

# Propuesta de Modelo de Procesos para una Ingeniería de Explotación de Información: MoProPEI

Sebastian Martins<sup>1,2</sup>, Patricia Pesado<sup>1,3</sup>, Ramón García-Martínez<sup>1,2</sup>

1. Programa de Doctorado en Ciencias Informáticas. Facultad de Informática. Universidad Nacional de La Plata

2. Laboratorio de Investigación y Desarrollo en Ingeniería de Explotación de Información. Grupo de Investigación en Sistemas de Información. Universidad Nacional de Lanús

3. Instituto de Investigaciones en Informática LIDI. Facultad de Informática. UNLP - CIC  
martinssebastian@yahoo.com.ar, ppesado@lidi.info.unlp.edu.ar, rgm1960@yahoo.com

**Resumen**—Los modelos de procesos para proyectos de explotación de información (en ocasiones mencionados como minería de datos) existentes, carecen de la visión de gestión necesaria para llevar a cabo este tipo de proyectos de manera exitosa ó presentan una estructura la cual no se adapta completamente a las necesidades de este tipo de proyectos. En este trabajo se propone un modelo de procesos para proyectos de explotación de información que guíe el desarrollo del proyecto, considerando tanto los aspectos de gestión, así como los técnicos, con el fin de generar piezas de conocimiento las cuales sirvan como soporte para la toma de decisiones. Complementariamente, se reestructuran las actividades técnicas favoreciendo el progreso fluido del proyecto.

**Palabras Claves**— Modelo de Proceso, Ingeniería de Explotación de información, Minería de Datos, Gestión de Proyectos.

## I. INTRODUCCIÓN

La Inteligencia de Negocio propone un abordaje interdisciplinario (dentro del que se encuentra la Informática), que tomando todos los recursos de información disponibles y el uso de herramientas analíticas y de síntesis con capacidad de transformar la información en conocimiento, se centra en generar a partir de éstos, conocimiento que contribuya con la toma de decisiones de gestión y generación de planes estratégicos en las organizaciones [1; 2].

El termino Minería de Datos está fuertemente ligado al concepto de base de datos y se remonta a la definición de algoritmos de búsqueda de patrones en grandes bases de datos [3]. Sin embargo, hoy existen líneas de investigación en campos tales como: minería de textos [4], minería de imágenes [5], minería de patrones en flujos de información [6], minería en la web [7], entre otras. En este contexto, se conviene utilizar el término “explotación de información” como referencia genérica a cualquiera de los tipos de minería precitados.

A partir del párrafo previo, se define a la Explotación de Información como la sub-disciplina de los Sistemas de Información que aporta a la Inteligencia de Negocio [8] las herramientas para la transformación de información en conocimiento [9].

Con base en que la Ingeniería de Software ha sido definida en el SWEBOOK [10] como: “la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo, operación y mantenimiento de software, y el estudio de estos enfoques, es decir, la aplicación de la ingeniería al software”; se conviene en definir a la Ingeniería de Explotación de Información como la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo de proyectos de explotación de información, y el estudio de este enfoque, es decir, la aplicación de la ingeniería a la explotación de

información. La ingeniería de explotación de información entiende que los procesos y las metodologías son utilizadas para: ordenar, controlar y gestionar la tarea de encontrar patrones de conocimiento en masas de información [11].

Una de las lecciones aprendidas sobre desarrollo de software en Informática derivada de los estadios tempranos de la disciplina, es que, la ausencia de una ingeniería de software conllevaba a un desarrollo artesanal de los artefactos software [12]. El desarrollo artesanal implicaba la imposibilidad de poder establecer dentro de valores racionales, parámetros tales como: [a] cantidad y calificación de los recursos humanos a emplear en el proyecto, [b] tiempos de desarrollo, [c] modelos de proceso que guiaran el desarrollo y permitieran establecer hitos de entrega, [d] formalismos de documentación que dieran cuenta de lo hecho en el proyecto de desarrollo del artefacto software y de las decisiones de diseño asumidas, constituyendo el punto de partida para futuras ampliaciones de funcionalidades, [e] modelos de costo de proyecto [13], entre otros. De hecho la estimación de estos parámetros se hacía en base a la experiencia de individuos sin ninguna base ingenieril; y lo que para un grupo de desarrollo podía hacerse en meses, para otro podía hacerse en años.

En [14], se definen a los procesos como un marco de trabajo de las tareas que se requieren para construir un producto final de alta calidad, cuya visión ingenieril incorpora los métodos técnicos y de gestión, y herramientas a utilizar. Algunos de los objetivos principales de definir un marco de trabajo común en la organización son:

- Planificar y definir cada una de las actividades o tareas requeridas para el desarrollo de un proyecto, así como establecer las responsabilidades de las mismas.
- Definir y producir resultados estandarizados (documentos, productos software, etc.), los cuales puedan ser comprendidos, comparados, repetidos y reutilizados por cualquier miembro de la organización, en cualquier momento.
- Definir hitos de evaluación y control de los productos obtenidos, así como de la viabilidad y desvíos del proyecto.
- Planificar y producir conocimiento a partir de las experiencias registradas en los distintos proyectos que la organización realice.

De forma complementaria, en [15], define la necesidad de aplicar una metodología para el desarrollo de proyectos de explotación de información, para evitar obtener dos resultados indeseados: aprender conocimientos que no son verdaderos y aprender conocimiento poco útil. Dado que el objetivo para el cual se realiza un proyecto de explotación de información, es la generación de conocimiento para dar soporte a la toma de

decisiones, el riesgo que conlleva basar dichas decisiones sobre conocimientos no válidos, el primero de los resultados no deseados, es mayor que el segundo tipo de resultados indeseados y puede ocasionar grandes pérdidas a la organización.

Los proyectos de ingeniería de explotación de información poseen características muy distintas a los proyectos de desarrollo de software tradicional [16], sobre todo en la parte operativa del proyecto. La diferencia se presenta en los procesos de desarrollo y mantenimiento en los cuales el ciclo de fases de un proyecto de software tradicional: inicio, requisitos, análisis y diseño, construcción, integración y pruebas no resultan naturales en un proyecto de explotación de información [17]. Por otra parte, al evaluar las principales metodologías existentes en el mercado para los proyectos de explotación de información [18; 19; 20], se observa la falta de herramientas que permitan soportar de forma completa la fase de administración de proyectos.

Es objetivo de este artículo, proponer un Modelo de Procesos para Proyectos de Explotación de Información. Para comprender la necesidad del mismo, se describe en la sección II el estado actual de la cuestión, presentando las propuestas existentes. En la sección III, se explicitan los problemas identificados, causantes de la necesidad de definir un nuevo Modelo de Proceso. En la sección IV, se define la notación utilizada en la solución. En la sección V se presenta la propuesta de solución. Por último, se definen las conclusiones (sección VI).

## II. ESTADO DEL ARTE

En esta sección se abordan las propuestas de modelos de procesos relevantes en el área. En la sección A, se presenta al primer modelo de proceso KDD, Descubrimiento de Conocimiento en Bases de Datos, (del inglés Knowledge Discovery in DataBase), a partir del cual gran parte de las propuestas posteriores se inspiraron. En la sección B, se describe al modelo de proceso CRISP-DM, (del inglés Cross Industry Standard Process for Data Mining), al ser el modelo de mayor uso en la industria. Finalmente se presenta una propuesta que amplía los enfoques previos, incorporando aspectos asociados con la gestión de proyectos: el modelo de proceso propuesto por Marbán (sección C).

### A. KDD

En [3] se presenta la primera aproximación a un modelo de procesos, definiendo las principales actividades técnicas orientadas a estructurar, ordenar y guiar el desarrollo de las tareas necesarias para producir conocimiento a partir de los datos almacenados por una organización.

El objetivo principal es proveer un conjunto de herramientas para automatizar el proceso de análisis de datos y el artesanal proceso estadístico de selección de hipótesis. El objetivo subyacente al concepto de KDD, es el de diferenciar a la minería de datos, entendiendo a esta como la actividad de aplicar distintos algoritmos en los datos para obtener patrones, del proceso necesario para generar conocimiento a partir de los datos. Es decir, entender a la minería de datos como un subproceso que integra a un proceso general destinado a obtener patrones de conocimiento. Mediante la incorporación de KDD, se agregan una serie de pasos destinados a favorecer y garantizar el éxito de los resultados obtenidos en la aplicación de minería de datos. No basta con obtener patrones a partir de la ejecución de algoritmos de Minería de Datos, sino que se debe preparar los datos, así como el personal para una

correcta aplicación de los métodos y posteriormente analizar los patrones obtenidos para generar conocimiento (el cual pueda ser comprendido) y reutilizar este conocimiento para retroalimentar el proceso y para ser utilizado por los interesados. “Los pasos adicionales a KDD, tales como Preparación de los Datos, Selección de los Datos, Limpieza de los Datos, Incorporación del Conocimiento Previo Adecuado y Correcta Interpretación de los Resultados de la Minería de Datos, son esenciales para garantizar la extracción de conocimiento útil desde los datos. La implementación de los métodos de Minería de Datos a ciegas, puede ser una actividad peligrosa, la cual fácilmente puede concluir en el descubrimiento de patrones inválidos o sin sentido.” [3].

En [3] se definió al proceso KDD como un conjunto de pasos interactivos e iterativos en los cuales el usuario debe tomar una serie de decisiones. La figura 1, ilustra los pasos asociados al proceso KDD y las salidas de los mismos.

Los pasos que integran al proceso son:

1. Comprender el dominio de negocio, los conocimientos previos relevantes e identificar las metas del proyecto desde el punto de vista del cliente.
2. Seleccionar el conjunto de datos, centrándose en el subconjunto de variables o registros sobre los cuales se realizará el descubrimiento.
3. Limpiar y Pre-procesar los datos, con el objetivo de eliminar los ruidos y campos inexistentes para favorecer al proceso de descubrimiento.
4. Reducción y proyección de los datos, donde se analiza la posibilidad de reducir la cantidad de variables bajo consideración.
5. Relacionar el objetivo del proceso, con un método particular de Minería de Datos.
6. Análisis exploratorio y selección del modelo e hipótesis. En este paso se eligen los algoritmos y se seleccionan los métodos a utilizar para encontrar patrones.
7. Implementar Minería de Datos: A partir de la implementación correcta de los pasos previos, se asiste significativamente a la obtención de mejores resultados en este paso.
8. Interpretar los resultados. Dicho paso puede generar la necesidad de iterar con alguno de los pasos previos.
9. Actuar sobre el conocimiento descubierto, es decir implementar el conocimiento en algún sistema, o simplemente documentarlo para su uso. Además incluye la confrontación del conocimiento identificado con previos conocimientos o creencias.

### B. CRISP-DM

En [18], se define a CRISP-DM como una metodología descrita en términos de modelo de proceso jerárquico, la cual se conforma por un conjunto de tareas descritas en cuatro niveles de abstracción (fases, tareas generales, tareas específicas e instancia del proceso), las cuales se ejecutan en base al modelo de ciclo de vida definido.

Este modelo de proceso, ampliamente utilizado para el desarrollo de proyectos de explotación de información, se centra en el desarrollo técnico del proyecto, considerando levemente algunas actividades vinculadas con la administración del mismo. Dicha metodología se desarrolla a partir de un ciclo de vida, compuesto por seis fases (cada una compuesta por varias actividades), las cuales interactúan y se retroalimentan entre ellas (figura 2).

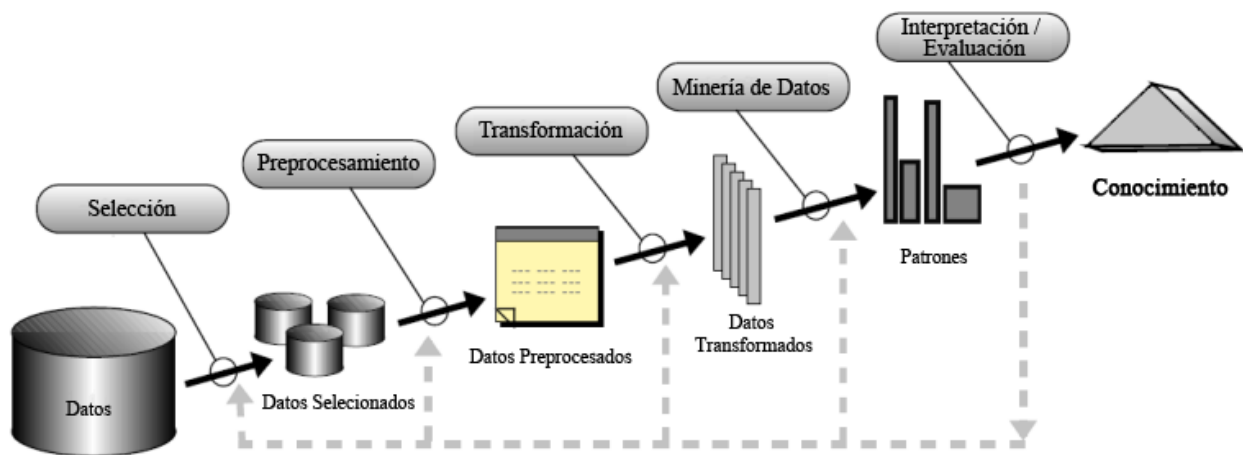


Fig. 1. Modelo de Proceso KDD [3]

El círculo exterior simboliza la naturaleza de la Minería de Datos, la retroalimentación de los resultados encadenando nuevas preguntas mejor enfocadas al negocio, y por consiguiente nuevas ejecuciones del ciclo de vida.

