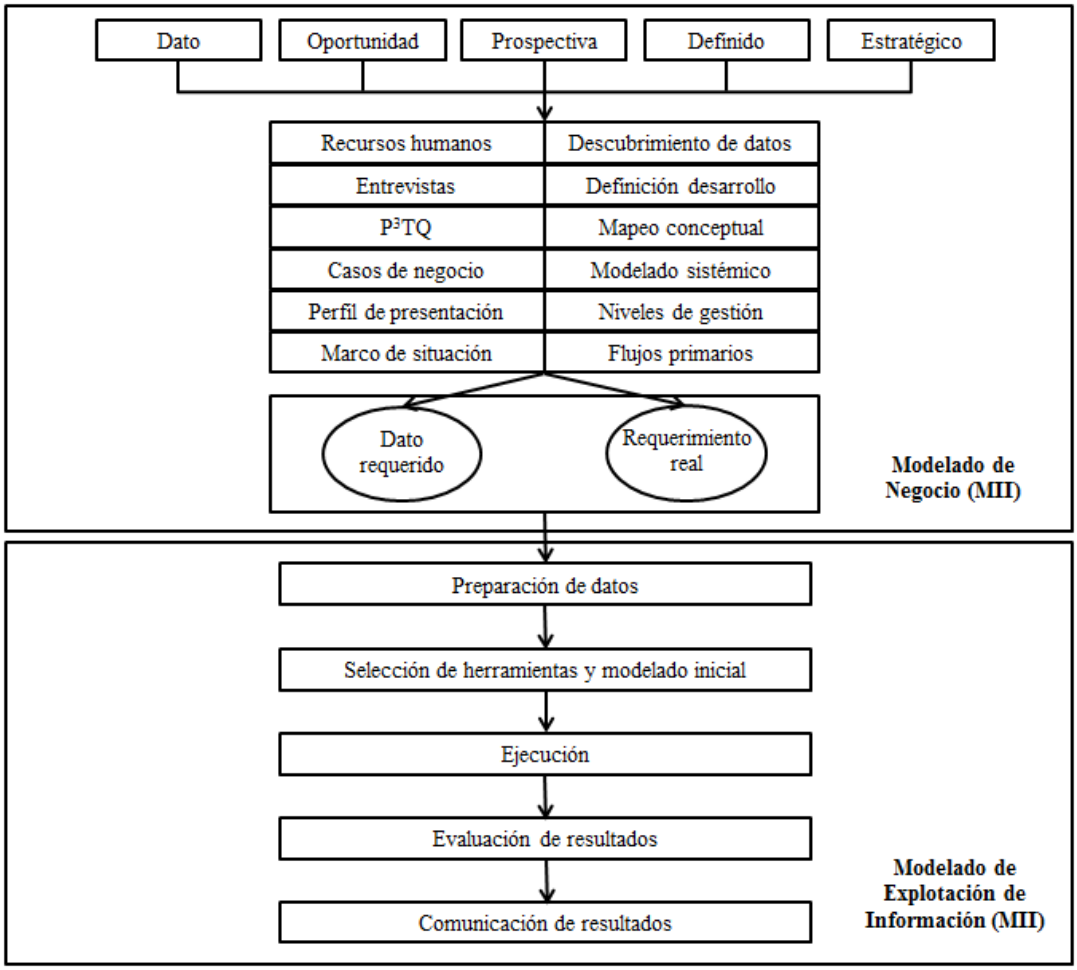

Figura 2.4. *Fases de la metodología P³TQ y sus componentes.* Extraído de [Britos, 2008].

El nombre de esta terminología corresponde al acrónimo de las cinco fases básicas del proceso: *Sample* (Muestreo), *Explore* (Exploración), *Modify* (Modificación), *Model* (Modelado), *Assess* (Valoración) [Britos, 2008; Moine et *al.*, 2011; SAS, 2012].

En la *Figura 2.5*, se observan las cinco fases básicas del proceso contemplado por la metodología SEMMA [Britos, 2008]:

Figura 2.5. *Fases componentes de la metodología SEMMA.* Extraído de [Britos, 2008].

De acuerdo a lo expresado por [Moine et *al.*, 2011], las actividades de análisis y comprensión del problema abordado son excluidas de la metodología SEMMA, ya que la misma se encuentra enfocada fundamentalmente en aspectos técnicos. Dicha metodología fue propuesta inicialmente para trabajar con el software de Explotación de Datos de la compañía SAS.

En la *Figura 2.6*, se visualizan las cinco fases de la metodología SEMMA y la dinámica general de la misma [Britos, 2008].

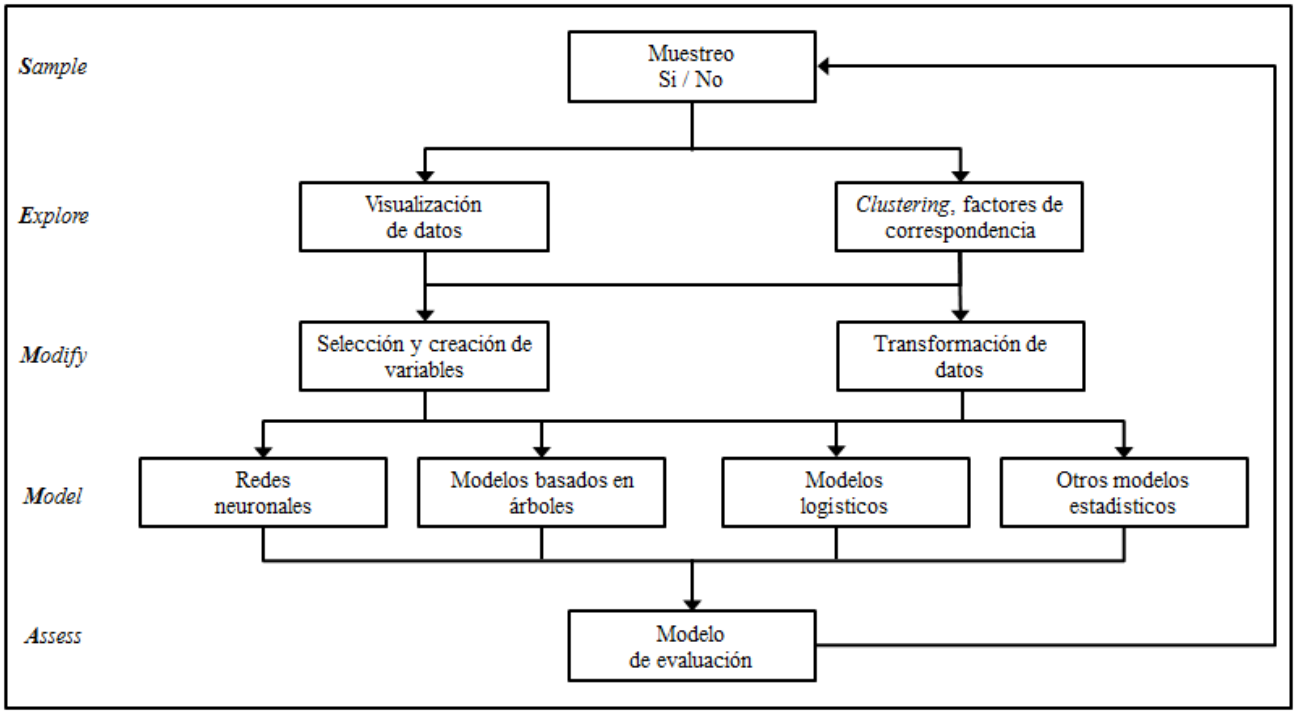

Figura 2.6. *Dinámica general de la metodología SEMMA.* Extraído de [Britos, 2008].

