

Modelo de Procesos para la Gestión de Requerimientos en Proyectos de Explotación de Información

Pollo-Cattaneo, M. F.^{1,2}, Mansilla, D.², Vegega, C.², Pesado, P.³, García-Martínez, R.⁴, P. Britos, P.⁵

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

² Grupo de Estudio en Metodologías de Ingeniería de Software. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional. Argentina.

³ Instituto de Investigaciones en Informática LIDI. Facultad de Informática. Universidad Nacional de La Plata. Argentina.

⁴ Grupo Investigación en Sistemas de Información. Departamento Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús. Argentina

⁵ Grupo de Investigación en Explotación de Información. Laboratorio de Informática Aplicada. Universidad Nacional de Río Negro. Argentina.

fpollo@posgrado.frba.utn.edu.ar, rgarcia@unla.edu.ar, paobritos@gmail.com

Resumen. Los Proyectos de Explotación de Información tienen por objetivo la transformación de los datos que recopila el proceso de negocio, en conocimiento útil para la toma de las decisiones. Los procesos asociados a los proyectos tradicionales de construcción de software no pueden utilizarse, debido a que estos procesos están orientados a la construcción de un producto diferente, que es el software. En este contexto, se presenta un proceso de gestión de requerimientos para Proyectos de Explotación de Información, que hace hincapié en las fases de definición de proyectos, educación del negocio, conceptualización del negocio e identificación de los procesos de explotación de información asociados al Proyecto de Explotación de Información.

Palabras Claves: Elicitación de Requerimientos. Metodología. Proceso. Explotación de Información. Ingeniería de Requerimientos.

1. Introducción

La Inteligencia de Negocios propone un abordaje interdisciplinario (dentro del que se encuentra la Informática) que se centra en generar conocimiento que contribuya con la toma de decisiones de gestión y generación de planes estratégicos en las organizaciones [1]. La Explotación de Información (EdI) es la sub-disciplina de la Informática que aporta a la Inteligencia de Negocios las herramientas de análisis y síntesis para extraer conocimiento no trivial que se encuentra (implícitamente) en los datos disponibles de diferentes fuentes de información [2]. Para un experto, o para el responsable de un Sistema de Información, normalmente no son los datos en sí lo más relevante, sino el conocimiento que se encierra en sus relaciones, fluctuaciones y dependencias.

Si bien existen metodologías que acompañan el desarrollo de Proyectos de Explotación de Información que se consideran probadas y tienen un buen nivel de madurez en cuanto al desarrollo del proyecto entre las cuales se destacan CRISP-DM [4], P3TQ [5] y SEMMA [6], estas metodologías dejan de lado aspectos operativos y de gestión de proyecto [7]. Así, por ejemplo, en la metodología CRISP-DM la primera fase busca identificar y comprender los aspectos del negocio relacionados al proyecto que se está realizando, pero no define técnicas, métodos ni herramientas para obtener esta información ni los medios necesarios para realizar su documentación.

En este contexto, este trabajo tiene como objetivo sistematizar el cuerpo de conocimientos existente en la Ingeniería en Software y la Ingeniería del Conocimiento para sentar las bases para el desarrollo de una Ingeniería de Requisitos con particular énfasis en Proyectos de Explotación de Información. De esta manera se propone un Modelo de Procesos para la Gestión de Requisitos en Proyectos de EdI. Para ello primero se describe el problema detectado (sección 2). Luego se definen las soluciones existentes junto con la nueva propuesta (sección 3), aplicando esta última en una prueba de concepto para su validación (sección 4). Finalmente se indican las conclusiones obtenidas y futuras líneas de trabajo (sección 5).