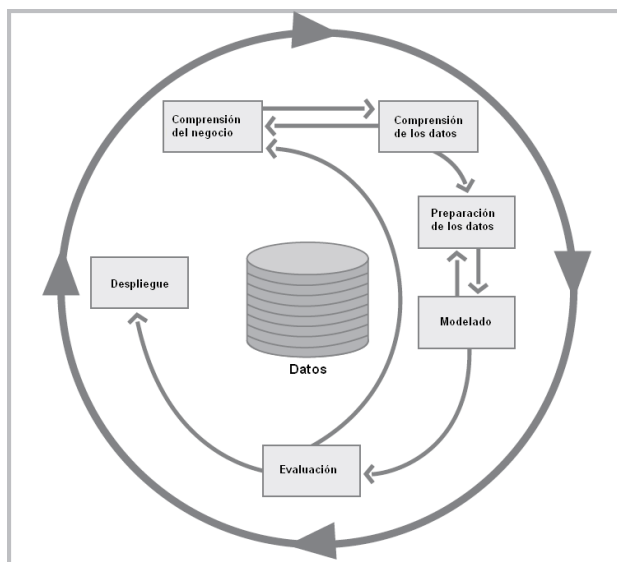


Fig. 2. Modelo de Proceso: CRISP-DM [18]

Como se observa en la figura 2, las fases que componen al modelo de proceso son:

1. **Comprensión del Negocio**, cuyo objetivo es comprender los objetivos y requerimientos del proyecto desde la perspectiva del negocio, así como identificar el problema de minería de datos y realizar la planificación del proyecto. Las actividades asociadas son: Determinar los Objetivos del Negocio, Evaluación de la Situación, Determinar las Metas de Minería de Datos y Producir el Plan del Proyecto.
2. **Comprensión de los Datos**, donde se realiza una recolección inicial de los datos y se los analiza con el objetivo de identificar problemas de calidad y posibles subconjuntos de interés para distintas hipótesis. Las actividades generales que lo integran son: Recolección Inicial de los Datos, Descripción de los Datos, Exploración de los Datos y Verificación de la Calidad de los Datos.
3. **Preparación de los Datos**, conlleva la ejecución de todas las actividades necesarias para favorecer la calidad de los resultados. Las actividades generales que lo integran son: Seleccionar los Datos, Limpieza de los Datos, Construcción

de los Datos, Integración de los Datos y Formateo de los Datos.

4. **Modelado**, se seleccionan y configuran las técnicas de modelado a utilizar. Las actividades generales que componen dicha fase son: Seleccionar de las Técnicas de Modelado, Generar el Diseño de las Pruebas, Construir el Modelo y Evaluar el Modelo.
5. **Evaluación**, se analiza el modelo generado para garantizar que este cumpla con los objetivos del negocio. Las actividades generales que lo integran son: Evaluar los Resultados, Revisar el Proceso y Determinar Próximos Pasos.
6. **Despliegue**, la cual abarca las actividades de integrar el conocimiento obtenido a algún sistema, o documentarlo para su posterior uso. Las actividades generales que conforman la última fase son: Planificar la Implementación, Planificar el Monitoreo y Mantenimiento, Producir el Reporte Final y Revisión del proyecto.

### C. IEEE Std. 1074+ CRISP-DM

El modelo de proceso propuesto por Marbán [21], consiste fundamentalmente en integrar a la propuesta CRISP-DM el estándar para modelado de sistemas convencionales IEEE Std 1074 (aunque también integra elementos de la ISO 12207) con el objetivo de solventar la mayor debilidad identificada al estándar de facto para proyectos de explotación de información.

La figura 3, ilustra la estructura del modelo de proceso integrado, el cual se compone de 4 procesos generales: organizacional, administración del proyecto, desarrollo e integral.

La propuesta toma en su totalidad al estándar IEEE, alterando únicamente el núcleo de desarrollo con las etapas identificadas en CRISP-DM, e incorpora de la ISO 12207 los procesos de adquisición y suministro [21].

### III. DESCRIPCIÓN DEL PROBLEMA

En [21], se indica que la tasa de fracaso en proyectos pertenecientes a la disciplina supera el 60%, siendo CRISP-DM la metodología más utilizada.

A partir del desarrollo de variados proyectos de explotación de información, implementando los distintos modelos de procesos de explotación de información existentes, se identifican una serie de deficiencias y carencias para la ejecución exitosa del proyecto.

Al igual que los modelos de proceso de software tradicional, en sus primeras etapas se enfocaban en las actividades asociadas a la algoritmia y programación del software [21], los modelos de procesos de explotación de información, se centran en su actividad equivalente (Minería de datos).

El modelo de proceso KDD tiene como principal carencia la falta de especificaciones con respecto a las actividades necesarias para administrar un proyecto de explotación de información.

CRISP-DM si bien identifica algunas pocas actividades vinculadas a la administración de un proyecto, este está focalizado en el desarrollo de las actividades técnicas del proyecto, sin considerar aspectos como métricas, estimación, tercerización, gestión del cambio, entre otros. [21]

Si bien Marbán con su modelo de proceso integrado, busca subsanar las omisiones de su modelo base (CRISP-DM), el abordaje que propone no se ajusta a las necesidades de un proyecto de explotación de información, en el cual incluye etapas y actividades que no aplican a dicho tipo de proyectos, como son:

- La necesidad de construir un producto software como elemento asociado al desarrollo del proyecto de explotación de información. El producto de interés en este tipo de proyectos son las piezas de conocimiento, mientras que la construcción de un producto software que incorpore dichas piezas, corresponde a un proyecto externo el cual utiliza dicho conocimiento.
- Por consiguiente del párrafo anterior, la transición e instalación del sistema software, no es requerido en este tipo de proyectos. Las piezas de conocimiento obtenidas, son las que deben transmitirse y explicarse al cliente.
- El mantenimiento, no es una actividad aplicable a este tipo de proyectos, dado que como se mencionó en los puntos anteriores, no hay sistema informático que mantener, y además las piezas de conocimiento generadas, no pueden ser mantenidas dado que una vez implementadas dichas piezas en el negocio, su realidad se ve modificada, por lo cual los datos anteriores pierden su representatividad con el nuevo estado o situación en el que la organización se encuentra.

Adicionalmente, los modelos de procesos mencionados, presentan deficiencias en la estructuración de las tareas generales generando iteraciones innecesarias que causan demoras y aumentos en los costos. La preparación de los datos antes de definir las herramientas y algoritmos a utilizar conlleva a una posible necesidad de iterar nuevamente hacia la etapa de preparación de los datos, siendo posible identificar dichos elementos de manera previa. La selección de los algoritmos de minería de datos a utilizar es dejada a la experticia del ingeniero de minería de datos. Se omiten las vinculaciones y dependencias entre los elementos producidos a lo largo del proceso.

De las insuficiencias identificadas surge como problema la necesidad de desarrollar un modelo de proceso de desarrollo de proyectos de ingeniería de explotación de información que guie de forma precisa tanto en las cuestiones técnicas del proyecto, así como las de gestión del mismo, siendo necesario para ello identificar las fases, actividades, tareas, técnicas de representación y procedimientos de ejecución los cuales permitan sistematizar el desarrollo de proyectos en el área.

#### IV. NOTACIÓN UTILIZADA EN LA SOLUCIÓN

En esta sección se describen los distintos formalismos utilizados para describir cada uno de los subprocesos que integran al modelo de procesos de explotación de información propuesto. Como se mencionó con anterioridad, MoProPEI está integrado por 2 subprocesos. Las tareas pertenecientes al subproceso a describir, se identifican con rectángulos de borde en puntas rectas en cuyo interior se asigna el nombre de la tarea (figura 4). Las tareas del subproceso restante o secundario, se ilustran con rectángulos de borde en puntas rectas dividido en dos mitades horizontales, en cuya parte superior se identifica la actividad a la que pertenece y en la parte inferior el nombre de la tarea (figura 5).

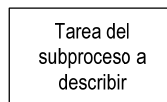


Fig. 4. Representación de la Tarea del subproceso a describir

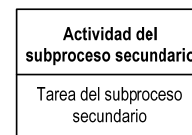


Fig. 5. Representación de la Tarea del subproceso secundario

Las actividades pertenecientes al subproceso a describir, se identifican con el formalismo ilustrado en la figura 7, mientras que la figura 8 representa a una actividad del subproceso secundario, la cual se vincula con alguna actividad del subproceso a describir.

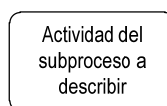


Fig. 6. Representación de la Actividad del subproceso a describir

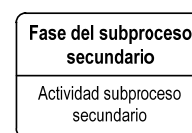


Fig. 7. Representación de la Actividad del subproceso secundario

Las actividades del subproceso a describir se ilustran con un rectángulo con bordes redondeados en cuyo interior se identifica el nombre de la actividad, mientras que las actividades del subproceso secundario, se ilustran con un rectángulo con bordes redondeados dividido en dos mitades horizontales, en cuya parte superior se identifica la fase a la que pertenece y en la parte inferior el nombre de la actividad.

Las actividades del subproceso a describir se agrupan a la fase que pertenecen utilizando llaves (figura 8), en donde todas las actividades definidas en el mismo nivel horizontal que la llave, pertenecen a la fase identificada en el lado externo de la misma.

Las tareas pertenecientes a una actividad a describir, se agrupan a la actividad que pertenecen utilizando corchetes (figura 9), en donde todas las tareas definidas en el mismo nivel horizontal que el corchete, pertenecen a la actividad identificada en el lado externo de la misma.

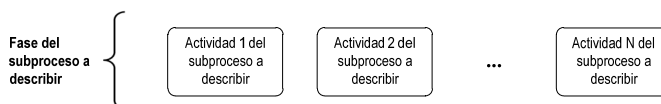


Fig. 8. Representación del agrupador de actividades para la fase del subproceso a describir

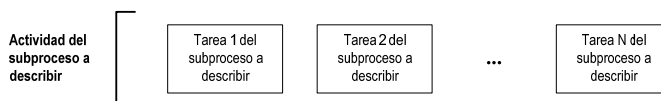


Fig. 9. Representación del agrupador de tareas para la actividad del subproceso a describir

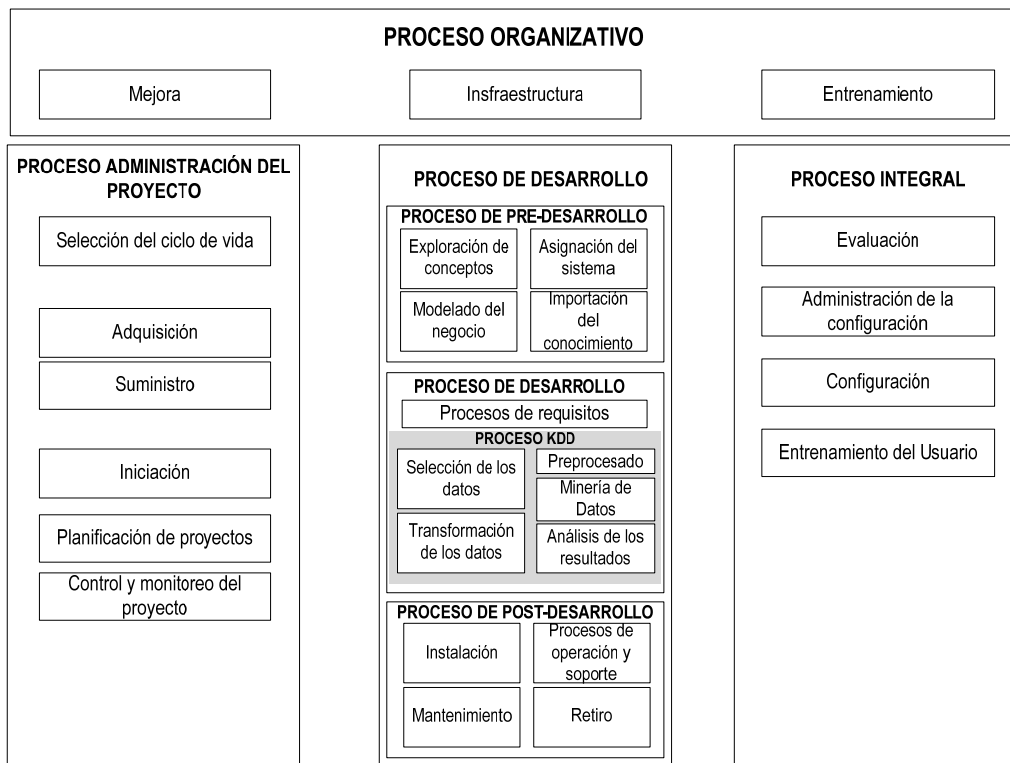


Fig. 3. Modelo de Proceso: Marbán [21]

Adicionalmente se identifican los elementos de entrada (inputs) y salida (outputs) que cada actividad y/o tarea posee, determinando las dependencias entre las mismas.