En [Britos, 2008] se describen las fases de la metodología SEMMA:

- Fase I – "Muestreo *(Sample)"* → Extracción de una muestra representativa.

En esta primera fase de la metodología, se realiza la extracción de un conjunto de datos (población muestral) sobre la que se va a llevar a cabo el análisis. La muestra debe ser representativa de la población, caso contrario los resultados obtenidos no son válidos para el proceso en cuestión. El método de muestreo más común se denomina "muestreo aleatorio simple", en el que cada elemento en la población tiene la misma probabilidad de ser seleccionado. En esta metodología, para cada una de las muestras escogidas se debe asociar un determinado nivel de confianza.

- Fase II – "Exploración *(Explore)"* → Exploración de los datos de la muestra seleccionada.

En esta fase, se realiza un análisis de los datos extraídos en la muestra, para lo cual se propone el uso de herramientas de visualización o de diferentes técnicas estadísticas para la exploración de la información seleccionada, que contribuyan a poner de manifiesto relaciones entre variables. Esto permite simplificar el problema y optimizar la eficiencia del modelo, ayudando a refinar los procesos de descubrimiento de información en las fases subsiguientes del proceso en cuestión.

- Fase III – "Modificación *(Modify)"* → Modificación de los datos.

La tercera fase de la metodología, involucra la modificación de los datos que van a ser ingresados al modelo para que tengan el formato adecuado, mejorando la definición de los mismos.

- Fase IV – "Modelado *(Model)"* → Modelación de los datos.

En esta fase, se procede a modelar el conjunto de datos, permitiendo al software realizar una búsqueda completa de combinaciones de datos que ayudarán a predecir los resultados esperados de manera confiable. El objetivo de esta fase es establecer una relación entre las variables objeto del estudio y las variables explicativas, de manera tal que posibiliten inferir el valor de las mismas con un nivel de confianza determinado.

Las técnicas utilizadas para el modelado de los datos incluyen técnicas adaptativas, lógica difusa, reglas de asociación, árboles de decisión, redes neuronales y computación evolutiva; como así también involucran métodos estadísticos tradicionales.

- Fase V – "Valoración *(Assess)"* → Evaluación de los datos.

La última fase de la metodología SEMMA, consiste en la valoración de los datos obtenidos para determinar el grado de confiabilidad de los mismos y así poder evaluar el modelo, mediante la comparación con otros métodos estadísticos o con nuevas poblaciones muestrales.

2.2.4. Estudio Comparativo de las Principales Metodologías de Proyectos de Explotación de Información

Según una publicación realizada en el año 2007 por la comunidad KDnuggets (*Data Mining Community's Top Resource*), CRISP-DM ha sido la metodología más utilizada en el área de Explotación de Datos [Moine et *al.*, 2011].

De acuerdo a los resultados presentados en [Moine et *al.*, 2011], la metodología SEMMA inicia el proyecto a partir de un conjunto de datos; mientras que CRISP-DM comienza el proceso con el análisis del negocio. Por tanto, la metodología P^3TQ considera cinco escenarios posibles como punto de partida (siendo más completa en este aspecto).

Teniendo en cuenta la estructura de fases del proceso de Explotación de Datos, en la *Tabla 2.3* se observa como SEMMA excluye la fase de análisis y comprensión del problema, y se enfoca principalmente en aspectos técnicos; mientras que CRISP-DM y P^3TQ contemplan dicha fase antes de comenzar el proceso de Explotación de Datos. En las tres metodologías se contemplan la fase de la selección y preparación de los datos, la fase de modelado y la fase de evaluación de los patrones

obtenidos. Sin embargo, en SEMMA el proceso de evaluación e interpretación de dichos patrones se realiza sobre el desempeño del modelo en cuestión, mientras que en las otras dos metodologías restantes la evaluación se realiza teniendo en cuenta la utilidad que se aporta al dominio de la aplicación o al problema de la organización en cuestión. En SEMMA, no se considera la fase de implementación. Por tanto, CRISP-DM adiciona además una planificación para controles futuros y un análisis de finalización del proyecto denominado 'análisis postmortem' [Moine et *al.*, 2011].

FASES	CRISP – DM	SEMMA	P^3TQ
Análisis y comprensión del negocio	Comprensión del negocio		Modelado del negocio
Selección y preparación de los datos	Entendimiento de los datos	Muestreo Comprensión	Preparación de los datos
	Preparación de los datos	Modificación	
Modelado	Modelado	Modelado	Selección de herramientas y modelado inicial
Evaluación	Evaluación	Valoración	Refinamiento del modelo
Implementación	Despliegue		Comunicación

Tabla 2.3. *Fases del proceso de Explotación de Datos de las metodologías CRISP-DM, SEMMA y P^3TQ.* Síntesis extraída de [Moine et *al.*, 2011].

En el trabajo citado en el párrafo anterior, también se considera el nivel de detalle en las tareas llevadas a cabo en cada fase; SEMMA, propone sólo los pasos generales del proyecto de Explotación de Datos, sin especificar exactamente las tareas que deben llevarse a cabo en cada una de sus fases. Aquellas organizaciones que apliquen la metodología SEMMA en sus proyectos, deberán ser responsables de establecer las tareas y las actividades que se implementarán en cada etapa, ya que sólo se definen las fases generales en dicha metodología. En cambio, CRISP-DM y P^3TQ, especifican con mayor detalle las actividades del proceso en cuestión, inclusive la metodología P^3TQ indica además 'cómo' deben realizarse las mismas.

En referencia a las actividades para la gestión del proyecto, en la *Tabla 2.4* se observa como las mismas no están incluidas en SEMMA. Por su parte, las metodologías P^3TQ y CRISP-DM incorporan actividades de planificación para las diferentes áreas de la gestión del proyecto, pero no proponen tareas de control y monitoreo [Moine et *al*., 2011].