2. Definición del Problema

Aunque la Ingeniería en Software y la Ingeniería del Conocimiento [8] proveen muchos métodos, técnicas y herramientas, estos no son útiles ya que no se ocupan de los aspectos específicos que poseen los Proyectos de Explotación de Información. Mientras, que las herramientas tradicionales de elicitación de la Ingeniería en Software se enfocan en la descripción de los diferentes tipos de requerimientos haciendo hincapié en las características (funcionales o no) que debe cumplir el producto software final [9], un proyecto EdI no busca la construcción del sistema software sino la aplicación de un proceso que convierta los datos disponibles en conocimiento. Por lo tanto, los requerimientos en este tipo de proyecto, se encontrarán relacionados a identificar y describir los objetivos del proyecto junto con su relación con los objetivos del negocio de la organización donde se está realizando el proyecto. Además es necesario realizar un reconocimiento inicial de las fuentes de información disponibles en la organización identificando cuáles fuentes se encuentran informatizadas (en repositorios de datos) y cuáles no. Como resultado del análisis conjunto de los objetivos del proyecto y los repositorios de datos identificados, es posible delimitar el alcance del proyecto en un conjunto de objetivos de requisitos que podrán ser luego resueltos a través de la aplicación de procesos de EdI [10].

Sin embargo, y como sucede en otras ingenierías, al comienzo de un proyecto de EdI se tiene la dificultad adicional de no manejar el vocabulario del negocio y de los datos de los miembros de la organización para poder lograr un mejor entendimiento sobre los aspectos del proyecto.

En este sentido, ante la carencia de métodos, técnicas y herramientas asociadas a la ejecución de las tareas relacionadas a la gestión de requerimientos en proyectos de EdI, se ha detectado la necesidad de ofrecer los siguientes elementos:

- un proceso de elicitación de requerimientos para identificar las principales necesidades del cliente, sus expectativas, restricciones y los principales repositorios de datos que son necesarios para realizar el proyecto.
- un conjunto de plantillas que permita documentar todos los requerimientos para que puedan ser, luego, consultados durante la realización del proyecto.
- un proceso de formalización de requerimientos que indique la forma en que se deben completar las plantillas en base a los requerimientos educidos.

3. Solución Propuesta

Para dar respuesta al problema detectado en la sección anterior, se han propuesto varias soluciones (descriptas en la sección 3.1) que buscan dar solución a aspectos puntuales del problema pero que presentan problemas al intentar aplicarlas en un todo. A partir del análisis de estas soluciones parciales existentes se propone su integración en un Modelo de Procesos para la Gestión completa de los Requerimientos en Proyectos de EdI. La propuesta de este modelo se realiza en la sección 3.2.

3.1. Soluciones Existentes

En [11] se realiza una propuesta para la solución del problema definiendo tanto un conjunto de plantillas asociadas al proyecto, los requisitos, los datos y la terminología de la organización así como un proceso general que sirva de guía para la obtención de la información necesaria que debe ser documentada en dichas plantillas. Sin embargo, este proceso se encuentra demasiado vinculado a la documentación de los requerimientos, dejando sin definir las actividades necesarias para educir y entender los requerimientos. De todas formas, ha servido como punto de partida para la definición de las propuestas realizadas a continuación:

- 1) A partir de la revisión de las plantillas propuestas se ha realizado un ajuste en las mismas con el objetivo de tener un mayor entendimiento y se ha definido un conjunto de normas que indican cómo se deben escribir los requerimientos [12].
- 2) Se ha definido un proceso de formalización para transformar los requerimientos educidos de forma de poder ser documentados en las plantillas revisadas (a través de la utilización de técnicas de representación de conocimiento provistas por la Ingeniería del Conocimiento [13]).
- 3) Adaptando el ciclo de vida propuesto por [14] para iniciativas de Data Warehouse & Business Intelligence, se ha propuesto un modelo de procesos para elicitación de requerimientos en proyectos de EdI [15, 16] el cual define un conjunto de fases con las tareas, técnicas y métodos que se deben aplicar, pero que no considera la utilización del proceso de formalización para completar las plantillas correspondientes.