Se identifican 3 categorías de elementos:

- Externos: cuya proveniencia es ajena a la empresa que realiza el proyecto o los miembros que la conforman, los cuales se identifican mediante líneas intermitentes cuyos extremos son puntos rellenos (figura 10).
- Internos: elementos generados por la empresa que desarrolla el proyecto o miembros que la conforman, identificándose mediante líneas continuas cuyos extremos son puntos rellenos (figura 11).
- Del proyecto: elementos generados a partir de la ejecución de alguna tarea asociada al proyecto, identificado con una flecha cuya línea es continua si la actividad o tarea de donde proviene es anterior a la que la recibe como input (figura 12) o intermitente si es posterior (figura 13).

Se considera relevante señalar nuevamente que el orden en que se presentan las actividades, no es el orden de ejecución de las mismas. Adicionalmente, en las figuras 10 a 13 el uso de actividades o tareas es indistinto.

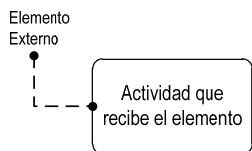


Fig. 10. Formalismo de representación para elementos externos

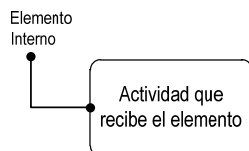


Fig. 11. Formalismo de representación para elementos Internos

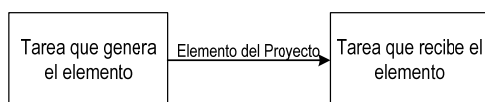


Fig. 12. Formalismo de representación para elementos del proyecto cuya tarea que lo genera es anterior a la tarea que lo recibe

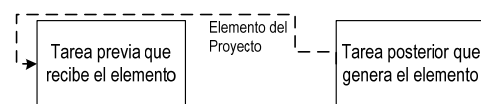


Fig. 13. Formalismo de representación para elementos del proyecto cuya tarea que lo genera es posterior a la tarea que lo recibe

## V. PROPUESTA DE SOLUCIÓN

En esta sección se describe el modelo de procesos de explotación de información propuesto. En la sección A se introducen los aspectos generales asociados al Modelo de Proceso para Proyectos de Explotación de Información (MoProPEI). En la sección B, se introduce la estructura general del mismo. En las últimas dos secciones se describen detalladamente los subprocesos que integran al modelo: el subproceso Gestión (sección C), asociado a las tareas de administración y control de un proyecto, y el subproceso Desarrollo (sección D), integrado por las tareas técnicas del proyecto.

### A. Aspectos Generales

Como ya se ha definido previamente en este artículo, la explotación de información se centra en la generación de piezas de conocimiento a partir de la implementación de algoritmos pertenecientes al campo de la Minería de Datos, u otros tipos de minería.

En [3] se define a la Minería de Datos, como una de las etapas que integran al proceso, la cual se basa en la ejecución de algoritmos sobre los datos para obtener patrones, mientras que el producto final del proceso es descubrir conocimiento útil a partir de los datos.

A partir de los párrafos previos, se desprende que el producto central de este tipo de proyectos son las piezas de **Conocimiento**. El éxito o fracaso del mismo no solo se basa en la generación de piezas de conocimiento nuevas e interesantes, sino también en la adecuada transmisión de las mismas al cliente. No basta con obtener patrones de la información, sino

que es necesario validarlos con el experto y adaptarlos para que estos puedan ser comprendidos por el demandante del proyecto (pudiendo o no tener experticia en el área abordada en el proyecto).

De forma complementaria, para alcanzar el éxito por parte de la empresa que desarrolla el proyecto, no solo es esencial satisfacer al cliente, sino también alcanzar dicho objetivo con el menor costo y plazo posible, focalizándose además en la documentación del nuevo conocimiento que puede surgir durante el desarrollo de un proyecto, utilizando este como valor agregado para futuros proyectos. Es decir, aprender de los errores previos y reutilizar las estrategias previamente desarrolladas.

La producción o adaptación de un producto software, como resultado de las piezas de conocimiento producidas en un proyecto de explotación de información, es un producto ajeno al desarrollo de este tipo de proyectos.

A partir de las piezas de conocimiento generadas se apalancan acciones correctivas en el dominio de estudio. Como consecuencia al hacer un posterior análisis, las acciones aplicadas generaron cambios en el dominio en cuestión, siendo necesario adaptar el proyecto a las nuevas realidades.

A partir de estos postulados, es que se concibe al modelo de proceso el cual presenta las siguientes características generales:

- Centrado en la generación del conocimiento: el objetivo final del proyecto en producir piezas de conocimiento relevantes para lo toma de decisiones.
- Adaptación a las necesidades del proyecto: permite ajustar la ejecución del proyecto a partir de las características que el mismo presenta.
- Enfocado en la gestión: fortalece la planificación, administración, documentación de todos los aspectos necesarios para el desarrollo de un proyecto de explotación de información.
- Reducción del trabajo redundante: a través de la planificación, y el ordenamiento de las tareas, reduce la cantidad de iteraciones necesarias para el desarrollo del proyecto.
- Orientado al Aprendizaje Continuo: genera, registra y reutiliza el conocimiento producido durante el desarrollo de los proyectos.
- Orientado a la sistematización: identifica y define las tareas a realizar de forma específica, contribuyendo a lograr el objetivo final.

### B. Estructura General

El modelo de proceso se estructura de forma jerárquica mediante cuatro niveles, cada uno de los cuales presenta un mayor nivel de especificidad. Estos niveles son:

- **Subprocesos:** Es la división de mayor nivel de generalidad, la cual está integrada por fases, y permite distinguir entre las tareas técnicas y las de gestión del proyecto.
- **Fases:** Subdivide a las tareas en base a sus finalidades generales. Estas están compuestas por actividades cuyos objetivos se encuentran asociados.
- **Actividades:** Subdivide a las tareas en base a sus objetivos específicos.
- **Tareas:** trabajo a realizarse el cual está asociado a un conjunto de técnicas posibles, las cuales generan a partir de un conjunto de elementos de entrada, uno o más elementos de salida.

Es importante aclarar que la distribución de las actividades en el modelo de proceso, no tiene como objetivo identificar el orden de aplicación, sino que el mismo se definirá a partir de las tareas destinadas a dicho propósito. De acuerdo al modelo de ciclo de vida seleccionado y las actividades oportunamente identificadas para generar el mapa de actividades, se establecerá el orden de ejecución e iteración de las tareas a realizar.

El modelo de proceso, como se mencionó previamente, se compone de dos subprocesos: Desarrollo, enfocado a las actividades técnicas y Gestión, orientado al control y la administración del proyecto.

Cada subproceso está compuesto por un conjunto de fases, las cuales agrupan a las actividades según las finalidades de las mismas (Figura 4).

La fase Gestión contiene cinco fases: Iniciación del Proyecto, Planificación del Proyecto, Soporte, Gestión del Control y Calidad, y Gestión de la Entrega.

La fase Desarrollo está integrada por seis fases: Entendimiento del Dominio, Entendimiento de los Datos, Modelado, Preparación de los Datos, Implementación, y Evaluación y Presentación.

La ejecución ambos subprocesos no es secuencial, sino que debe realizarse en paralelo, siendo el proceso de gestión un proceso de soporte para las actividades principales destinadas al desarrollo del producto final (piezas de conocimiento).

La figura 14, solo busca presentar los componentes generales del proceso, y las vinculaciones externas entre los dos subprocesos. Dicha imagen ilustra las vinculaciones y dependencias existentes entre las fases que componen a los subprocesos, identificando con una flecha la misma, cuyo extremo es un triángulo relleno, si dicha fase insume al menos un elemento generado por la fase donde se origina la flecha.

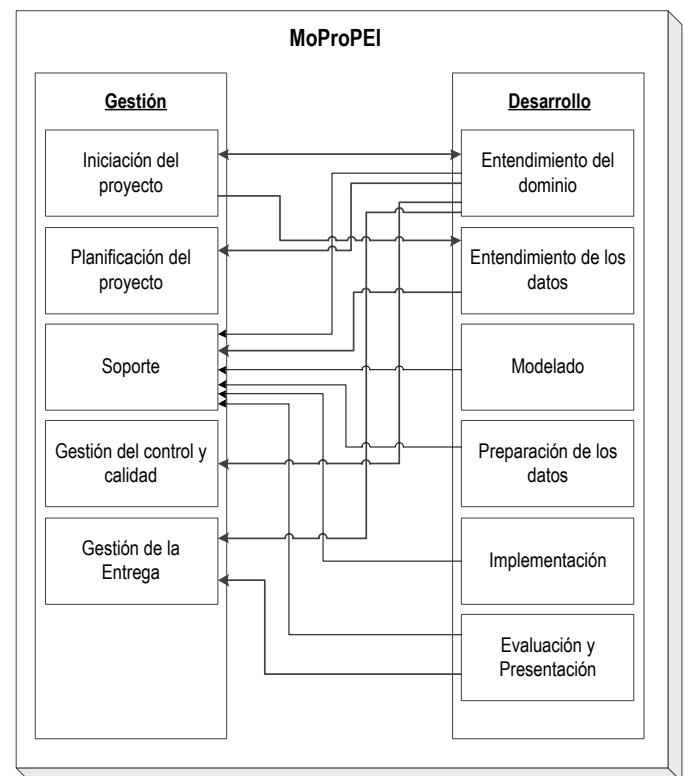


Fig. 14. Estructura General - MoProPEI

En caso que ambas fases tengan se retroalimenten, la relación contendrá un triángulo relleno en ambos extremos. La fase de entendimiento del negocio se relaciona con las fases

iniciación del proyecto, planificación del proyecto, soporte, gestión del control y calidad, y gestión de la entrega; las fases entendimiento de los datos, modelado, preparación de los datos e implementación se relacionan con la fase de soporte. La fase evaluación y presentación se vincula con las fases soporte y gestión de la entrega. De las fases de gestión, la única que se relaciona con fases de desarrollo es iniciación del proyecto, la cual se vincula con las fases entendimiento del dominio y entendimiento de los datos.

C. Sub-Proceso: Gestión

El subproceso de gestión abarca todas aquellas actividades asociadas a la interacción inicial con el cliente, la estructuración del proyecto, planificación y administración de los recursos, control y calidad del desarrollo del proyecto.

Como se mencionó en la sección anterior, la figura 4, presenta las relaciones y dependencias desde la perspectiva de las fases, abstrayéndose del elemento específicamente vinculado y las asociaciones internas entre las actividades pertenecientes a una misma fase. La figura 15, amplía la

descripción de las relaciones para las distintas actividades que componen al subproceso de gestión.

En el sector izquierdo de la figura 15 se observan las fases que componen al subproceso, y a su derecha en el mismo nivel horizontal se encuentran cada una de las actividades que las integran, junto con sus elementos de entrada y salida.

*Iniciación del proyecto* compuesto por 4 actividades: definición de la comunicación cuyos elementos de entrada son el discurso del cliente (externo), y la política de la organización contratante (externo), la política de la organización (interno), la experiencia de proyectos previos (interno) y el discurso del líder del proyecto (interno); exploración de conceptos iniciales, cuyos elementos de entrada son el discurso del cliente y el protocolo de comunicación externa; evaluación de la situación, alimentado por el análisis de los recursos existentes (externo) y el reporte de exploración inicial, producido en la actividad previa; y la definición del ciclo de vida, cuyas entradas son el documento de requisitos (actividad requisitos del subproceso de desarrollo) y Proyectos Guías, obtenido en la actividad previa.