Dichos autores señalan que en el caso de la metodología P^3TQ, las actividades de planificación se llevan a cabo en el "Modelado del Negocio (MII)". Tomando como escenario inicial un problema u oportunidad organizacional, la planificación del riesgo está presente en la tarea "Describir la situación del negocio para el proceso de Explotación". Por tanto, la planificación del alcance, tiempo, costo y recursos humanos, se ven reflejadas en la tarea "Armar el caso de negocio".

En la metodología CRISP-DM, las actividades de planificación del riego se proponen en la tarea "Evaluación de la situación", y aquellas actividades vinculadas a la planificación de alcance, tiempo, costo y recursos humanos, se presentan en la tarea "Crear un plan para el proyecto de Explotación de Datos". Si bien no se detallan tareas de monitoreo y control, el plan del proyecto debe ser examinado, y en caso de ser necesario reformado, antes del inicio de cada fase del proceso [Moine et *al*., 2011].

FASES	CRISP – DM	SEMMA	P^3TQ
Gestión del alcance	Planificación del alcance en la tarea "Crear un plan de proyecto de Explotación de Datos"		Planificación del alcance en la tarea "Armar el caso de negocio" del Modelado del Negocio
Gestión del tiempo	Planificación del tiempo en la tarea "Crear un plan de proyectos de Explotación de Datos"		Planificación del tiempo en la tarea "Armar el caso de negocio" del Modelado del Negocio
Gestión del costo	Planificación del costo en la tarea "Crear un plan de proyecto de Explotación de Datos"		Planificación del costo en la tarea "Armar el caso de negocio" del Modelado del Negocio

Tabla 2.4. *Actividades para la gestión del proyecto de las metodologías CRISP-DM, SEMMA y P^3TQ.* Síntesis extraída de [Moine et *al.*, 2011].

Gestión del riesgo	Planificación del riesgo en la tarea "Evaluación de la situación"	Planificación del riesgo en la tarea "Describir la situación del negocio para el proceso de Explotación"
Gestión de los recursos humanos	Planificación de los recursos humanos en la tarea "Crear un plan de proyecto de Explotación de Datos"	Planificación de los recursos humanos en la tarea "Armar el caso de negocio" del Modelado de Negocio.

Tabla 2.4. (Continuación). *Actividades para la gestión del proyecto de CRISP-DM, SEMMA y P^{3}TQ.* Síntesis extraída de [Moine et *al.*, 2011].

[Britos, 2008; Pollo Cattaneo et *al.*, 2010] indican que las tres metodologías identifican técnicas de Explotación de Información (TEI) utilizables. A diferencia de CRISP-DM, las metodologías P^{3}TQ y SEMMA no identifican problemas de inteligencia de negocio (PIN), ni realizan una caracterización abstracta de los mismos (CRISP-DM hace una caracterización parcialmente abstracta de dichos problemas). P^{3}TQ y SEMMA, tampoco identifican relaciones entre problemas de inteligencia de negocio y técnicas de Explotación de Información, ni procesos de Explotación de Información (CRISP-DM esboza parcialmente los procesos a desarrollar) (*Tabla 2.5*).

METODOLOGÍA / CARACTERÍSTICA	CRISP-DM	SEMMA	P^{3}TQ
Identifica problemas de inteligencia de negocio (PIN).	●	X	X
Identifica una caracterización abstracta de PIN.	○	X	X
Identifica técnicas de Explotación de Información (TEI) utilizables.	●	●	●
Identifica relaciones entre las TEI y los PIN.	○	X	X
Identifica procesos de Explotación de Información (proceso PIN x TEI).	○	X	X

Tabla 2.5. *Conceptos de inteligencia de negocio, técnicas y procesos de Explotación de Información abarcados por las metodologías CRISP-DM, SEMMA y P^{3}TQ (● = SI, ○ = parcialmente, **X** = NO).* Extraído de [Britos, 2008].

Un estudio comparativo entre las principales metodologías de Explotación de Información realizado por [Mendez y Rodríguez, 2009], permitió concluir a los autores que la metodología P^{3}TQ es la más completa de las tres mencionadas. Dicha metodología, analiza más profundamente un mayor número de variables con respecto a las otras dos metodologías. Si bien CRISP-DM es más abierta y completa que SEMMA, no llega a obtener el nivel de detalle de P^{3}TQ, ya que no se analizan los pasos, resultados y situaciones que se pueden presentar dentro de cada fase (*Tabla 2.6*).

	SEMMA	CRISP-DM	P^{3}TQ
Permite elección totalmente libre de herramientas.	NO	SI	SI
Cantidad de fases.	5	6	5 (1 MII y 4 MIII)
Todas las fases pueden relacionarse.	NO	SI	SI
Considera motivo del proyecto.	NO	NO	SI
Considera naturaleza del interés de las partes.	NO	NO	SI
Considera otros aspectos no técnicos.	NO	SI	SI
Identifica claramente las variables sobre las que el proyecto tiene impacto.	NO	NO	SI (Product, Place, Price, Time, Quantity)
Está detallada paso a paso cada etapa del método.	NO	NO	SI

Tabla 2.6. *Comparación de las principales metodologías de Explotación de Información*. Extraído de [Mendez y Rodríguez, 2009].

2.2.5. Modelo de Procesos para Proyectos de Explotación de Información (Vanrell)

El modelo de procesos para proyectos de Explotación de Datos en cuestión, fue propuesto por Juan Pablo Vanrell en el año 2012 [Vanrell, 2012]. Según el autor, para llevar a cabo esta solución, se utilizó como base el modelo de procesos Competisoft, utilizado para el desarrollo de proyectos de software clásico. Al mismo tiempo, se utilizó la metodología CRISP-DM, utilizada para el desarrollo de proyectos de Explotación de Información, con el propósito de aportar todas las herramientas necesarias requeridas para realizar todos los ajustes y adaptaciones necesarias en Competisoft, y adecuar el mismo para su uso en proyectos de tales características.

El mismo autor señala, que el modelo de procesos Competisoft utiliza MoProSoft como modelo base, y posee 3 categorías en su estructura con sus respectivos procesos y sub-procesos, reflejando la estructura organizacional. La categoría “Alta Dirección (DIR)” contiene el proceso de “Gestión de Negocio”; la categoría “Gerencia (GER)” involucra los procesos de “Gestión de Procesos”,

"Gestión de Proyectos" y "Gestión de Recursos", el cual se divide en tres sub-procesos: el sub-proceso de "Recursos Humanos y Ambiente de Trabajo", el de "Bienes, Servicios e Infraestructura" y el sub-proceso de "Conocimiento de la Organización". Por último, la categoría "Operaciones (OPE)", contiene los procesos de "Administración de un Proyecto Específico", "Desarrollo de Software" y "Mantenimiento de Software" (*Figura 2.7*).