3.2. Modelo de Procesos Propuesto

La solución propuesta consiste en un modelo de procesos que permite relacionar las diferentes propuestas asociadas a la elicitación de requerimientos en proyectos de EdI. El modelo consiste en definir una relación transversal entre las propuestas realizadas en [12], [13], [15] y [16].

El proceso se divide en cuatro fases principales: Definición del Proyecto, Educción del Negocio, Conceptualización del Negocio e Identificación de Procesos de Explotación de Información. Cada fase tiene definido un conjunto de actividades y un conjunto de procesos de formalización asociados a dichas actividades.

La figura 1 muestra el mapa conceptual completo del Modelo de Procesos propuesto, dividido en los diferentes niveles (Fase, Actividad y Proceso de Formalización).

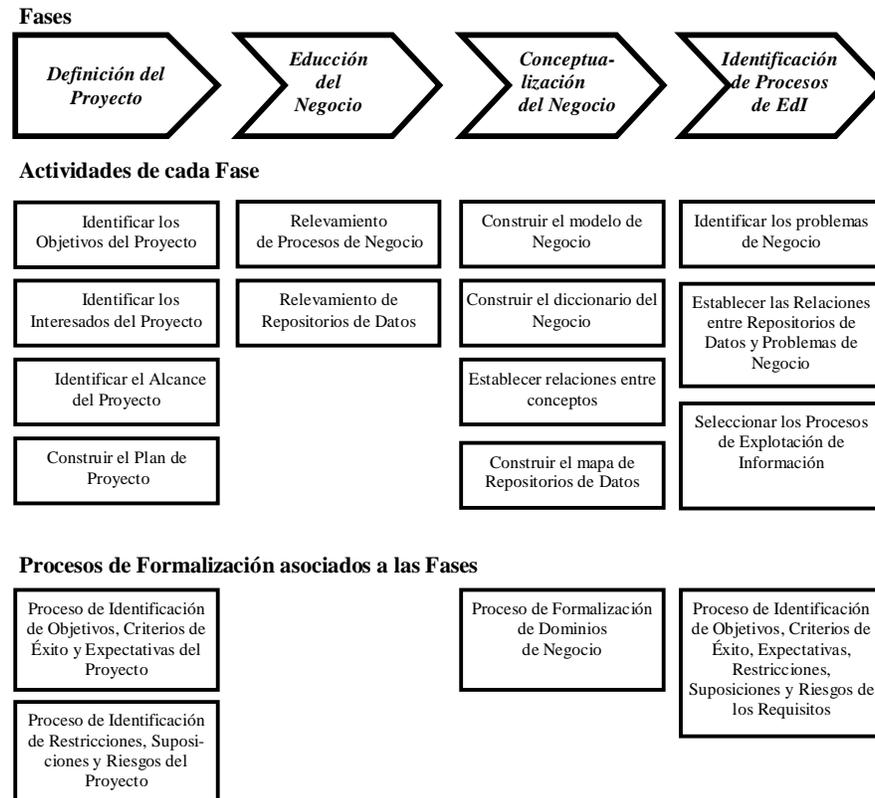


Fig.1 Mapa Conceptual del Modelo de Procesos Propuesto

La fase de Definición del Proyecto tiene por objetivo definir el alcance del proyecto, los interesados y los objetivos que se deben alcanzar. En esta fase se realiza la planificación de actividades de educación de requisitos.

La fase de Educación del Negocio tiene por objetivo comprender el idioma utilizado en el negocio, descubrir las palabras específicas del mismo y cuál es el significado que el negocio le da a esas palabras específicas.

La fase de Conceptualización del Negocio tiene por objetivo definir el negocio en términos de conceptos utilizados en el mismo, vocabulario y repositorios donde se almacena la información de los diferentes procesos del negocio.

La fase de Identificación de Procesos de Explotación de Información define los procesos de minería de datos que se pueden utilizar para resolver los problemas identificados en el proceso de negocio.