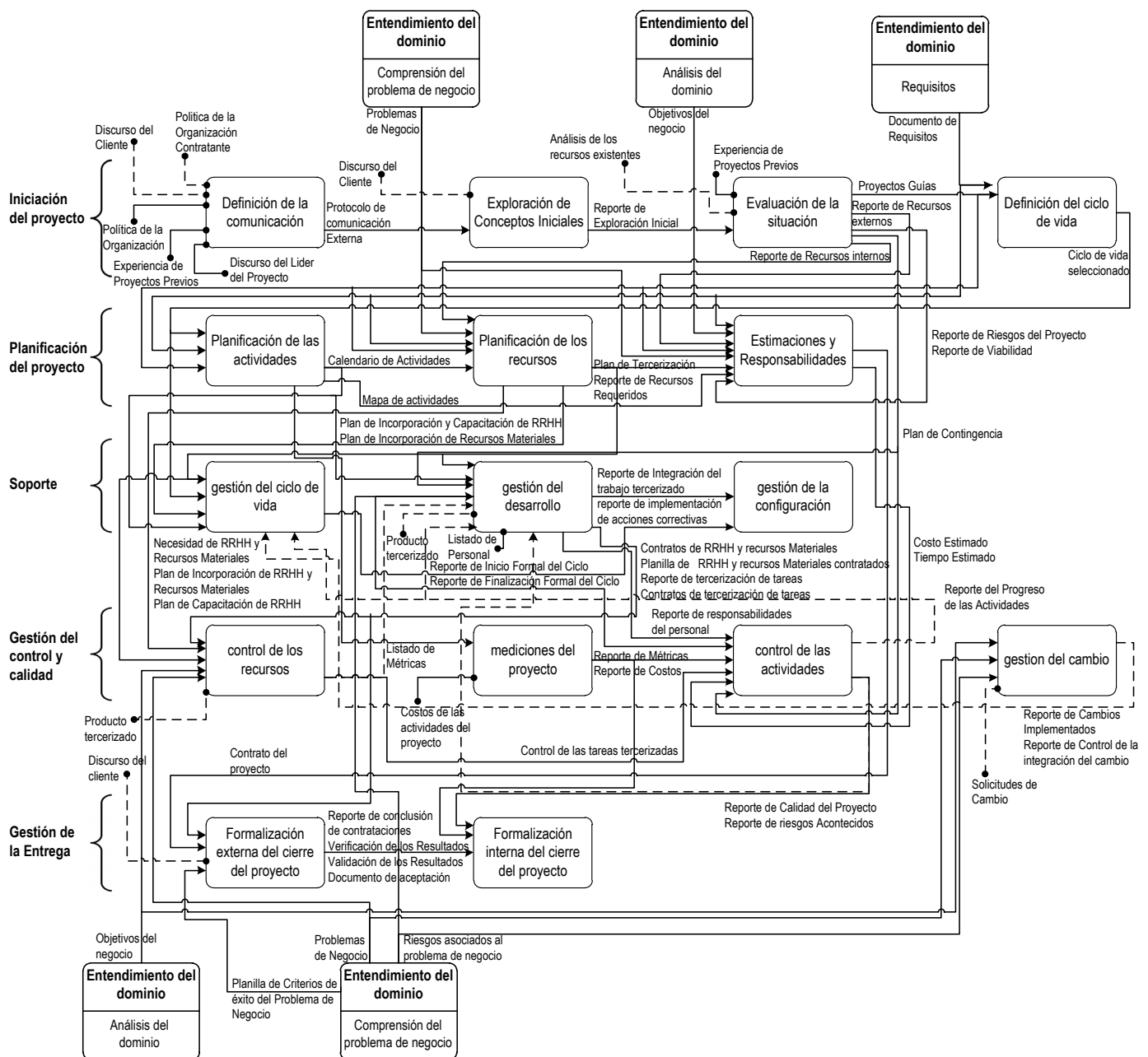


Fig. 15. Estructura del subproceso: Gestión

*Planificación del proyecto*, integrado por 3 actividades: planificación de las actividades, cuyos elementos de entrada son el ciclo de vida seleccionado (obtenido en la última actividad de la fase previa), el documento de requisitos (actividad requisitos del subproceso de desarrollo) y Proyectos Guías (obtenido en la actividad evaluación de la situación); planificación de los recursos, alimentado por el calendario de actividades (producido en la actividad previa), el problema de negocio (identificado en la actividad comprensión del problema de negocio perteneciente al subproceso desarrollo), el documento de requisitos (actividad requisitos del subproceso de desarrollo), el reporte de recursos internos (evaluación de la situación) y Proyectos Guías (obtenido en la actividad evaluación de la situación); y estimaciones y responsabilidades, cuyos elementos de entrada son el documento de requisitos (actividad requisitos del subproceso de desarrollo), los objetivos del negocio (identificados en la actividad análisis del dominio perteneciente al subproceso de desarrollo), el reporte de recursos externos, reporte de riesgos del proyecto, reporte de viabilidad y Proyectos Guías (producidos en la actividad evaluación de la situación), los problemas de negocio (identificado en la actividad comprensión del problema de negocio perteneciente al subproceso desarrollo), el mapa de actividades (planificación de las actividades), el plan de tercerización, y el reporte de recursos requeridos (producido en la actividad anterior).

*Soporte* compuesto por 3 actividades: gestión del ciclo de vida, cuyos elementos de entrada son el plan de tercerización y el reporte de recursos requeridos (planificación de los recursos), el ciclo de vida seleccionado (definición del ciclo de vida), los planes de incorporación y capacitación de recursos humanos y de incorporación de recursos materiales (planificación de los recursos) y el calendario de actividades (planificación de las actividades); gestión del desarrollo, cuyos elementos de entrada son el listado de personal (interno), el producto tercerizado (externo), los reportes de calidad del proyecto y de riesgos acontecidos, y el reporte del progreso de las actividades (producidos en una actividad posterior: control de las actividades), control de las tareas tercerizadas (producidos en una actividad posterior: control de los recursos), riesgos asociados al problema de negocio (identificados en la actividad comprensión del problema de negocio perteneciente al subproceso de desarrollo), plan de contingencia (evaluación de la situación), mapa de actividades (planificación de las actividades), el plan de tercerización y el reporte de recursos requeridos (planificación de los recursos); y gestión de la configuración, alimentada por los reportes de integración del trabajo tercerizado y de implementación de acciones correctivas (obtenidos en la actividad previa) y los reportes de inicio y finalización formal del ciclo (gestión del ciclo de vida).

*Gestión del control y calidad*, está conformada por 4 actividades: control de recursos, cuyos elementos de entrada son el producto tercerizado (externo), los problemas de negocio (generados en la actividad comprensión del problema de negocio perteneciente al subproceso de desarrollo), los objetivos del negocio (perteneciente a la actividad análisis del dominio del subproceso de desarrollo), Contratos de RRHH y recursos Materiales, la Planilla de RRHH y recursos Materiales contratados, el Reporte de tercerización de tareas y los Contratos de tercerización de tareas (gestión del desarrollo), Plan de Incorporación y Capacitación de RRHH, y Plan de Incorporación de Recursos Materiales (planificación de los recursos) y el plan de tercerización y el reporte de recursos

requeridos (planificación de los recursos); mediciones del proyecto cuyos elementos de entrada son el listado de métricas (planificación de las actividades) y los costos de las actividades del proyecto (interno); control de las actividades, alimentado por el reporte de responsabilidades del personal (gestión del desarrollo), los riesgos asociados al problema de negocio (comprensión del problema de negocio subproceso desarrollo), reporte de métricas y de costos (mediciones del proyecto), control de las tareas tercerizadas (control de recursos), costos y tiempo estimados (estimaciones y responsabilidades) y el plan de contingencia (evaluación de la situación); y gestión del cambio, cuyos elementos de entrada son las solicitudes de cambio (externo), el problema de negocio y los riesgos asociados al problema de negocio (comprensión del problema de negocio subproceso desarrollo) y los objetivos del negocio (análisis del dominio subproceso desarrollo).

*Gestión de la entrega*, integrada por 2 actividades: formalización externa del cierre del proyecto, cuyos elementos de entrada son el discurso del cliente (externo), la planilla de criterios de éxito del problema de negocio (comprensión del problema de negocio subproceso desarrollo), contrato del proyecto (estimaciones y responsabilidades) y los Contratos de RRHH y de recursos Materiales, la Planilla de RRHH y de recursos Materiales contratados, el Reporte de tercerización de tareas, y los Contratos de tercerización de tareas (gestión del desarrollo); y formalización interna del cierre del proyecto, cuyos elementos de entrada son el Reporte de conclusión de contrataciones, la Verificación y validación de los Resultados y el Documento de aceptación (obtenidos en la actividad previa), los reportes de calidad del proyecto y de riesgos acontecidos (control de actividades) y los reportes de métricas y costos (mediciones del proyecto).

Las relaciones identificadas en la figura 15 se abstraen de las vinculaciones internas existentes entre las tareas que integran a una misma actividad, es decir, se trata la actividad como una caja negra.

En los siguientes párrafos se procede a describir cada una de las actividades previamente mencionadas, identificando sus objetivos y las tareas que la componen.

### *C.1. Iniciación del Proyecto*

La fase de iniciación del proyecto está compuesta por 4 actividades: definición de la comunicación, exploración de conceptos iniciales, evaluación de la situación y definición del ciclo de vida.

*Definición de la comunicación*: tiene como objetivo definir los medios formales que serán utilizados para realizar las comunicaciones (tanto orales como escritas) entre las partes intervinientes en el proyecto. Dicha actividad está integrada por 4 tareas: definir el protocolo de documentación externa, definir el protocolo de comunicación externa, definir el protocolo de documentación interna y definir el protocolo de comunicación interna.

*Exploración de Conceptos Iniciales*: cuyo objetivo es realizar un primer estudio del negocio y las necesidades del cliente. Dicha actividad está compuesta por 3 tareas: planificar la adquisición de conocimiento, implementar técnicas de adquisición de conocimiento, y generar el reporte de exploración inicial.

*Evaluación de la situación*: en la cual se analizan las características y requerimientos del proyecto, las posibilidades de la organización para realizar el mismo y se determina su factibilidad. La actividad está compuesta por 7 tareas:



Evaluación de proyectos previos, Identificación de los recursos externos, Identificación de los recursos internos, Identificación de riesgos del proyecto, Definición del plan de contingencia, Identificar la posibilidad de tercerización y Determinación de la Viabilidad del proyecto.

*Definición del ciclo de vida:* cuyo objetivo es identificar las posibles formas de *estructurar* el proyecto, las cuales favorezcan el desarrollo y éxito del proyecto. La actividad está integrada por 2 tareas: identificación de alternativas de ciclos de vidas y selección del ciclo de vida.

### C.2. Planificación del proyecto

La fase planificación del proyecto está compuesta por 3 actividades: planificación de las actividades, planificación de los recursos y estimaciones y responsabilidades.

*Planificación de las actividades:* tiene como objetivo seleccionar y determinar el tiempo en el cual se implementarán *cada* una de las tareas asociadas al proyecto, de acuerdo a sus características. Esta actividad está compuesta por 3 tareas: definir las actividades asociadas al proyecto, identificar las métricas a realizar y establecer el calendario de actividades.

*Planificación de los recursos:* su objetivo es identificar, determinar y planificar que recursos el proyecto requerirá, a partir de sus características, estableciendo el modo mediante el cual se satisfarán dichas necesidades. Dicha actividad está integrada por 6 tareas: planificar la necesidad de recursos, planificación de la tercerización, prever los recursos, planificar la incorporación de RRHH, planificar la capacitación de RRHH y planificar la adquisición de recursos materiales.

*Estimaciones y Responsabilidades:* establece los tiempos y costos del proyecto, y los alcances y obligaciones de las partes intervinientes en el proyecto. Dicha actividad está compuesta por 3 tareas: estimar el tiempo de desarrollo del proyecto, estimar el costo del proyecto y definir las responsabilidades de las partes.

### C.3. Soporte

La fase soporte está compuesta por 3 actividades: gestión del ciclo de vida, gestión del desarrollo y gestión de la configuración.

*Gestión del Ciclo de Vida:* tiene como objetivo formalizar los alcances de cada iteración del proyecto y establecer los logros alcanzados y reajustar los alcances de la próxima iteración en caso que fuese necesario. La actividad está integrada por 2 tareas: formalizar el inicio del ciclo y formalizar el cierre del ciclo.

*Gestión del Desarrollo:* abarca aquellas actividades asociadas con la ejecución de tareas para el desarrollo del producto, las cuales abarcan distintos aspectos como establecer los objetivos a realizar por cada una de las partes que desarrollarán el proyecto, integrar los subproductos resultantes, comunicar los avances del proyecto, entre otros. Las tareas que componen a dicha actividad son 6: definir responsabilidades del personal, definir tercerización de tareas, definir las contrataciones de recursos, integración del trabajo tercerizado, comunicación del progreso de las actividades e implementar acciones correctivas.

*Gestión de la Configuración:* está orientada a documentar la información relevante del proyecto, así como mantener su trazabilidad. Dicha actividad está integrada por 2 tareas: integrar los resultados al documento del proyecto y mantener el versionado del documento.

### C.4. Gestión del control y calidad

La fase gestión del control y calidad está compuesta por 4 actividades: planificación de las actividades, planificación de los recursos, estimaciones y responsabilidades y estimaciones y responsabilidades.

*Control de los Recursos:* cuyo objetivo es comprobar el cumplimiento de las planificaciones realizadas respecto a la incorporación de recursos (materiales, productos y humanos). Dicha actividad está integrada por 4 tareas: controlar la adquisición de recursos materiales, controlar la incorporación de RRHH, controlar la capacitación de RRHH y controlar el cumplimiento de tareas tercerizadas.

*Mediciones del Proyecto:* tiene como objetivo cuantificar los elementos de interés para el análisis del proyecto. Dicha actividad abarca las siguientes 2 tareas: implementar métricas del proyecto y control de los costos del proyecto.

*Control de las Actividades:* tiene como objetivo evaluar las distintas características del proyecto (progreso de las actividades, riesgos y calidad) con el fin de identificar a tiempo posibles desvíos que pongan en riesgo el desarrollo exitoso del proyecto. Dicha actividad está integrada por 3 tareas: control de ejecución de actividades, análisis y control de riesgos y control de calidad del proyecto.

*Gestión del Cambio:* abarca todas las actividades orientadas a evaluar, implementar y controlar los cambios requeridos durante el desarrollo del proyecto. Dicha actividad está integrada por 3 tareas: evaluar las solicitudes de cambio, implementación del cambio y control del cambio.