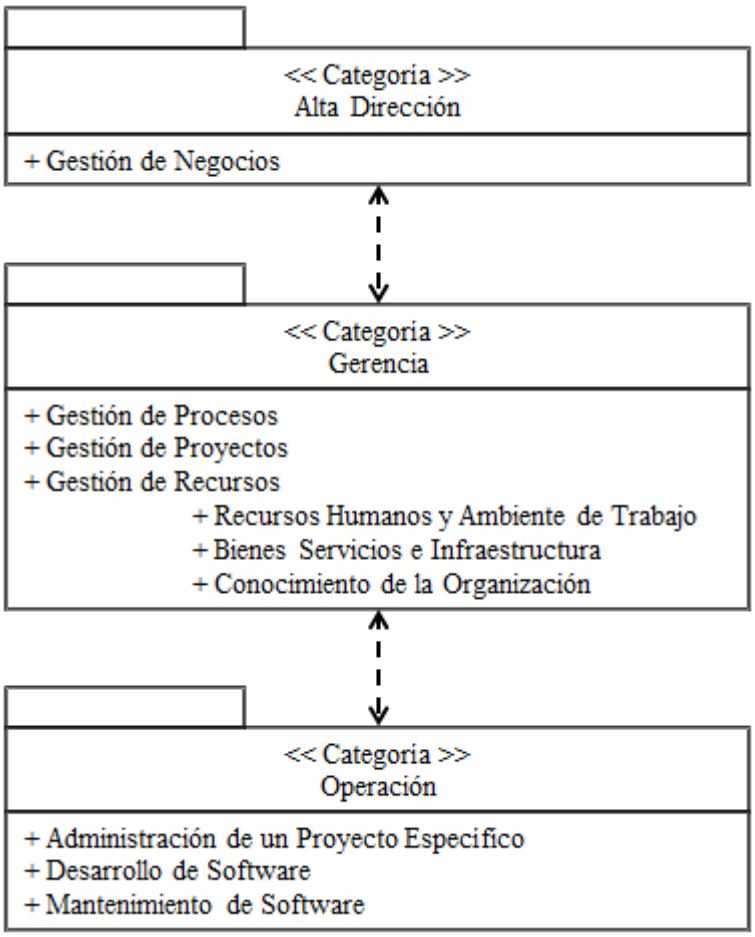

Figura 2.7. *Esquema de categorías de procesos*. Extraído de [Vanrell, 2012].

Las categorías de "Alta Dirección" y "Gerencia", son mantenidas sin cambios en el modelo desarrollado por Vanrell; el autor sostiene que las mismas están definidas a un nivel empresarial y pueden ser aplicadas a cualquier tipo de proyecto. Por tanto, se mantienen los procesos de la categoría "Operación" definidos en Competisoft, como "Administración de Proyectos" y "Desarrollo de Proyecto", adaptando los mismos a los proyectos de Explotación de Información [Vanrell, 2012].

En [Vanrell, 2012], se menciona que en el nuevo modelo propuesto, el proceso de desarrollo de Competisoft se reemplazó por las fases de desarrollo aportadas por la metodología CRISP-DM, incluyendo un conjunto de actividades y herramientas (o técnicas) faltantes en dichas fases. A partir

de las fases de desarrollo planteadas por CRISP-DM, se realizó tanto la división de los sub-procesos como las tareas a realizar en cada uno de ellos, organizando las mismas en un proceso independiente. Se incorporaron nuevas actividades al modelo en cuestión, más precisamente en los sub-procesos "Preparación de los datos", "Evaluación" y "Entrega".

Dicho autor, además señala que se agregaron recomendaciones de las técnicas a emplear para aquellas tareas definidas en los sub-procesos del proceso de "Desarrollo del Proyecto". Entre ellas se destacan:

- Generación de informes: para formalizar las tareas incluyendo una descripción de los datos que estos deben contener.
- Modelos de Entidad-Relación, Entrevistas Estructuradas y Técnicas Estadísticas: para describir el conjunto de datos.
- Diferentes Procesos de Explotación de Información (Algoritmos TDIDT, Mapas Auto Organizados, Redes Bayesianas o SOM): para la construcción de modelos.
- Método Unificado de Transformación: para llevar adelante las transformaciones de los datos.

Considerando las distintas naturalezas de los proyectos de Explotación de Información y los de software clásico, una serie de nuevos elementos se incorporaron en el proceso de administración de Competisoft, para contrarrestar las diferencias surgidas entre ambos. También se incluyeron tareas de la metodología CRISP-DM relacionadas a los procesos de administración, como así también se presentaron actividades y herramientas (o técnicas) nuevas que fueron consideradas de utilidad para el modelo de procesos [Vanrell, 2012].

En [Vanrell, 2012], se menciona que las tareas definidas en las fases de CRISP-DM, vinculadas a los procesos de administración y que fueron incorporadas al modelo descripto son las de "Determinar los objetivos del Negocio", que se relaciona con la elicitación de requerimientos y "Evaluación de la Situación", ambas tareas contempladas en el sub-proceso de "Planificación/Entendimiento del Negocio"; y por último, la tarea "Planear la entrega", la cual fue situada dentro del sub-proceso de "Cierre/Entrega". El resto de las tareas definidas en el proceso de administración del nuevo modelo pertenecen al modelo Competisoft.

Adicionalmente, este autor señala que se agregaron una serie de actividades en aquellos casos en que no estaban definidas las mismas, para llevar a cabo cada una de las etapas especificadas en el proceso de administración, para las tareas del sub-proceso de "Planificación/Entendimiento del Negocio", como son "Definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento", "Definir ciclos y actividades con base en la descripción del proyecto y en el proceso específico", "Determinar tiempo estimado para cada actividad", "Elaborar

plan de adquisiciones y capacitación", "Establecer el calendario de actividades", "Producir un Plan de Proyecto" y "Formalizar el inicio de un nuevo ciclo del proyecto".

Siguiendo con la misma línea de trabajo, el autor menciona que en el sub-proceso de "Realización" se definieron un conjunto de actividades para las siguientes tareas: "Acordar las tareas con el equipo de trabajo", "Acordar la distribución de información", "Revisar con el responsable la descripción del producto, el equipo de trabajo y el calendario", "Revisar cumplimiento del plan de adquisiciones y capacitación", "Administrar subcontratos", "Recolectar reportes de actividades y mediciones y sugerencias de mejora y productos de trabajo", "Registrar costo real del proyecto", "Revisar el registro de rastreo basado en los productos de trabajo recolectados", "Revisar los productos terminados durante el proyecto", "Recibir y analizar las solicitudes de cambio del cliente" y "Realizar reuniones con el equipo de trabajo y cliente para reportar avances y tomar acuerdos".