En esta instancia, se identifican los roles de las personas que participan en este modelo de procesos. Dichos roles, junto con sus responsabilidades son indicados en la tabla 1.

Rol	Responsabilidades
<i>Líder de Proyecto</i>	Gestionar las acciones para que se lleven a cabo las actividades del proyecto y se cumplan los compromisos del proyecto.
<i>Analista Funcional</i>	Relevar y analizar los diferentes procesos del negocio.
<i>Especialista de Datos</i>	Relevar y analizar las diferentes fuentes de información. Debe tener las capacidades técnicas necesarias para recuperar datos de dichas fuentes.
<i>Analista en Explotación de Información</i>	Establecer relaciones entre datos y relacionar procesos de explotación de información con los problemas del negocio detectados. Aplicar algoritmos de minería de datos sobre los datos relacionados al proceso de negocio para obtener el conocimiento que permita resolver los problemas del negocio detectados.

Tabla 1. Roles del Proceso de Gestión de Requerimientos

3.2.1 Fase de Definición del Proyecto

Durante esta fase se realizan las tareas asociadas a la planificación del proyecto y a establecer el alcance y las personas interesadas en el mismo. Estas actividades son la base de todo proyecto [17]. La figura 2 muestra las actividades y procesos de formalización asociados a esta fase.

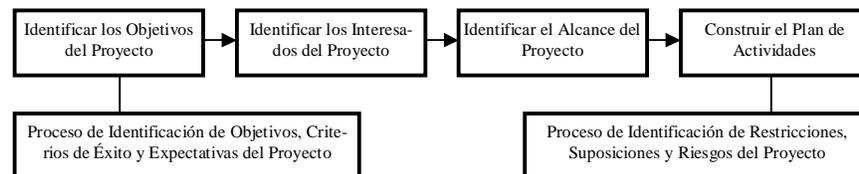


Fig.2 Actividades y Procesos de Formalización de la Fase de Definición del Proyecto

El Líder de Proyecto será el responsable de las diferentes actividades de esta fase. Comienza con la identificación de los objetivos del proyecto, utilizando el Proceso de Identificación de Objetivos, Criterios de Éxito y Expectativas del Proyecto. Con los

objetivos definidos, se identifica la lista de interesados del proyecto (actividad “Identificar los Interesados del Proyecto”) en la que figuran quienes permiten definir el alcance del proyecto de EdI, actividad en la que también participa el Analista Funcional. Por último, con el objetivo, alcance e interesados del proyecto definidos, el Lider de Proyecto utiliza el Proceso de Identificación de Restricciones, Suposiciones y Riesgos del Proyecto para construir el Plan de Actividades que servirá como planificación para la ejecución de las diferentes actividades del proceso (actividad de “Construir Plan de Actividades”). Con el plan construido comienzan las actividades asociadas a las siguientes fases.

3.2.2 Fase de Educación del Negocio

Durante esta fase se realizan las tareas que permiten relevar los diferentes procesos de negocio. El Analista Funcional será el responsable de llevar a cabo las actividades definidas para esta Fase. La figura 3 muestra las actividades que la componen.

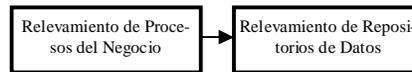


Fig.3 Actividades de la Fase de Educación del Negocio

El Analista Funcional, en la actividad de “Relevamiento de Procesos del Negocio”, recopilará la información de los diferentes procesos, mediante la utilización de técnicas tradicionales de educación, como las que se presentan en [9] y [18]. Esta información documentada será utilizada como referencia en la siguiente fase. Con la ayuda de un Especialista de Datos, deberá identificar los diferentes repositorios de información que existen en la organización (actividad de “Relevamiento de Repositorios de Datos”) y que pueden, o no, estar informatizados.