### C.5. Gestión de la Entrega

La fase gestión de la entrega está compuesta por 2 actividades: formalización externa del cierre del proyecto y formalización interna del cierre del proyecto.

*Formalización Externa del Cierre del Proyecto:* tiene como objetivo garantizar el logro de los objetivos requeridos por el cliente y finalizar con los compromisos entre las partes intervinientes en el proyecto. Dicha actividad está integrada por 2 tareas: verificación y validación del proyecto y finalización de los contratos.

*Formalización Interna del Cierre del Proyecto:* su objetivo es concluir con las tareas asociadas al análisis y evaluación del desarrollo del proyecto, con la finalidad que sirvan como información útil para futuros proyectos. Dicha actividad está integrada por 2 tareas: generar el reporte de mediciones y sugerencias de mejora.

La tabla 1 describe en detalle cada una de las tareas que integran al subproceso gestión mediante la siguiente estructura:

- *Objetivos:* Se describen sus metas generales, es decir, el trabajo que se realiza para generar los elementos de salida, a partir de los inputs.
- *Inputs:* Se describen los insumos que requiere la tarea para ser iniciada.
- *Outputs:* Son los elementos de salida, los cuales pueden ser insumos de otras tareas.

### D. Sub-Proceso: Desarrollo

El proceso de desarrollo abarca todas las actividades asociadas a la identificación de patrones relevantes para la organización, de calidad y novedosas, así como su análisis y comprensión para la generación de piezas de conocimiento validables e interesantes que aporten valor agregado a la misma.

TABLA I. LISTADO DE TAREAS PERTENECIENTES AL SUBPROCESO DE GESTIÓN

Fase	Actividad	Tarea	Objetivo	INPUT	OUTPUT	
Iniciación del proyecto	Definición de la comunicación	Definir Protocolo de documentación externa	Determinar los formalismos que se deberán utilizar para toda la documentación que se entregará al cliente generado a lo largo del desarrollo del proyecto, adaptando las mismas a las necesidades y políticas del cliente	Discurso del Cliente Política de la Organización Contratante	Protocolo de documentación Externa	
		Definir Protocolo de comunicación externa	Pautar las formas/canales de comunicación con el cliente, que se deberán utilizar a lo largo del desarrollo del proyecto, de acuerdo a las posibilidades/necesidades y políticas del cliente.	Discurso del Cliente Política de la Organización Contratante	Protocolo de comunicación Externa	
		Definir Protocolo de documentación interna	Determinar los formalismos que se deberán utilizar para toda la documentación interna del desarrollo del proyecto. Los mismos pueden haber sido establecidos y utilizados en proyectos anteriores.	Política de la Organización Experiencia de Proyectos Previos Discurso del Líder del Proyecto	Protocolo de documentación Interna	
		Definir Protocolo de comunicación interna	Pautar las formas/canales de comunicación internas, que se utilizarán a lo largo del proyecto. Las mismas pueden haber sido establecidas e implementadas previamente al proyecto.	Política de la Organización Experiencia de Proyectos Previos Discurso del Líder del Proyecto	Protocolo de comunicación Interna	
	Exploración de Conceptos Iniciales	Planificar la adquisición de conocimiento	Realizar un estudio superficial de las características de la organización contratante, identificando conceptos relevantes que guíen las posteriores actividades de educación, análisis y comprensión del negocio. Adicionalmente busca adentrarse en los conceptos teóricos relevantes para el dominio del negocio.	Discurso del cliente Conceptos teóricos asociados al dominio Estudio de la organización Protocolo de comunicación Externa	Plan de Adquisición de conocimiento	
		Implementar técnicas de adquisición de conocimiento	Implementar la planificación realizada en la tarea previa, con el objetivo de comprender en detalle las características de la organización contratante.	Plan de Adquisición de conocimiento Protocolo de comunicación Externa	Conocimiento Adquirido	
		Generar el reporte de exploración Inicial	Documentar todos los conocimientos relevantes adquiridos en las tareas previas, como base para realizar las tareas posteriores del proyecto.	Discurso del cliente Conceptos teóricos asociados al dominio Estudio de la organización Conocimiento Adquirido	Reporte de Exploración Inicial	
	Evaluación de la situación	Evaluación de proyectos previos	Reutilizar el aprendizaje realizado en proyectos previos para resolver los desafíos asociados al proyecto en progreso, sirviendo de guías para la ejecución de las futuras actividades del mismo.	Reporte de Exploración Inicial Experiencia de Proyectos Previos	Proyectos Guías	
		Identificación de los recursos externos	Determina los recursos, tanto materiales como humanos, existentes en la organización contratante, de interés para el desarrollo del proyecto.	Reporte de Exploración Inicial Análisis de los recursos existentes	Reporte de Recursos externos	
		Identificación de los recursos internos	Determina los recursos, tanto materiales como humanos, existentes en la organización que realiza el proyecto, de interés para el desarrollo del mismo.	Reporte de Exploración Inicial Análisis de los recursos existentes	Reporte de Recursos internos	
		Identificación de riesgos del proyecto	Determinar y caracterizar los riesgos existentes en el desarrollo del proyecto, de acuerdo a las particularidades del negocio y de la experiencia acumulada en proyectos previos.	Reporte de Exploración Inicial Proyectos Guías Reporte de Recursos externos	Reporte de Riesgos del Proyecto	
		Definición del plan de contingencia	Determinar posibles acciones a tomar en caso que ocurran los riesgos previstos, con el objetivo de subsanar, en la medida de lo posible, los daños asociados a dicho riesgo.	Reporte de Exploración Inicial Proyectos Guías Reporte de Riesgos del Proyecto Riesgos asociados al problema de negocio	Plan de Contingencia	
		Identificar la posibilidad de tercerización	Determinar la posibilidad de que un equipo externo a la organización realice el proyecto o parte del mismo. Dicha necesidad puede surgir por distintos motivos: reducir costos o tiempos del proyecto, carencia de recursos (humanos y/o materiales), entre otros.	Reporte de Exploración Inicial Proyectos Guías Reporte de Recursos externos	Reporte de Posibilidad de Tercerización	
		Determinación de la Viabilidad del proyecto	Definir la posibilidad de éxito del proyecto, detallando el nivel de factibilidad del proyecto y el nivel de riesgos asociados. En esta tarea se definirá si se continua o no con el proyecto.	Reporte de Exploración Inicial Proyectos Guías Reporte de Recursos externos Reporte de Posibilidad de Tercerización	Reporte de Viabilidad	
		Definición del ciclo de vida	identificación de alternativas de ciclos de vidas	Determinar las posibles estructuras del proyecto que favorecen el desarrollo del mismo, definiendo el orden y la transición en la cual las fases serán realizadas a lo largo del proyecto, identificando además iteraciones entre las mismas.	Documento de Requisitos	Alternativas de ciclo de vida
					Proyectos Guías	

		selección del ciclo de vida	Determinar la mejor solución de las identificadas en la tarea previa, de acuerdo a cuál de las alternativas se ajusta más a las necesidades del proyecto.	Alternativas de ciclo de vida	Ciclo de vida seleccionado
Planificación del proyecto	Planificación de las actividades	Definir las actividades asociadas al proyecto	Seleccionar el conjunto de actividades necesarias para el desarrollo exitoso del proyecto, descartando aquellas que fueran irrelevantes.	Documento de Requisitos	Mapa de actividades
				Proyectos Guías	
				Ciclo de vida seleccionado	
		Identificar las Métricas a realizar	Determinar el conjunto de características que serán medidas y evaluadas en el transcurso del proyecto, así como aquellas que serán utilizadas a posterior.	Documento de Requisitos	Listado de Métricas
				Proyectos Guías	
		Establecer el calendario de actividades	Planificar el tiempo en el cual se realizarán cada una de las actividades asociadas al proyecto.	Documento de Requisitos	Calendario de Actividades
	Proyectos Guías				
	Mapa de actividades				
	Ciclo de vida seleccionado				
	Planificación de los recursos	planificar la necesidad de recursos	Determinar la necesidad de recursos, tanto materiales como humanos, para el desarrollo del proyecto.	Documento de Requisitos	Reporte de Recursos Requeridos
				Proyectos Guías	
				Calendario de Actividades	
		Planificación de la tercerización	Determinar los tiempos de inicio y entrega de los productos tercerizados.	Documento de Requisitos	Plan de Tercerización
				Proyectos Guías	
				Calendario de Actividades	
				Reporte de Posibilidad de Tercerización	
		prever los recursos	Determinar la necesidad de incorporar recursos en base a las características del proyecto y los recursos actualmente disponibles	Reporte de Recursos Requeridos	Necesidad de Recursos Materiales
				Plan de Tercerización	Necesidad de Recursos Humanos
				Reporte de Recursos internos	
		planificar la Incorporación de RRHH	Planear la contratación de recursos humanos de acuerdo a las necesidades del proyecto y experiencias previas	Documento de Requisitos	Plan de Incorporación de RRHH
				Proyectos Guías	
Calendario de Actividades					
Reporte de Recursos Requeridos					
Necesidad de Recursos Humanos					
planificar la Capacitación de RRHH	Planear la capacitación del recurso humano acorde a las necesidades del proyecto y experiencias previas	Documento de Requisitos	Plan de Capacitación de RRHH		
		Proyectos Guías			
		Calendario de Actividades			
		Reporte de Recursos Requeridos			
		Necesidad de Recursos Humanos			
		Plan de Incorporación de RRHH			
planificar la Adquisición de recursos materiales	Planear la compra de recursos materiales en base a las necesidades del proyecto y experiencias previas	Documento de Requisitos	Plan de Incorporación de Recursos Materiales		
		Proyectos Guías			
		Calendario de Actividades			
		Reporte de Recursos Requeridos			
		Necesidad de Recursos Materiales			
Estimaciones y Responsabilidades	Estimar el tiempo de desarrollo del Proyecto	Determinar el tiempo necesario para el desarrollo del proyecto a partir de las características del mismo y de experiencias previas	Documento de Requisitos	Tiempo Estimado	
			Proyectos Guías		
			Mapa de actividades		
			Reporte de Recursos externos		
			Objetivos del negocio		
	Estimar el costo del proyecto	Determinar el costo de desarrollo del proyecto en base a las características del mismo y experiencias previas	Determinar el costo de desarrollo del proyecto en base a las características del mismo y experiencias previas	Documento de Requisitos	Costo Estimado
				Proyectos Guías	
				Mapa de actividades	
				Objetivos del negocio	
				Problemas de Negocio	
Definir las responsabilidades de las partes	Definir los derechos y obligaciones de cada una de las partes que intervienen en el desarrollo del proyecto.	Definir los derechos y obligaciones de cada una de las partes que intervienen en el desarrollo del proyecto.	Documento de Requisitos	Contrato del proyecto	
			Proyectos Guías		
			Objetivos del negocio		

				Problemas de Negocio	
				Tiempo Estimado	
				Costo Estimado	
				Reporte de Riesgos del Proyecto	
				Reporte de Viabilidad	
				Plan de Tercerización	
Soporte	gestión del ciclo de vida	formalizar el inicio del ciclo	Definir formalmente el inicio del desarrollo del proyecto o una iteración del mismo, estableciendo alcances del mismo.	Calendario de Actividades	Reporte de Inicio Formal del Ciclo
				Plan de Tercerización	
			Ciclo de vida seleccionado		
			Plan de Incorporación de RRHH		
			Plan de Capacitación de RRHH		
			Plan de Incorporación de Recursos Materiales		
			Reporte de Finalización Formal del Ciclo		
		formalizar el cierre del ciclo	Definir formalmente el fin del desarrollo del proyecto o iteración, estableciendo las metas alcanzadas y el estado de las no alcanzadas.	Reporte de Inicio Formal del Ciclo	Reporte de Finalización Formal del Ciclo
				Reporte del Progreso de las Actividades	
				Reporte de Cambios Implementados	
				Reporte de Control de la integración del cambio	
	gestión del desarrollo	definir responsabilidades del personal	Determinar las responsabilidades de cada miembro del equipo a lo largo del proyecto.	Calendario de Actividades	Reporte de responsabilidades del personal
				Reporte de Recursos Requeridos	
				Listado de Personal	
		definir tercerización de tareas	Establecer formalmente las obligaciones, responsabilidades y alcances de cada una de las partes intervinientes en la tercerización del producto.	Calendario de Actividades	Contratos de tercerización de tareas
				Plan de Tercerización	Reporte de tercerización de tareas
		definir las contrataciones de recursos	Establecer formalmente la contratación de recursos, tanto materiales como humanos, para el desarrollo del proyecto.	Plan de Incorporación de RRHH	Contratos de recursos Materiales
					Planilla de recursos Materiales contratados
				Plan de Incorporación de Recursos Materiales	Contratos de RRHH
					Planilla de RRHH contratados
integración del trabajo tercerizado		Incorporar los productos tercerizados al proyecto, verificando la correcta integración de los mismos.	Producto tercerizado	Resultados Integrados	
			Documentación del Producto tercerizado		
			Plan de Tercerización	Reporte de Integración del trabajo tercerizado	
		Reporte de tercerización de tareas			
		Control de las tareas tercerizadas			
Comunicación del progreso de las actividades	Formalizar los progresos del proyecto a las partes involucradas en el mismo.	Reporte del Progreso de las Actividades	Comunicar el progreso al equipo/cliente		
Implementar acciones correctivas	Implementar acciones que modifiquen el curso del proyecto a causa de los riesgos acontecidos	Reporte de Riesgos del Proyecto	reporte de implementación de acciones correctivas		
		Plan de Contingencia			
		Riesgos asociados al problema de negocio			
		Reporte de riesgos Acontecidos			
gestión de la configuración	integrar los resultados al documento del proyecto	Establecer todos los resultados generados a lo largo del proyecto en un documento único que facilite la administración, el análisis y la sistematización del desarrollo del proyecto	Versionado de la documentación del proyecto	documentación del proyecto	
	mantener el versionado del documento	Determinar el estado actual del sistema, permitiendo mantener la trazabilidad de los cambios realizados a lo largo del proyecto	Reporte de Inicio Formal del Ciclo	Versionado de la documentación del proyecto	
		Reporte de Finalización Formal del Ciclo			
		Reporte de Integración del trabajo tercerizado			
		reporte de implementación de acciones correctivas			
Gestión del control y calidad	control de los recursos	controlar la adquisición de recursos materiales	Contratos de recursos Materiales	Control de la adquisición de recursos materiales	
			Planilla de recursos Materiales contratados		
			Necesidad de Recursos Materiales		
			Plan de Incorporación de Recursos Materiales		
	controlar la incorporación de RRH	Controlar la adquisición de los recursos humanos con respecto a lo planificado, permitiendo identificar desvíos	Contratos de RRHH	Control de la incorporación de RRHH	
			Planilla de RRHH contratados		
Plan de Incorporación de RRHH					
		Necesidad de Recursos Humanos			
controlar la capacitación de RRH	Controlar la capacitación de los recursos humanos con respecto a lo planificado, permitiendo identificar desvíos	Plan de Capacitación de RRHH	Control de la capacitación de RRHH		
		Necesidad de Recursos Humanos			