Del mismo modo [Vanrell, 2012], señala que en el sub-proceso de "Evaluación y control", se definieron actividades para las tareas "Evaluar el cumplimiento del plan de proyecto y plan de desarrollo" y "Analizar y controlar los riesgos".

Por último, el creador del modelo comenta que se definieron una serie de actividades para las tareas "Formalizar la terminación del proyecto o ciclo", "Llevar a cabo el cierre del contrato con subcontratistas" y "Planear la entrega" en el sub-proceso de "Cierre/Entrega".

En tanto, se recomendaron un conjunto de técnicas para ser aplicadas durante la realización de las actividades mencionadas en el proceso de "Administración del Proyecto".

Estas técnicas incluyen [Vanrell, 2012]:

- Análisis de PERT → tareas: "Determinar tiempo estimado para cada actividad" y "Establecer el calendario de actividades".
- Diagramas de Gantt → tareas: "Elaborar plan de adquisiciones y capacitación" y "Establecer el calendario de actividades".
- Diagramas de Gantt, Análisis de PERT y Análisis de Camino Crítico → tareas: "Producir un Plan de Proyecto" y "Producir un Plan de Desarrollo".
- DM-COMO o Técnicas Empíricas de Estimación → tarea: "Calcular el costo estimado del proyecto".
- Diferentes tipos de Entrevistas (Estructuradas y No Estructuradas), Taxonomías de Riesgos, Análisis de Riesgo Económico, Análisis de Riesgo Técnico, Análisis de Finanzas, Retorno de la Inversión, Análisis de Riesgo Operativo y de Soporte, Análisis de Riesgo del Programa y uso de Glosarios → tarea: "Evaluación de la situación".
- Entrevistas Estructuradas → tarea: "Generar el reporte de mediciones y sugerencias de mejora".

- Generación de informes, → para especificar las tareas incluyendo una descripción de los datos que estos deben contener.
- Organigramas y Redes de Expertos → tareas: "Determinar los objetivos del negocio" y "Evaluación de la situación".
- Taxonomías de Riesgo → tarea: "Analizar y controlar los riesgos".

2.2.6. Modelo Vanrell vs. CRISP-DM, SEMMA y P^{3}TQ

A continuación se presenta un análisis comparativo entre el modelo de procesos para proyectos de Explotación de Información comentado en la sección 2.2.5 y las metodologías CRISP-DM, SEMMA y P^{3}TQ, destacando semejanzas, diferencias y otros aspectos relevantes, con el propósito de avanzar en la construcción de un mapa de actividades para este tipo de proyectos y contemplar los cambios necesarios para soportar las actividades correspondientes, utilizando como base el modelo de proceso en cuestión.

- Categoría de Alta Dirección

De acuerdo a las conclusiones aportadas por [Vanrell, 2012], el proceso de "Gestión de Negocio" dentro de la categoría de procesos de "Alta Dirección" de Competisoft no es considerado en las metodologías CRISP-DM, SEMMA y P^{3}TQ. Dicho proceso contempla la planificación estratégica, los preparativos para la realización de la estrategia, la valoración y el mejoramiento continuo de la organización en cuestión. En el modelo de Explotación de Datos propuesto por dicho autor, se usa el proceso definido en Competisoft dentro de la categoría de procesos de "Alta Dirección". Por tanto, en base a lo expresado y al análisis de las metodologías en cuestión, se concluye que la categoría de "Alta Dirección" definida en el modelo Vanrell, y sus elementos componentes (tareas, actividades, etc.), no están contemplados en las metodologías CRISP-DM, SEMMA y P^{3}TQ.

- Categoría Gerencia

En [Vanrell, 2012], se menciona que se usan los procesos definidos en Competisoft dentro de la categoría de procesos de "Gerencia" en el modelo propuesto. Asimismo, el autor señala que los procesos "Gestión de Procesos", "Gestión de Proyectos" y "Gestión de Recursos" dentro de la mencionada categoría en Competisoft, no se encuentran contemplados por la metodología SEMMA. El proceso "Gestión de Procesos", involucra las actividades de planificación de procesos, los preparativos para la implantación y la evaluación y control de procesos; por su parte, el proceso "Gestión de Proyectos", comprende las actividades de planificación, realización y evaluación de desempeño; por último, el proceso "Gestión de Recursos", contempla las actividades internas de planificación, seguimiento y control de recursos e investigación de tendencias tecnológicas. En base

a lo expresado y al estudio de dicha metodología, se concluye que la categoría de "Gerencia" definida en el modelo Vanrell, y sus elementos constitutivos (tareas, actividades, etc.), no están contemplados en SEMMA.

Siguiendo con el análisis de la categoría de procesos de "Gerencia", el mismo autor señala que las metodologías P^3TQ y CRISP-DM, no contemplan los procesos "Gestión de Procesos" y "Gestión de Proyectos". Con respecto a la relación entre los sub-procesos contenidos en el proceso de "Gestión de Recursos" de Competisoft y las metodologías P^3TQ y CRISP-DM, existen diversas alternativas mencionadas por el autor. El sub-proceso "Bienes y Servicios e Infraestructura", que involucra las actividades de preparación de instrumentación y la generación de reportes, no se haya contemplado por P^3TQ pero si por CRISP-DM, ya que la misma posee una actividad identificada como "Inventario de Recursos" en la fase de "Comprensión del Negocio". En referencia al sub-proceso "RRHH y Ambiente de Trabajo", que incluye las actividades de preparación de la instrumentación y generación de reportes; y al sub-proceso "Conocimiento de la Organización", que comprende las actividades de planificación, realización y seguimiento y control, ambos sub-procesos se relacionan con P^3TQ, más precisamente con el "Modelado del Negocio" contemplado en dicha metodología. Por tanto, CRISP-DM no posee procesos o tareas relacionadas al sub-proceso "RRHH y Ambiente de Trabajo"; mientras que en relación al sub-proceso "Conocimiento de la Organización", se definen las tareas "Determinación de las metas del proyecto de Explotación de Información", "Evaluación de la situación" y "Determinación de los objetivos", las cuales se encuentran fuertemente relacionadas con dicho sub-proceso de Competisoft [Vanrell, 2012].

Por tanto, en base al análisis comparativo del autor y al estudio de las metodologías en cuestión, se concluye que la categoría de procesos de "Gerencia" del modelo Vanrell, se encuentra relacionada parcialmente con las metodologías P^3TQ y CRISP-DM, específicamente a través del proceso de "Gestión de Recursos" y sus tareas y actividades específicas, sin mostrar relación con los procesos "Gestión de Procesos" y "Gestión de Proyectos" y el conjunto de tareas y actividades asociadas.