3.2.3 Fase de Conceptualización del Negocio

En la fase de conceptualización se construye el modelo de negocio que se utiliza como base del Proyecto de EdI. El Analista Funcional será el responsable de las actividades de esta fase que incluyen las actividades definidas en la figura 4.

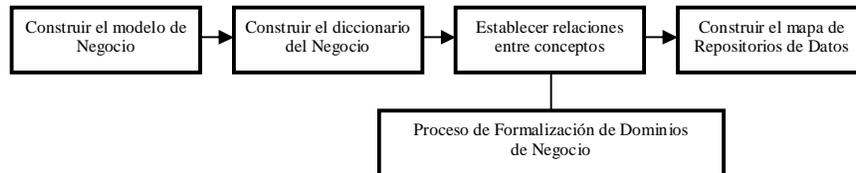


Fig.4 Actividades y Procesos de Formalización de la Fase de Conceptualización del Negocio

Con la información obtenida en la “Educción del Negocio”, durante la actividad de “Construir el modelo de Negocio”, el Analista Funcional modela los casos de uso de negocio asociados al alcance del proyecto. A partir de este modelo, como parte de la actividad “Construir el Diccionario del Negocio”, se analiza el vocabulario utilizado en el negocio y se construye el diccionario de términos asociados a los diferentes procesos de negocio. El Analista de Explotación de Información, en la actividad de “Establecer las Relaciones entre Conceptos”, sigue el Proceso de Formalización de Dominios de Negocio, establecido en [13] para generar el Diagrama de Entidad-Relación (DER) asociado a los conceptos. Por último, el rol de Especialista de Datos será el de ser el responsable de relevar las diferentes fuentes de información asociadas a los procesos modelados, las que permitirán establecer las relaciones entre los Casos de Uso identificados y los repositorios de datos existentes. Esta relación se establece en la actividad de “Construir el Mapa de Repositorio de Datos”.

3.2.4 Fase de Identificación de Procesos de Explotación de Información

Esta fase es la conclusión del trabajo realizado por el equipo de proyecto y define como resultado final los Procesos de Explotación de Información a utilizar en el proyecto. La figura 5 representa las actividades y los procesos de formalización asociados a esta fase.

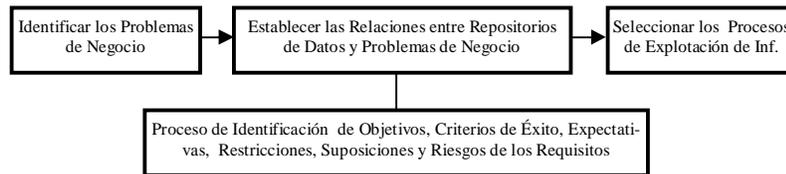


Fig.5 Actividades y Procesos de Formalización de la Fase de Identificación de Procesos de Explotación de Información

El Analista Funcional se encargará de construir una lista que contiene los problemas que el negocio desea resolver (actividad “Identificar los Problemas de Negocio”), y en conjunto con el Analista de Explotación de Información establecerán las relaciones existentes entre los problemas identificados y los repositorios que contienen información útil para la resolución de los problemas (actividad “Establecer las Relaciones entre Repositorios de Datos y Problemas de Negocio”) utilizando el Proceso de Proceso de Identificación de Objetivos, Criterios de Éxito, Expectativas, Restricciones, Suposiciones y Riesgos de los Requisitos. Con esta información, y utilizando técnicas propuestas en [15], [16] se podrá/n seleccionar el o los procesos de Explotación de Información que utilizará el proyecto (actividad “Seleccionar Procesos de Explotación de Información”).

4. Prueba de Concepto

La prueba de concepto se realizó aplicando el proceso propuesto, al caso de estudio “Detección de Patrones de Daños y Averías” [19].

Como primera fase del proceso se identifican los objetivos que el negocio desea cumplir, los interesados del proyecto y el alcance que se desea alcanzar. Estas tareas se planifican y formalizan en el plan de actividades.