Gestión de la Entrega	mediciones del proyecto	controlar el cumplimiento de tareas tercerizadas	Controlar el progreso de los productos tercerizados en base a lo planificado, permitiendo identificar desvíos	Producto tercerizado Documentación del Producto tercerizado Plan de Tercerización Reporte de tercerización de tareas Contratos de tercerización de tareas Objetivos del negocio Problemas de Negocio	Control de las tareas tercerizadas	
		Implementar Métricas del proyecto	Documentar los valores de las características a analizar durante el desarrollo del proyecto	Listado de Métricas	Reporte de Métricas	
		control de los costos del proyecto	Documentar los costos del proyecto	Costos de las actividades del proyecto	Reporte de Costos	
		control de las actividades	control de ejecución de actividades	Documentar el estado de las tareas realizadas en base a lo planificado	Reporte de responsabilidades del personal Calendario de Actividades Control de las tareas tercerizadas	Reporte del Progreso de las Actividades
			Análisis y control de riesgos	Evaluar los posibles riesgos y desvíos del proyecto, permitiendo identificarlos y corregirlos reduciendo el impacto que estos generen	Calendario de Actividades	Reporte de riesgos Acontecidos
					Costo Estimado	
	Tiempo Estimado					
	Reporte de Métricas					
	Reporte de Costos					
	Reporte del Progreso de las Actividades					
	control de calidad del proyecto		Evaluar el estado del proyecto en base a las métricas definidas	Plan de Contingencia	Reporte de Calidad del Proyecto	
				Control de las tareas tercerizadas		
		Reporte de Métricas				
	gestión del cambio	Evaluar las solicitudes de cambio	Analizar la correspondencia de los pedidos de cambio en base a las características del proyecto	Solicitudes de Cambio	Reporte de solicitudes de Cambios	
				Documento de Requisitos	Documento de Requisitos	
				Problemas de Negocio	Problemas de Negocio	
				Riesgos asociados al problema de negocio	Riesgos asociados al problema de negocio	
		implementación del cambio	Documentar la ejecución de las acciones correctivas	Reporte de solicitudes de Cambios	Reporte de Cambios Implementados	
		control del cambio	Evaluar los cambios implementados garantizando que no introduzcan inconvenientes en el desarrollo del proyecto		Reporte de solicitudes de Cambios	Reporte de Control de la integración del cambio
	Reporte de Cambios Implementados					
Documento de Requisitos						
Problemas de Negocio						
Formalización externa del cierre del proyecto	Verificación y Validación del proyecto	Evaluar el proyecto con respecto a los criterios de éxito, los alcances y lo requerido por el cliente	Contrato del proyecto	Documento de aceptación		
			Planilla de Criterios de éxito del Problema de Negocio			
			Discurso del cliente			
	Finalización de los contratos	Formalizar la finalización de las contrataciones realizadas		Contratos de tercerización de tareas	Reporte de conclusión de contrataciones	
				Contratos de recursos Materiales		
				Contratos de RRHH		
				Contrato del proyecto		
	Formalización interna del cierre del proyecto	Generar el reporte de mediciones	Formalizar los resultados obtenidos en base a las mediciones realizadas	Reporte de Calidad del Proyecto	Reporte interno de desarrollo del proyecto	
				Reporte de Métricas		
				Reporte de Costos		
sugerencias de mejora		Formalizar dificultades enfrentadas durante el desarrollo del proyecto, las acciones evaluadas/ejecutadas y los resultados obtenidos		Reporte interno de desarrollo del proyecto	Reporte de sugerencia de mejoras	
				Reporte de conclusión de contrataciones		
				Verificación de los Resultados		
				Validación de los Resultados		
Documento de aceptación	Reporte de Calidad del Proyecto	Reporte de riesgos Acontecidos				

Como se mencionó en la sección anterior, la figura 4, presenta las relaciones y dependencias desde la perspectiva de las fases, abstractándose del elemento específicamente vinculado y las asociaciones internas entre las actividades pertenecientes a una misma fase. La figura 16, amplía dicho concepto para las distintas actividades que componen al subproceso de desarrollo.

En la parte izquierda de la figura 16 se observan las fases que componen al subproceso, y a su derecha en la misma posición horizontal cada una de las actividades que las integran, junto con sus elementos de entrada y salida.

*Entendimiento del dominio*, compuesta por 3 actividades: *requisitos*, que tiene como entrada el discurso del experto, elemento externo obtenido por el experto en el área del negocio, y el reporte de exploración inicial, obtenido en el subproceso de gestión, fase de iniciación del proyecto, y actividad exploración de conceptos iniciales; análisis del dominio, cuyo elemento de entrada es el documento de requisitos producido en la actividad previa; y comprensión del

problema de negocio, que se alimenta del documento de requisitos y de los objetivos del negocio identificados en la actividad previa.

*Entendimiento de los datos*, compuesta por 3 actividades: análisis de los datos, cuyos elementos de entrada son el problema de negocio, obtenido en la última actividad de la fase anterior, el reporte de recursos externo, generado por la actividad evaluación de la situación perteneciente al subproceso de gestión y la fuente de datos (externo); exploración de los datos, que se alimenta de el diagrama de relación de fuentes de datos, la descripción de las fuentes de datos, los campos relacionados con el problema de negocio, la descripción de los campos (todos elementos de la actividad previa) y las fuentes de datos no digitales (externo); y evaluación de los datos, cuyos elementos de entrada son el reporte de datos explorados, obtenido en la actividad previa, y los campos relacionados con el problema de negocio, producido en la primera actividad de la fase.

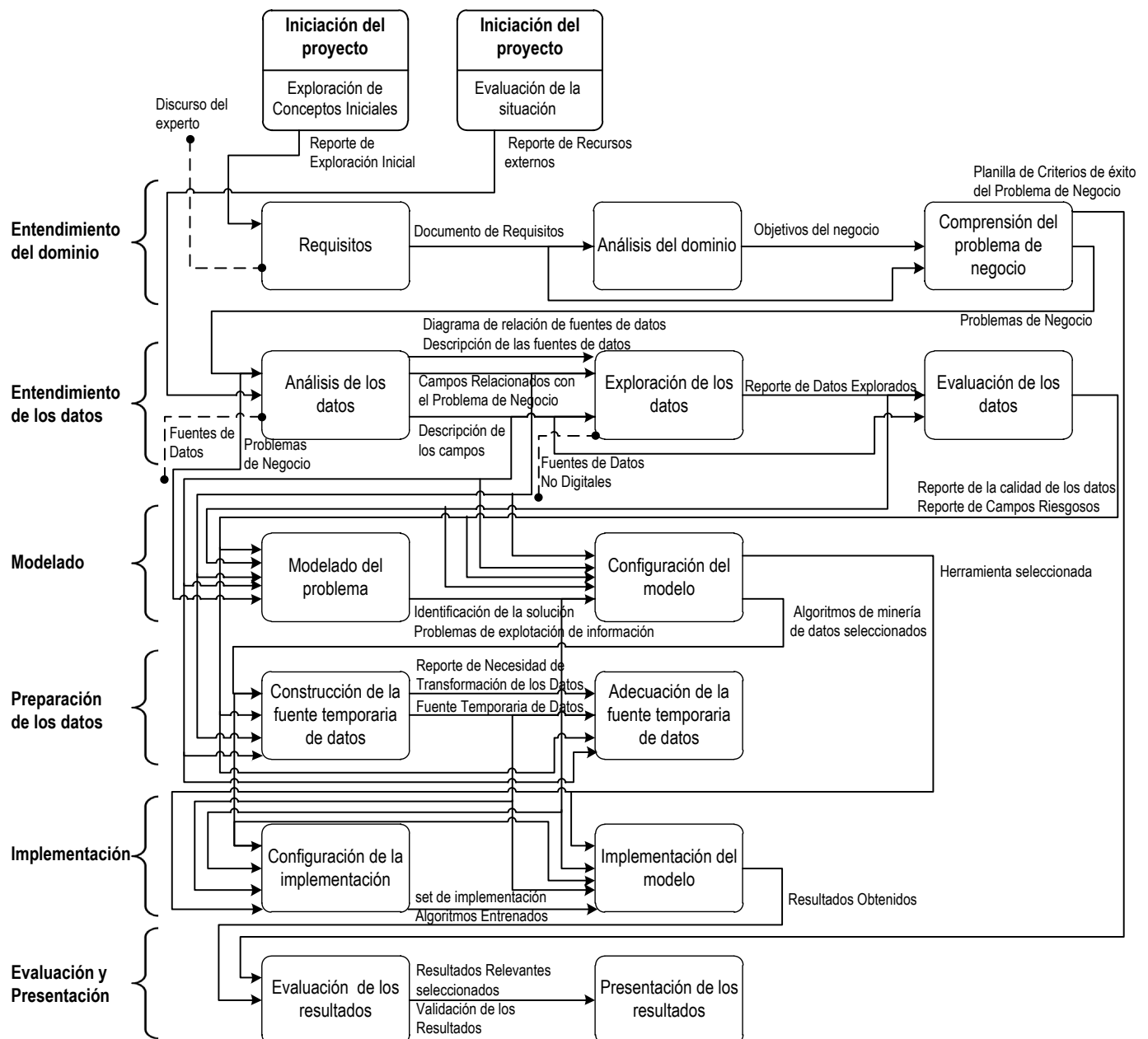


Fig. 16. Estructura del subproceso: Desarrollo

*Modelado* compuesta por 2 actividades: Modelado del problema, cuyos elementos de entrada son los reportes de calidad de los datos y de campos riesgosos, obtenidos en la última actividad de la fase anterior, el reporte de datos explorados, producido por la actividad de exploración de los datos, campos relacionados con el problema de negocio y descripción de los campos, producidos por la primera actividad de la fase previa y el problema de negocio, generado por la última actividad de la primera fase; y configuración del modelo, cuyos elementos de entrada son los reportes de calidad de los datos y de campos riesgosos, obtenidos en la última actividad de la fase anterior, el reporte de datos explorados, producido por la actividad de exploración de los datos, campos relacionados con el problema de negocio y descripción de los campos, producidos por la primera actividad de la fase previa y la identificación de la solución y los problemas de explotación de información, obtenidos de la actividad previa.

*Preparación de los datos*, conformada por 2 actividades: Construcción de la fuente temporaria de datos, alimentada por los algoritmos de minería de datos seleccionados (última actividad de la fase previa), los reportes de calidad de los datos y de campos riesgosos, producidos por la última actividad de la fase entendimiento de los datos, y los campos relacionados con el problema de negocio y la descripción de los campos, producidos por la primera actividad de la fase entendimiento de los datos y el problema de negocio; y adecuación de la fuente temporaria de datos, la cual recibe como entrada la fuente de temporaria de datos y el reporte de necesidad de transformación de los datos (obtenidos en la actividad anterior), los reportes de calidad de los datos y campos riesgosos (producidos en la actividad evaluación de los datos), y los campos relacionados con el problema de negocio (generado en la actividad de análisis de los datos).

*Implementación*, integrado por 2 actividades: configuración de la implementación, cuyos elementos de entrada son los algoritmos de minería de datos seleccionados y la herramienta seleccionada (configuración del modelo), identificación de la solución y los problemas de explotación de información (modelado del problema), fuente temporaria de datos (construcción de la fuente temporaria de datos); e implementación del modelo, alimentado por el set de implementación y los algoritmos entrenados, generados en la actividad previa, identificación de la solución y los problemas de explotación de información (modelado del problema), los algoritmos de minería de datos seleccionados y la herramienta seleccionada (configuración del modelo) y la fuente temporaria de datos (construcción de la fuente temporaria de datos).

*Evaluación y Presentación*, compuesta por 2 actividades: evaluación de los resultados cuyos elementos de entrada son los resultados obtenidos, generados en el paso previo, y la planilla de criterios de éxito del problema de negocio (comprensión del problema de negocio); y presentación de los resultados, cuyo elemento de entrada son los resultados relevantes seleccionados.

Las actividades desplegadas en la figura 16, presentan una capa de abstracción la cual agrupa distintas tareas, en base a su objetivo) y sus vinculaciones.

#### D.1. Entendimiento Del Dominio

La fase gestión de la entrega está compuesta por 3 actividades: requisitos, análisis del dominio y comprensión del problema de negocio.

*Requisitos*: cuyo objetivo es analizar en detalle las características y necesidades generales del proyecto, dejando dichos contenidos *propiamente* documentados. Dicha actividad está compuesta por 2 tareas: Educación de Requisitos y Desarrollo del Documento de Requisitos.

*Análisis del dominio*: tiene como meta identificar y profundizar los conocimientos de la organización con respecto a los objetivos del negocio y la comprensión de los términos específicos del mismo. Dicha actividad está integrada por 2 tareas: Descripción de la terminología e Identificación de objetivos.

*Comprensión del problema de negocio*: cuya finalidad es comprender y detallar la problemática a resolver asociadas con el proyecto, identificando los riesgos posibles y los criterios bajo los cuales se considerará satisfactoriamente resuelto el problema. Dicha actividad está conformada por 4 tareas: Identificación de los problemas de negocio, identificación de los expertos, identificación de riesgos del problema e identificación de criterios de éxito del problema de negocio.

#### D.2. Entendimiento de los datos

La fase entendimiento de los datos está compuesta por 3 actividades: análisis de los datos, exploración de los datos y evaluación de los datos.

*Análisis de los Datos*: cuyo objetivo es identificar y describir las fuentes de datos existentes en la organización de interés para el proyecto, y para los problemas de negocio, y la tipología de los datos que la conforman. Dicha actividad está integrada por 4 tareas: descripción de las vinculaciones entre las fuentes de datos, descripción de las fuentes de datos descripción de los campos e identificar los campos asociados al problema de negocio.

*Exploración de los Datos*: tiene como objetivo migrar los datos previamente identificados a una fuente de datos digital (estructural y tecnológicamente compatibles), y realizar un análisis detallado de los datos (identificar rango de valores existentes, media, mediana y promedio, etc.). Dicha actividad está compuesta por 2 tareas: integrar los datos a un medio digital y explorar los datos.

#### D.3. Modelado

La fase modelado está compuesta por 2 actividades: modelado del problema y configuración del modelo.

*Modelado del Problema*: cuya meta es definir el problema de negocio mediante vocabulario técnico, y determinar las técnicas de explotación de información a implementar para dar solución a los problemas identificados. Dicha actividad está integrada por 2 tareas: definir el problema de explotación de información y modelar el problema de explotación de información.

*Configuración del Modelo*: tiene como objetivo establecer las herramientas y algoritmos de explotación de información a aplicar en base al problema de explotación de información y las técnicas previamente identificadas. Dicha actividad está compuesta por 4 tareas: identificación de las herramientas alternativas, selección de la herramienta identificación de los algoritmos de minería de datos y selección de los algoritmos de minería de datos.

#### D.4. Preparación de los datos

La fase preparación de los datos está compuesta por 2 actividades: construcción de la fuente temporaria de datos y adecuación de la fuente temporaria de datos.

*Construcción de la fuente temporaria de datos:* cuyo objetivo es establecer la fuente de datos sobre la cual se implementarán las herramientas de explotación de información. Dicha actividad se compone de 3 tareas: seleccionar los datos, definir la fuente temporaria de datos y generar la fuente temporaria de datos.

*Adecuación de la fuente temporaria de datos:* su finalidad es ajustar la fuente de datos con el objetivo de adaptar los mismos a los requerimientos de las herramientas y algoritmos, así como favorecer la calidad de los resultados. Dicha actividad está integrada por 2 tareas: limpiar los datos y formatear los datos.

#### D.5. Implementación

La fase implementación está compuesta por 2 actividades: configuración de la implementación e implementación del modelo.

*Configuración de la implementación:* cuyo objetivo es determinar los conjuntos de datos a utilizar y ajustar los algoritmos para obtener el mejor resultado. Dicha actividad está conformada por 3 tareas: seleccionar el set de entrenamiento, configurar los algoritmos de explotación de información y entrenar los algoritmos de explotación de información.

*Implementación del modelo:* en la cual se ejecutan los algoritmos de explotación de información sobre el conjunto de datos seleccionado. Dicha actividad está integrada por una tarea: implementación de algoritmos de explotación de información.

#### D.6. Evaluación y Presentación

La fase evaluación y presentación está compuesta por 2 actividades: evaluación de los resultados y presentación de los resultados.

*Evaluación de los resultados:* cuyo objetivo es analizar los resultados obtenidos en la implementación del modelo y corroborar los mismos con los criterios de éxito y el cliente. Dicha actividad está conformada por 3 tareas: Seleccionar los resultados relevantes, verificar los resultados y validar los resultados.

*Presentación de los resultados:* su objetivo es generar y presentar al cliente el reporte final con los resultados obtenidos, garantizando la adecuada apropiación de los mismos. Dicha actividad está integrada por 2 tareas: generar el reporte final y presentar el reporte final.

La tabla II describe en detalle cada una de las tareas que integran al subproceso desarrollo, implementando la misma estructura que la utilizada para la tabla I.

## VI. CONCLUSIONES

Este trabajo parte de la premisa que los modelos de procesos existentes para proyectos de explotación de información, no consideran de forma completa las actividades requeridas para el desarrollo de este tipo de proyectos. La alta tasa de fracaso de los proyectos en el área, reflejan las carencias existentes en las distintas metodologías actuales, cuya falta de detalle sobre las actividades a realizar es una de las principales problemáticas.

El modelo de proceso más utilizado (CRISP-DM), así como su modelo base (KDD), se encuentran completamente enfocados en el proceso de desarrollo, asociado con las

actividades técnicas (asociadas con la evaluación y preparación de los datos, y la minería de datos). La falta de control de recursos, de planificación de las tareas, junto con la posibilidad de adaptación del proceso a las necesidades del proyecto (mediante la estructuración del progreso de las etapas y/o las actividades a realizar) incrementa la posibilidad de fracaso del mismo.

A partir de la identificación de dichas dificultades, Marbán propone un bosquejo de modelo de proceso para proyectos de explotación de información (o minería de datos en palabras del autor), el cual acopla el estándar más utilizado para proyectos de explotación de información (CRISP-DM), con los modelos de procesos para proyectos software convencionales (IEEE Std 1074 e ISO 12207). Dicha aproximación, no es completa y presenta carencias en la adaptación y definición detallada de las actividades. Adicionalmente el modelo de proceso incluye actividades pertinentes para la construcción de proyectos software tradicionales, las cuales no aplican para proyectos de explotación de información.

A partir de las dificultades identificadas durante los proyectos realizados, surge la necesidad de generar un modelo de proceso el cual solventa las problemáticas de los modelos previos, y brinde una guía detallada del progreso de las actividades, sus vinculaciones, así como los elementos asociados a las mismas.

En este contexto, se introduce un modelo de procesos para proyecto de explotación de información enfocado en la gestión, el cual guía al experto en el área respecto del conjunto de actividades, tareas y técnicas asociadas con este tipo de proyectos. La propuesta está conformada por 2 subprocesos:

El subproceso de *gestión*, definido a partir de las necesidades de este tipo de proyectos, donde se establecen todas las actividades asociadas a la interacción inicial con el cliente, la estructuración del proyecto, planificación y administración de los recursos, control y calidad del desarrollo del proyecto.

El subproceso de *desarrollo*, siendo este sobre el cual se enfocaron más los modelos de procesos previos, el cual toma sus bases del estándar de facto en la industria, y se reestructura para lograr un mejor acoplamiento entre las fases que lo componen. Además se incorporaron nuevas tareas, las cuales brindan un mayor nivel de sistematización, calidad y éxito del proyecto.

Entre los aportes que se realizan, se destacan:

- La incorporación del concepto de modelo de proceso, como instrumento para modelar la solución a implementar en base a los problemas de negocio y sus respectivos problemas de explotación de información, además de brindar una mayor sistematización del proyecto.
- La reestructuración de las fases de modelado y preparación de los datos, brindando un mejor acoplamiento entre las mismas y reduciendo el tiempo y costo asociados a iteraciones innecesarias a causa de una planificación deficiente o inexistente.
- La incorporación de un proceso destinado a dar soporte del desarrollo del proyecto, el cual se enfoca en la planificación, adaptación, gestión, control y reutilización del proyecto y sus productos asociados.
- Enfatiza (al igual que KDD), que el producto final del proyecto, es la generación de piezas de conocimiento interesantes que sirvan para dar soporte a la toma de decisiones.



TABLA II. LISTADO DE TAREAS PERTENECIENTES AL SUBPROCESO DE DESARROLLO

Fase	Actividad	Tarea	Objetivo	INPUT	OUTPUT	
Entendimiento del dominio	Requisitos	Educción de requisitos	Interactuar con el cliente/experto para profundizar los conocimientos vinculados al proyecto.	Discurso del experto Reporte de Exploración Inicial	Requisitos	
		Desarrollo del documento de requisitos	Documentar los conocimientos, necesidades y requerimientos del proyecto	Reporte de Exploración Inicial Requisitos	Documento de Requisitos	
	Análisis del dominio	Descripción de la terminología	Detallar los conceptos relevantes asociados al proyecto	Documento de Requisitos	Glosario de términos	
		Identificación de objetivos	Determinar los objetivos generales del proyecto	Documento de Requisitos	Objetivos del negocio	
	Comprensión del problema de negocio	Identificación de los problemas de negocio	Determinar los problemas de negocio, a partir de las características del proyecto	Documento de Requisitos Objetivos del negocio	Problemas de Negocio	
		Identificación de los expertos en el problema	Identificar a los expertos involucrados en el problema de negocio con quienes interactuar para validar los resultados	Problemas de Negocio	Listado de expertos en el Problema de Negocio	
		Identificación de riesgos del problema	Determinar los posibles riesgos asociados a la resolución de los problemas de negocio previamente identificados	Documento de Requisitos Problemas de Negocio	Listado de expertos en el Problema de Negocio	Riesgos asociados al problema de negocio
				Listado de expertos en el Problema de Negocio		
		Identificación de criterios de éxito del Problema de Negocio	Determinar bajo qué criterios se considera exitoso los resultados obtenidos para cada problema de negocio	Documento de Requisitos Problemas de Negocio	Listado de expertos en el Problema de Negocio	Planilla de Criterios de éxito del Problema de Negocio
				Listado de expertos en el Problema de Negocio		
	Entendimiento de los datos	Análisis de los datos	descripción de las vinculaciones entre las fuentes de datos	Identificar las asociaciones entre los datos contenidos por las distintas fuentes de datos, en caso que existiesen	Fuentes de Datos Reporte de Recursos externos	Diagrama de relación de fuentes de datos
			Descripción de las fuentes de datos	Detallar las características que presentan las fuentes de datos, desde su tecnología, versión, tipo de soporte, etc.	Fuentes de Datos Reporte de Recursos externos Diagrama de relación de fuentes de datos	Descripción de las fuentes de datos
Descripción de los campos			Detallar los campos que componen a las fuentes de datos existentes, identificando tipología, asociaciones, valores posibles, etc.	Fuentes de Datos Reporte de Recursos externos Diagrama de relación de fuentes de datos Descripción de las fuentes de datos	Descripción de los campos	
Identificar los Campos asociados al problema de negocio			Determinar el conjunto de datos relevantes asociados a los problemas de negocio	Problemas de Negocio Diagrama de relación de fuentes de datos Descripción de las fuentes de datos Descripción de los campos	Campos Relacionados con el Problema de Negocio	
Exploración de los datos		Integrar los datos a un medio digital	Convertir los datos de interés para el proyecto, a un mismo medio digital que permita implementar las técnicas de explotación de información	Fuentes de Datos No Digitales	Fuentes de Datos Digitales Reporte de Migración a Fuentes de Datos Digitales	
		Explorar los datos	Describir de forma detallada las características de los campos de interés para el proyecto.	Diagrama de relación de fuentes de datos Descripción de las fuentes de datos Descripción de los campos Fuentes de Datos Digitales Reporte de Migración a Fuentes de Datos Digitales Campos Relacionados con el Problema de Negocio	Reporte de Datos Explorados	
Evaluación de los datos		Verificación de la calidad de los datos	Evaluar la calidad de los campos de interés para el proyecto, identificando valores nulos, valores fuera de rango, etc.	Descripción de los campos Reporte de Datos Explorados	Reporte de la calidad de los datos	
				Descripción de los campos Reporte de Datos Explorados		
		Identificación de campos riesgosos	Determinar los campos analizados que presentan dificultades las cuales pueden perjudicar su utilización para el proyecto	Descripción de los campos Reporte de Datos Explorados Reporte de la calidad de los datos	Reporte de Campos Riesgosos	