- Categoría Operación

El modelo desarrollado, mantiene los procesos de la categoría de "Operación", definidos en Competisoft, como "Administración de Proyectos" y "Desarrollo de Proyectos", adaptándolos a los proyectos de Explotación de Información [Vanrell, 2012].

Según este autor, el proceso "Administración de Proyectos" del modelo objeto de estudio, consta de cuatro sub-procesos constitutivos: "Planificación/Entendimiento del negocio", "Realización", "Evaluación y control" y "Cierre/Entrega". Se incorporaron nuevos elementos en el proceso de administración de Competisoft y se incluyeron tareas definidas en las fases de la metodología

CRISP-DM relacionadas a los procesos de administración, y se agregaron una serie de actividades para el desarrollo del proceso "Administración del Proyecto" del modelo en cuestión.

El proceso "Desarrollo del Proyecto" del modelo propuesto por Vanrell contiene seis sub-procesos componentes: "Entendimiento del negocio", "Entendimiento de los datos", "Preparación de los datos", "Modelado", "Evaluación" y "Entrega". La división de los sub-procesos y las tareas correspondientes de cada uno de ellos en dicho proceso, fueron definidas a partir de las fases de desarrollo planteadas por la metodología CRISP-DM, con la incorporación de nuevas actividades no contempladas en dicha metodología [Vanrell, 2012].

De acuerdo a la discusión sobre las metodologías CRISP-DM, SEMMA y P^{3}TQ presentada por [Britos, 2008], las fases "Preparación de los datos" y "Evaluación" de la metodología CRISP-DM, se relacionan con las fases "Modificación (*Modify*)" y "Valoración (*Assess*)" de la metodología SEMMA respectivamente, y la sub-fases "Preparación de datos" y "Evaluación de resultados" de la fase de "Modelado de Explotación de Información (MIII)" de la metodología P^{3}TQ respectivamente. Por tanto, la fase de "Modelado" de CRISP-DM, se relaciona con la fase "Modelado (*Model*)" de SEMMA, y las sub-fases "Selección de herramientas y modelado inicial" y "Ejecución" de la fase de "Modelado de Explotación de Información (MIII)" de la metodología P^{3}TQ. Por su parte, las fases "Comprensión de negocio" e "Implementación" de la metodología CRISP-DM, se relacionan con la fase de "Modelado de negocio (MII)" y la sub-fase "Comunicación de resultados" de la fase "Modelado de Explotación de Información (MIII)" de la metodología P^{3}TQ respectivamente; sin embargo, estas fases no tienen relación con fases de la metodología SEMMA. Por tanto, la fase "Comprensión de los datos" de la metodología CRISP-DM, se relacionada con las fases "Muestreo (*Sample*)" y "Exploración (*Explore*)" de la metodología SEMMA; sin mostrar relación con fases de la metodología P^{3}TQ.

De manera análoga, se pueden establecer relaciones entre los sub-procesos correspondientes al proceso "Desarrollo del Proyecto" del modelo de Vanrell y las fases (y sub-fases) de las metodologías CRISP-DM, SEMMA y P^{3}TQ, teniendo en cuenta que dicho modelo incluyó el reemplazo del proceso de desarrollo de Competisoft por las fases de desarrollo descriptas en la metodología CRISP-DM.

En base a esto, se deduce que el sub-proceso "Entendimiento del negocio" del proceso "Desarrollo del Proyecto" del modelo de Vanrell, se relaciona con la fase de "Modelado de negocio (MII)" de la metodología P^{3}TQ, en aquellas cuestiones vinculadas al desarrollo del proyecto propiamente dicho, excluyendo los aspectos relacionados a la administración de dicho proyecto y que se encuentran relacionados al proceso "Administración de Proyectos", y no guarda relación con fases de la metodología SEMMA. Por tanto, el sub-proceso de "Entendimiento de los datos", se relacionada

con las fases "Muestreo (*Sample*)" y "Exploración (*Explore*)" de la metodología SEMMA; sin mostrar relación con fases de la metodología P^{3}TQ. El sub-proceso "Preparación de los datos", se relaciona con la fase "Modificación (*Modify*)" de la metodología SEMMA y la sub-fase "Preparación de datos" de la fase de "Modelado de Explotación de Información (MIII)" de la metodología P^{3}TQ. Por su parte, el sub-proceso "Modelado", se relaciona con la fase "Modelado (*Model*)" de SEMMA y las sub-fases "Selección de herramientas y modelado inicial" y "Ejecución" de la fase de "Modelado de Explotación de Información (MIII)" de la metodología P^{3}TQ. El sub-proceso "Evaluación", se relaciona con la fase "Valoración (*Assess*)" de la metodología SEMMA y la sub-fase "Evaluación de resultados" de la fase de "Modelado de Explotación de Información (MIII)" de la metodología P^{3}TQ respectivamente. Por último, el sub-proceso "Entrega", se relaciona con la sub-fase "Comunicación de resultados" de la fase "Modelado de Explotación de Información (MIII)" de la metodología P^{3}TQ; sin presentar relación con las fases de la metodología SEMMA.

Según el análisis comparativo de las metodologías CRISP-DM, SEMMA y P^{3}TQ realizado por [Moine et *at.*, 2012], P^{3}TQ es la metodología más completa en lo que respecta al análisis y comprensión del negocio, mientras que SEMMA excluye esta actividad del modelo. A estas conclusiones pueden añadirse las obtenidas por [Vanrell, 2012], mencionando que la metodología P³TQ define una evaluación completa del negocio denominada "Modelado del Negocio".

Según [Ochoa et *al.*, 2006], cualquier proyecto de software puede fracasar como consecuencia de una comprensión incorrecta del negocio. Por tal motivo, el entendimiento adecuado del mismo ayuda a definir el tipo de solución a desarrollar y condiciona el proyecto en cuestión.

Por consiguiente, en el caso de los sub-procesos "Planificación/Entendimiento del negocio" y "Entendimiento del negocio"; definidos en los procesos de "Administración de Proyectos" y "Desarrollo del Proyecto" respectivamente del modelo Vanrell; y teniendo en cuenta la relación existente con el "Modelado del Negocio" de P^{3}TQ y el análisis más completo del negocio contemplado en dicha metodología en comparación con CRISP-DM y SEMMA [Moine et *at.*, 2012], se procederá a considerar dicha información para avanzar en el desarrollo del mapa de actividades para proyectos de Explotación de Información, contemplando los cambios necesarios para soportar las mismas.

3. DESCRIPCIÓN DEL PROBLEMA

En este capítulo, se plantea la problemática de investigación del Trabajo de Especialidad, a partir de la complejidad que involucra la tarea de construcción de los mapas de actividades en proyectos de Explotación de Información (sección 3.1).

3.1. IDENTIFICACIÓN DEL PROBLEMA DE INVESTIGACIÓN

En referencia al trabajo de [Pressman, 2004], [Pytel, 2011] señala que la gestión de un proyecto de software se inicia a partir de una serie de actividades que se denominan planificación del proyecto. Antes del comienzo de un proyecto, se debe realizar una estimación: del trabajo a realizar, de los recursos necesarios y del tiempo que transcurrirá desde el inicio hasta el final de su realización.

Por su parte, los proyectos de Explotación de Información también requieren de un proceso de planificación que permita realizar estimaciones. Sin embargo, existen diferencias entre un proyecto de Explotación de Información y un proyecto convencional de construcción de software que limitan la utilización de los métodos tradicionales de estimación [Rodríguez et *al*., 2010].

Una metodología estándar, puede ser adaptada a un proyecto específico, mediante la construcción del mapa de actividades para ese proyecto en particular, teniendo en cuenta las características propias del mismo [Diez et *al*., 2003].

Los proyectos de Explotación de Información, necesitan elegir el ciclo de vida más adecuado y detallar el mapa de actividades. En función de las características del proyecto en cuestión, dichas actividades se seleccionan o eliminan [Britos, 2006]. Dependiendo del ciclo de vida elegido el mapa de actividades variará. El mapa de actividades es básicamente una tabla que describe qué actividades se van a ejecutar para un determinado proyecto [Juristo, 2003]. En él se detallan cuáles son las actividades más relevantes para el proyecto en cuestión. Sin embargo, la construcción de un mapa de actividades no es una tarea insignificante, ni tampoco puede ser generada de forma automática. Para llevar a cabo esta tarea de forma correcta, se requiere de capacidad análisis, conocimientos adecuados y experiencia en la aplicación de metodologías estándares de desarrollo [Diez, 2003].

El responsable del proyecto debe evitar la realización de actividades innecesarias, seleccionando aquellas que permitan cubrir las necesidades de ese proyecto. La selección errónea de segmentos de una metodología estándar podría provocar algunos inconvenientes durante el desarrollo del

proyecto e incluso dificultar el cumplimiento efectivo de los objetivos propuestos en el mismo. Una selección acertada de los segmentos genera confianza en el equipo de proyecto para el uso de la metodología. Si bien una metodología estándar puede aportar mapas de actividades predefinidos, éstos se utilizan como marco de referencia para el desarrollo del mapa de actividades del proyecto en cuestión, según las particularidades del mismo. El mapa de actividades constituye una guía para el responsable del proyecto, donde todos y cada uno de los miembros del equipo de proyecto conocen las actividades que deberán realizar durante la ejecución del mismo y las habilidades o conocimientos específicos de las que deberán hacer uso. La construcción del mapa de actividades de un proyecto en particular requiere de capacidades especiales. [Diez et *al*., 2003].

Por tanto, el mapa de actividades constituye el punto de partida para lograr una organización del proyecto que permita su propia gestión. A partir del mismo, se puede realizar una estimación del tiempo y costo de cada una de las actividades componentes, y por ende, del proyecto general; de la asignación de recursos para cada actividad en particular, etc. [Juristo, 2003].

En el presente trabajo se propone la construcción de un marco teórico para avanzar en la desarrollo de un mapa de actividades, tomando como base un Modelo de Procesos para proyectos de Explotación de Información para PyMEs y contemplando las modificaciones necesarias para soportar las mismas.

4. CONCLUSIONES

En este Capítulo, se presentan las conclusiones derivadas de la presente investigación conforme a la revisión de la literatura relacionada con el tema en cuestión (sección 4.1).

4.1. APORTACIONES DEL TRABAJO DE ESPECIALIDAD

El presente trabajo constituye una aproximación al estudio de los mapas de actividades para proyectos de Explotación de Información. Se espera que dicho trabajo aporte elementos que contribuyan al desarrollo de la temática objeto de estudio en el marco científico-académico.

Los proyectos de Explotación de Información se están transformando en proyectos de ingeniería; por tal motivo, los modelos de procesos actuales deben ser analizados para incorporar en su estructura el punto de vista de la misma. Por tanto, los procesos a aplicar para la resolución de problemas de tales dimensiones, deberán contar con todas las actividades y tareas necesarias en un proceso de ingeniería [Mariscal et *al*., 2007].

En los últimos años, los proyectos de Explotación de Datos han experimentado una fuerte expansión; incrementándose el número, la complejidad y la variedad de los mismos; por ende, resulta necesario que los diferentes procesos de desarrollo tengan que estandarizarse para lograr resultados que puedan ser reutilizados en un futuro [Mariscal et *al*., 2007]. En el contexto sobre la gestión de proyectos de Explotación de Datos, se han ido desarrollando algunas metodologías que permiten gestionar la complejidad de los mismos de manera uniforme [Britos, 2008].

Una metodología estándar puede ser adaptada a un proyecto específico, mediante la confección del mapa de actividades para el proyecto en cuestión, considerando las particularidades del mismo. Sin embargo, la tarea de construir un mapa de tales características no se puede efectuar de manera automática. Para realizarse de forma correcta, se requiere experiencia en la aplicación de metodologías estándares de desarrollo, capacidad de análisis y conocimientos adecuados en la materia, ya que no es una tarea simple de llevarse a cabo [Diez et *al*., 2003].

Los proyectos de explotación de información, como proyectos informáticos, necesitan elegir el ciclo de vida más apropiado y especificar el mapa de actividades [Britos et *al*., 2006]. La selección del ciclo de vida constituye el complemento lógico del mapa de actividades para adaptar una metodología estándar a un proyecto en particular [Diez et *al*., 2003]. El éxito de un proyecto dependerá del ciclo de vida seleccionado para desarrollar el mismo [Mariscal et *al*., 2007].

A partir de la información contenida en el mapa, el responsable del proyecto podrá gestionar cuidadosamente cada una de las actividades seleccionadas durante el desarrollo del mismo, con el propósito de minimizar posibles inconvenientes que se puedan presentar, y que afecten de forma negativa el cumplimiento efectivo de los objetivos planteados.

Disponer de una matriz de actividades permitirá al responsable del proyecto disminuir el margen de error a la hora de realizar ciertas estimaciones y proyecciones para las variables de interés; y por ende, permitirá realizar una evaluación global del proyecto más precisa. Realizar inferencias válidas a partir de esta clase de información, es sumamente beneficioso para cualquier organización, ya sea desde el punto de vista empresarial, económico, financiero, del planeamiento estratégico, etc.

En la primera parte de este trabajo, realizamos una descripción de los mapas de actividades, exponiendo un conjunto de características asociadas a ellos. A su vez, presentamos una serie de ventajas relacionadas a la confección y utilización de los mismos, justificando la creación e implementación de mapas de actividades para proyectos de Explotación de Información.

En la segunda parte, presentamos una descripción detallada de cada una de las tres metodologías principales que se utilizan para proyectos de Explotación de Información. Adicionalmente, realizamos un estudio comparativo de las metodologías expuestas, identificando similitudes, diferencias y otros aspectos relevantes inherentes a las mismas. El tratamiento de la información llevado a cabo en esta sección nos introduce en el estudio de las metodologías mencionadas, lo cual constituye una parte fundamental de nuestra base de conocimientos para abordar el presente trabajo y futuras líneas de investigación.

En la tercera etapa de este trabajo, efectuamos un estudio descriptivo del modelo de procesos para proyectos de Explotación de Información desarrollado por Vanrell. Como se puede apreciar, dicho modelo contempla dos procesos bien diferenciados: el proceso de “Administración de Proyectos” y el proceso de “Desarrollo de Proyectos”. El análisis realizado nos permite obtener importantes conocimientos referentes al modelo en cuestión, con el propósito de avanzar en el desarrollo de una matriz de actividades.

En la última parte de este trabajo, realizamos un análisis comparativo entre los elementos constitutivos de las principales metodologías de Explotación de Información: CRISP-DM, P^3TQ y SEMMA, y aquellos elementos que forman parte del modelo de procesos para proyectos de Explotación de Información objeto de estudio, identificando información relevante que va a ser utilizada para la construcción de un mapa de actividades a partir de los resultados obtenidos.

La investigación realizada en el presente trabajo justifica la construcción de un mapa de actividades para Proyectos de Explotación de Información, utilizando como base el Modelo de Procesos orientado a Pequeñas y Medianas Empresas desarrollado por Vanrell.

5. REFERENCIAS

Amón Uribe, I. y Jiménez Ramírez, C. (2009). *Hacia una Metodología para la Selección de Técnicas de Depuración de Datos*. Revista Avances en Sistemas e Informática. 6 (1): 185-190. ISSN 1657-7663.

Britos, P. (2008). *Procesos de Explotación de Información basados en Sistemas Inteligentes.* Tesis Doctoral. Universidad Nacional de La Plata, Facultad de Informática. La Plata, Argentina. http://www.iidia.com.ar/rgm/tesistas/td-pb-fi-unlp.pdf. Página web vigente al 21/03/2103.

Britos, P., Fernández, E., García-Martínez, R. (2006). *Propuesta Matriz de Actividades para un Ciclo de Vida de Explotación de Datos. Reportes Técnicos en Ingeniería del Software*. 8(2): 36-42. ISSN 1667-5002.

Chapman, P.; Clinton, J.; Keber, R.; Khabaza, T.; Reinartz, T.; Shearer, C.; Wirth, R. (2000). *CRISP-DM 1.0 Step by step BIguide*. Edited by SPSS.

Diez, E. (2003). *Sistema Generador del Mapa de Actividades de un Proyecto de Desarrollo de Software*. Tesis de Maestría en Ingeniería del Software (Instituto Tecnológico de Buenos Aires –Facultad de Informática de la Universidad Politécnica de Madrid).

Diez, E., Britos, P., Rossi, B., García-Martínez, R. (2003). *Generación Asistida del Mapa de Actividades de Proyectos de Desarrollo de Software*. Reportes Técnicos en Ingeniería del Software. (5)1:13-18. ISSN 1667-5002.

Juristo, N. (2003). *Proceso Software*. Material correspondiente a la Maestría en Ingeniería del Software del Instituto Tecnológico de Buenos Aires – Facultad de Informática de la Universidad Politécnica de Madrid.

Mariscal, G., Marbán, Ó.; González, A., Segovia, J. (2007). *Hacia la Ingeniería de Data Mining: Un Modelo de Proceso para el Desarrollo de Proyectos*. Proceedings V Taller de Minería de Datos y Aprendizaje (TAMIDA '07). Pág. 139-148. ISBN 978-84-9732-602-5

Mendez, P., Rodriguez, A. (2009). *Herramienta de Estudio de Viabilidad para Proyectos que Utilizan la Metodología P3TQ*. Trabajo Profesional de Ingeniería en Informática. Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. Universidad de Buenos Aires.

Moine, J.; Gordillo, S.; Haedo, A. (2011). *Análisis Comparativo de Metodologías para la Gestión de Proyectos de Minería de Datos*. Proceedings VIII Workshop Bases de Datos y Minería de Datos (WBDDM). Pág. 931-938.

Moine, J.; Gordillo, S.; Haedo, A. (2011). *Estudio comparativo de metodologías para minería de datos*. Proceedings XIII Workshop de Investigadores en Ciencias de la Computación. Pág. 278-281. ISBN 978-950-673-892-1.

Ochoa, M., Britos, P. y García-Martínez, R. (2006). *Una Protofase de Entendimiento del Negocio para Metodologías de Desarrollo de Sistemas*. XII Congreso Argentino de Ciencias de la Computación. San Luis. Argentina.

Pyle, D. (2003). *Business Modeling and Data Mining*. Morgan Kaufmann Publishers.

Pytel, P. (2011). *Método de Estimación de Esfuerzo para Proyectos de Explotación de Información. Herramienta Para Su Validación*. Tesis de Magister en Ingeniería del Software. Convenio Universidad Politécnica de Madrid e Instituto Tecnológico Buenos Aires.

Pollo-Cattaneo, F., Amatriain, H., Rodriguez, D., Pytel, P., Ciccolella, E., Vegega, C., Dearriba, M., Rodríguez Aubert, M., Bose, F., Giordano, L., Britos, P., García- Martínez, R. (2010). *Ingeniería de Proyectos de Explotación de Información*. Proceedings XII Workshop de Investigadores en Ciencias de la Computación. Pág.172-176.

Pressman, R. (2004). *Software Engineering: A Practitioner's Approach*. Editorial Mc Graw Hill.

Rodriguez, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. (2010). *Estimación Empírica de Carga de Trabajo en Proyectos de Explotación de Información*. Anales del XVI Congreso Argentino de Ciencias de la Computación. Pág. 664-673. ISBN 978-950-9474-49-9.

SAS (2012). *SAS Enterprise Miner: SEMMA*. http://www.sas.com/offices/europe/uk/technologies/analytics/datamining/miner/semma.html. Último acceso Abril 2012.

Vanrell, J. (2012). *Un Modelo de Procesos para Proyectos de Explotación de Información*. Tesis de Magister en Ingeniería de Sistemas de Información. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional.

Vilalta, J. A y Espinosa, M. (2008). *Metodología para el Diagnóstico de la Calidad de los Datos.* Journal Ingeniería Industrial. 29 (2): 1-6.