El trabajo a realizar define la lista de objetivos de negocio, entre los que podemos destacar los siguientes:

- OR-01 - Determinar la responsabilidad de siniestralidad en función del tipo de avería y el tipo de transporte.
- OR-02 - Identificar incidentes según el tipo de transporte.
- OR-03 - Identificar tipos de averías y/o daños.
- OR-04- Identificar partes averiadas y/o dañadas que muestren algún tipo de comportamiento.
- OR-05 - Identificar la gravedad de los daños y/o averías.
- OR-06 - Identificar los lugares donde se producen daños y/o averías tratando de definir patrones de comportamiento.
- OR-07 - Determinar en forma estadística: tipos de transporte que producen daños y/o averías como así también, partes, tipos de averías, gravedades, lugares donde se producen.

Durante la fase de “Educción de Negocio” se releva la información asociada a los procesos de negocio y a los repositorios de información asociados. La información relevada puede presentarse en un documento como el presentado en [19] o utilizando cualquier otro mecanismo de documentación tradicional como los indicados en [18].

En la Fase de “Conceptualización del Negocio” se trabaja con los modelos de negocio y de vocabularios, obteniendo el mapa de repositorios y conceptos, utilizando el Proceso de Formalización de Dominios de Negocio para establecer las relaciones entre conceptos. En [13] se presenta el trabajo realizado utilizando este proceso.

Por último, se trabaja en la “Fase de Identificación de Procesos de Explotación de Información”. En este caso, podemos identificar los siguientes problemas de negocio:

- Dificultad en identificar daños y averías producidos en unidades automotrices cero kilómetro desde que parten de la fábrica hasta el final del circuito.
- Establecer recursos consumidos por el movimiento de las unidades.
- Elevado costo en la distribución de repuestos para reparar las averías.

En [13] se define cómo establecer las relaciones entre los conceptos y los repositorios de datos. Como última actividad, se deben seleccionar los Procesos de Explotación de Información utilizando la información obtenida durante el proceso, se analizan los diferentes objetivos de requisito identificados y se establecen qué Procesos de Explotación de Información son acordes a cada problema identificado.

De dicho análisis, se puede concluir con las siguientes relaciones entre Procesos de Explotación de Información y objetivos:

- Para OR1, se selecciona el proceso de descubrimiento de reglas de pertenencia a grupos.

- Para OR2, se selecciona el proceso de descubrimiento de reglas de comportamiento usando “Transporte” como atributo objetivo.
- Para OR3, se selecciona el proceso de descubrimiento de reglas de comportamiento usando “Avería” como atributo objetivo.
- Para OR4, se selecciona el proceso de descubrimiento de reglas de comportamiento usando “Parte” como atributo objetivo.
- Para OR5, se selecciona el proceso de descubrimiento de reglas de comportamiento usando “Gravedad” como atributo objetivo.
- Para OR6, se selecciona el proceso de descubrimiento de reglas de comportamiento usando “Lugar” como atributo objetivo.
- Para OR7, se selecciona el proceso de ponderación de reglas de pertenencia a grupos.

Con este análisis se dan por concluidas las actividades del proceso y se continúa con la ejecución del proyecto de EdI.

5. Conclusiones

Este trabajo presenta una propuesta de Modelo de Procesos para la Gestión de Requerimientos en Proyectos de EdI que permite relacionar diferentes metodologías y herramientas propuestas en otros trabajos y dar un enfoque global a la Gestión de Proyectos de Explotación de Información. El proceso se descompone en cuatro fases, donde se gestiona en forma completa el alcance y los interesados del proyecto (Fase de Definición del Proyecto), se obtiene la información generada en los diferentes procesos de negocio (Educción del Negocio), se modelan los procesos y datos del negocio (Conceptualización del Negocio).