Modelado	Modelado del problema	definir el problema de explotación de información	Determinar mediante vocabulario técnico los problemas a resolver de acuerdo a los problemas de negocio identificados previamente	Problemas de Negocio Campos Relacionados con el Problema de Negocio Descripción de los campos Reporte de Datos Explorados Reporte de la calidad de los datos Reporte de Campos Riesgosos	Problemas de explotación de información	
		Modelar el problema de explotación de información	Representar mediante formalismos de conceptualización los conceptos asociados al problema de explotación de información, de forma que sea posible identificar las técnicas o algoritmos de explotación de información a utilizar	Campos Relacionados con el Problema de Negocio Reporte de Campos Riesgosos Problemas de explotación de información	Identificación de la solución	
		Identificación de las herramientas alternativas	Determinar el conjunto de herramientas existentes en el mercado que satisfacen las demandas del proyecto, analizando sus ventajas y desventajas	Campos Relacionados con el Problema de Negocio Descripción de los campos Reporte de Datos Explorados Reporte de Campos Riesgosos Problemas de explotación de información Identificación de la solución	Herramientas alternativas	
				Herramientas alternativas	Herramienta seleccionada	
	Configuración del modelo	Selección de la herramienta	Formalizar la selección de la/s herramientas elegidas para el desarrollo del proyecto	Herramientas alternativas	Herramienta seleccionada	
		identificación de los algoritmos de minería de datos	Determinar los posibles algoritmos de explotación de información a utilizar	Campos Relacionados con el Problema de Negocio Descripción de los campos Problemas de explotación de información Identificación de la solución Herramienta seleccionada	Algoritmos de minería de datos soportados	
				Campos Relacionados con el Problema de Negocio Descripción de los campos Problemas de explotación de información Identificación de la solución Herramienta seleccionada Algoritmos de minería de datos soportados	Algoritmos de minería de datos seleccionados	
		selección de los algoritmos de minería de datos	Determinar los algoritmos óptimos en base a las características del dominio y del problema a resolver	Campos Relacionados con el Problema de Negocio Descripción de los campos Problemas de explotación de información Identificación de la solución Herramienta seleccionada Algoritmos de minería de datos soportados	Algoritmos de minería de datos seleccionados	
				Campos Relacionados con el Problema de Negocio Descripción de los campos Problemas de explotación de información Identificación de la solución Herramienta seleccionada Algoritmos de minería de datos soportados	Datos Seleccionados	
	Preparación de los datos	Construcción de la fuente temporaria de datos	seleccionar los datos	Determinar a partir del análisis del estado de los datos asociados al problema de negocio, cuáles serán utilizados para la implementación de las técnicas escogidas.	Campos Relacionados con el Problema de Negocio Reporte de Campos Riesgosos	Datos Seleccionados
			Definir la fuente temporaria de datos	Determinar las características necesarias de la fuente de datos para la ejecución de los algoritmos de explotación de información	Campos Relacionados con el Problema de Negocio Reporte de Campos Riesgosos Datos Seleccionados Algoritmos de minería de datos seleccionados	Descripción de la Fuente Temporaria de Datos Reporte de Necesidad de Transformación de los Datos
			generar la fuente temporaria de datos	Integrar todos los campos seleccionados a una única fuente de datos, dejando constancia de las adaptaciones a realizar en la fuente de datos integrada	Descripción de los campos Descripción de la Fuente Temporaria de Datos Reporte de Necesidad de Transformación de los Datos	Fuente Temporaria de Datos Reporte de Generación de la Fuente Temporaria de datos Reporte de Transformación de los Datos Reporte de Limpieza de Datos
		Adecuación de la fuente temporaria de datos	Limpiar los datos	Adaptar o Eliminar la fuente temporaria de datos, para que esta posea únicamente valores dentro de los rangos posibles de cada campo	Descripción de los campos Reporte de la calidad de los datos Reporte de Campos Riesgosos Fuente Temporaria de Datos	Fuente Temporaria de Datos Reporte de Limpieza de Datos
					Descripción de los campos Reporte de Necesidad de Transformación de los Datos Fuente Temporaria de Datos	Fuente Temporaria de Datos Reporte de Transformación de los Datos
			Formatear los datos	Adaptar los valores de los campos al formato requerido	Descripción de los campos Reporte de Necesidad de Transformación de los Datos Fuente Temporaria de Datos	Fuente Temporaria de Datos Reporte de Transformación de los Datos
Implementación		Configuración de la implementación	seleccionar el set de entrenamiento	Determinar el conjunto de datos a utilizar para el entrenamiento y la ejecución de los algoritmos	Problemas de explotación de información Fuente Temporaria de Datos	Set de entrenamiento
			configurar algoritmos de Minería de datos	Determinar los parámetros de los algoritmos de explotación de información para generar el mejor resultado posible	Problemas de explotación de información Identificación de la solución Herramienta seleccionada	Set de implementación Reporte de Configuración de

		Entrenar algoritmos de Minería de datos	Ejecutar los algoritmos de explotación de información con la configuración definida en el conjunto de datos de entrenamiento con el objetivo de generar un modelo	Algoritmos de minería de datos seleccionados	Algoritmos Entrenados			
				Identificación de la solución				
				Herramienta seleccionada				
				Algoritmos de minería de datos seleccionados				
				Set de Entrenamiento				
	Implementación del modelo	Implementación de algoritmos de Minería de datos	Implementar los algoritmos de explotación de información entrados sobre el set de datos de implementación, obteniendo patrones existentes en los datos	Reporte de Configuración de Algoritmos	Reporte de Entrenamiento de Algoritmos			
				Identificación de la solución				
				Herramienta seleccionada				
				Algoritmos de minería de datos seleccionados				
				Fuente Temporal de Datos				
Evaluación y Presentación	Evaluación de los resultados	Seleccionar los resultados relevantes	Analizar y seleccionar aquellos resultados interesantes obtenidos en la tarea previa	Resultados Obtenidos	Resultados Relevantes seleccionados			
				Verificar los resultados		Contrastar los resultados seleccionados con los criterios de éxito garantizando que cumplan con los requerimientos del cliente respecto del problema de negocio	Resultados Relevantes seleccionados	Verificación de los Resultados
							Planilla de Criterios de éxito del Problema de Negocio	
	Validar los resultados	Contrastar los resultados con los expertos del negocio, identificando la relevancia de los mismos	Resultados Relevantes seleccionados	Validación de los Resultados				
	Presentación de los resultados	Generar el reporte final	Generar documentación de los resultados obtenidos para presentar al cliente	Presentar los resultados obtenidos al cliente	Resultados Relevantes seleccionados	Reporte del Proyecto		
					Validación de los Resultados			
		Presentar el reporte final	Reporte del Proyecto	Presentación del Proyecto				

#### FINANCIAMIENTO

Las investigaciones que se reportan en este trabajo han sido financiadas parcialmente por el Proyecto de Investigación 33A167 de la Secretaría de Ciencia y Técnica de la Universidad Nacional de Lanús (Argentina).

#### REFERENCIAS

- [1] Thomsen, E. 2003. *BI's Promised Land*. Intelligent Enterprise, 64: 21-25.
- [2] Negash, S., Gray, P. 2008. *Business Intelligence*. En *Handbook on Decision Support Systems 2*, ed. F. Burstein y C. Holsapple Heidelberg, Springer, Pág. 175-193.
- [3] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. 1996. *From data mining to knowledge discovery in databases*. AI Magazine, 17(3): 37-54.
- [4] Tan, A. 1999. *Text mining: The state of the art and the challenges*. In *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*. pp. 65-70.
- [5] Hsu, W., Lee, M., Zhang, J. 2002. *Image mining: Trends and developments*. Journal of Intelligent Information Systems, 19(1): 7-23.
- [6] Gaber, M., Zaslavsky, A. Krishnaswamy, S. 2010. *Data stream mining*. En Maimon, O. and Rokach, L., eds. *Data mining and knowledge discovery handbook*. Springer, Pág. 759-787. ISBN 978-0-387-09823-4.
- [7] Kosala, R., Blockeel, H. 2000. *Web mining research: A survey*. ACM SIGKDD Explorations Newsletter, 2(1): 1-15.
- [8] Langseth, J., Vivatrat, N. 2003. *Why Proactive Business Intelligence is a Hallmark of the Real-Time Enterprise: Outward Bound*. Intelligent Enterprise 518: 34-41.
- [9] Srivastava, J., Cooley, R., Deshpande, M., Tan, P. 2000. *Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data*. SIGKDD Explorations, 12: 12-23.
- [10] Abran, A., Moore, J. W., Bourque, P., Dupuis, R., Tripp, L. 2004. *Guide to the Software Engineering Body of Knowledge (2004 version)*. IEEE Computer Society Press. ISBN 0-7695-2330-7.
- [11] García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P., Vanrell, J. 2011. *Towards an Information Mining Engineering*. En *Software Engineering, Methods, Modeling and Teaching*. Sello Editorial Universidad de Medellín. ISBN 978-958-8692-32-6. Páginas 83-99.
- [12] Ochoa, A. Fernández, E., Britos, P., García-Martínez, R. 2008. *Metodologías de Ingeniería Informática*. Editorial Nueva Librería. ISBN 978-987-1104-54-3.
- [13] Boehm, 1981. B. W. BOEHM, *Software Engineering Economics*, Prentice Hall, 1981.
- [14] Pressman, R., 2002. *Ingeniería del Software. Un Enfoque Práctico*. Quinta Edición. ISBN: 84-481-3214-9.
- [15] Berry, M. J.A, Linoff, G. S., 2004. *Data Mining Techniques. For Marketing, Sales, and Customer Relationship Management*. Segunda Edición. ISBN 0-471-47064-3
- [16] 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. Pág. 674-682. ISBN 978-950-9474-49-9.
- [17] 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*. Pág. 46-52. ISBN 978-958-46-0577-1.

- [18] Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. 2000. CRISP-DM 1.0 Step by step BI guide. Edited by SPSS.
- [19] SAS, 2008. SAS Enterprise Miner: SEMMA. <http://www.sas.com/technologies/analytics/datamining/miner/semma.html>. Ultimo acceso 23/03/2014.
- [20] Pyle, D. 2003. Business Modeling and Business intelligence. Morgan Kaufmann Publishers.
- [21] Marbán, O., Mariscal, G., Menasalvas, E., Segovia, J. 2007. An Engineering Approach to Data Mining Projects. Lecture Notes in Computer Science, 4881: 578-588. Springer.



**Sebastian Martins.** Candidato del Programa de Doctorado en Ciencias Informáticas por la Universidad Nacional de La Plata (UNLP). Su investigación se centra en “Integración de Métodos en Ingeniería de Explotación de Información”. Es Licenciado en Sistemas por la Universidad Nacional de Lanús (UNLa). Es Investigador Asistente del Laboratorio de Investigación y

Desarrollo en Ingeniería de Explotación de Información del Grupo de Investigación en Sistemas de Información (UNLa GISI) y Docente de la Asignaturas Tecnologías de Explotación de Información, Ingeniería de Software I y III, Proyecto Software y Fundamentos de la Teoría de la Computación de la Licenciatura en Sistemas del Departamento de Desarrollo Productivo y Tecnológico de la Universidad Nacional de Lanús.



**Patricia Pesado.** Es Calculista Científica y Licenciada en Informática por la Universidad Nacional de La Plata. Es Profesora Titular Ordinaria del área de Ingeniería de Software, y Directora Adjunta del Instituto de Investigación en Informática (III-LIDI) de la Facultad de Informática de la Universidad Nacional de La Plata

(UNLP). Se desempeña en la Facultad de Informática de la Universidad Nacional de La Plata como Directora de la carrera Licenciatura en Sistemas y de la carrera Ingeniería en Computación. Es Docente Investigador Categoría II en el Programa de Incentivos del Ministerio de Educación.



**Ramón García Martínez.** Es Analista de Computación por la UNLP, es Licenciado en Sistemas de Información por la UNLP, es Master en Ingeniería Informática y Doctor en Informática por la Universidad Politécnica de Madrid. Es Profesor Titular Ordinario del área de Ingeniería de Software en la carrera de la Licenciatura en

Sistemas y Director del Grupo de Investigación en Sistemas de Información (GISI) de la Universidad Nacional de Lanús (UNLa). Se desempeña en la Facultad de Informática de la Universidad Nacional de La Plata como Profesor de Cursos de Doctorado en el Programa de Doctorado en Ciencias Informáticas. Es Docente Investigador Categoría I en el Programa de Incentivos del Ministerio de Educación.