Como futuras líneas de trabajo, se deberán definir o relacionar los procesos de formalización con la fase de Educción de Negocio, y se trabajará en la prueba del proceso en diferentes casos para obtener validaciones empíricas del mismo.

6. Referencias

1. Thomsen, E. (2003). *BI's Promised Land*. Intelligent Enterprise, 6(4): 21-25.
2. Schiefer, J., Jeng, J., Kapoor, S. & Chowdhary, P. (2004). *Process Information Factory: A Data Management Approach for Enhancing Business Process Intelligence*. Proceedings 2004 IEEE International Conference on E-Commerce Technology. Pág. 162-169.
3. Curtis, B., Kellner, M., Over, J. (1992). *Process Modelling*. Communications of the ACM, 35(9): 75-90.
4. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step Data Mining Guide*. <http://tinyurl.com/crispdm>. Último acceso Enero de 2013.
5. Pyle, D. (2003). *Business Modeling and Business intelligence*. Morgan Kaufmann Publishers.
6. SAS (2008). *SAS Enterprise Miner: SEMMA*. <http://tinyurl.com/semmaSAS>. Último acceso Enero de 2013.

7. 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.
8. García Martínez, R. & Britos, P. (2004). *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería.
9. Wiegers, K. (2003). *Software Requirements*. Microsoft Press.
10. García-Martínez, R., Britos, P., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P. (2011). *Information Mining Processes Based on Intelligent Systems*. Proceedings of II International Congress on Computer Science and Informatics (INFONOR-CHILE 2011). Pág. 87-94. ISBN 978-956-7701-03-2.
11. Pollo-Cattaneo, M. F., Britos, P., Pesado, P. & García-Martínez, R. (2010). *Proceso de Educación de Requisitos en Proyectos de Explotación de Información*. En Ingeniería de Software e Ingeniería del Conocimiento: Tendencias de Investigación e Innovación Tecnológica en Iberoamérica (Editores: R. Aguilar, J. Díaz, G. Gómez, ELéon). Pág. 01-11. Alfaomega Grupo Editor. ISBN 978-607-707-096-2.
12. Deroche, A. & Pollo-Cattaneo, M. F. (2011) *Guía de Buenas Prácticas para Completar las Plantillas de Requerimientos para Proyectos de Explotación de Información*. Reporte Técnico GEMIS-TD-2011-01-RT-2012-01. Grupo de Estudio de Metodologías para Ingeniería en Software, UTN-FRBA.
13. Vegega, C., Pytel, P., Ramón, H., Rodríguez, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. (2012). *Formalización de Dominios de Negocio para Proyectos de Explotación de Información basada en Técnicas de Ingeniería del Conocimiento*. Proceedings del XVIII Congreso Argentino de Ciencias de la Computación. Pag. 1049-1058. ISBN 978-987-1648-34-4.
14. Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., & Becker, B. (2011). *The data warehouse lifecycle toolkit*. Wiley & Sons.
15. Mansilla, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. (2012). *Modelo de Proceso para Elicitación de Requerimientos en Proyectos de Explotación de Información*. Proceedings Latin American Congress on Requirements Engineering and Software Testing. Pág. 38-45. ISBN 978-958-46-0577-1.
16. Mansilla, D., Pollo, F., Britos, P., García-Martínez, R. (2013). *A Proposal of a Process Model for Requirements Elicitation in Information Mining Projects*. Lecture Notes in Business Information Processing, 139: 165-173. ISBN 978-3-642-36610-9.
17. William, R. 1996. *A Guide To The Project Management Body Of Knowledge*. PMI Publishing.
18. Sommerville, Ian, Y Peter Sawyer. 1997. *Requirements Engineering: A Good Practice Guide*. Chichester, England: John Wiley & Sons
19. H. Flores (2009). *Detección de Patrones de Daños y Averías en la Industria Automotriz*. Tesis de Maestría en Ingeniería en Sistemas de Información. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional.