

Aplicación de técnicas de *Business Intelligence* para el almacenamiento y análisis de datos en la dirección de proyectos informáticos

Esther Hochsztain(*), Andrómaca Tasistro(**)

(*)Departamento de Métodos Cuantitativos

Facultad de Ciencias Económicas y de Administración. UDELAR

esther@ccee.edu.uy

(**) Agencia de Gobierno Electrónico y Sociedad de la Información (AGESIC)

andromaca.tasistro@agesic.gub.uy

Resumen

La dirección de proyectos informáticos debe enfrentar numerosos desafíos, entre ellos la identificación de los principales criterios que conducen al éxito (o no) de los mismos. Los gerentes de proyecto generan muchos datos, que se almacenan en diversos formatos, pero en la mayoría de las organizaciones su uso no se encuentra sistematizado con el objetivo de “aprender de los datos” y generar conocimiento reutilizable. El objetivo del artículo es presentar un framework basado en técnicas de Business Intelligence que contribuya a la mejora de la dirección de proyectos informáticos.

Palabras clave: dirección de proyectos Business Intelligence, Data Warehousing, PMI, CMMI

Áreas de conocimiento: Administración, Métodos Matemático-Cuantitativos, Computación

1. Introducción

El marco teórico de este trabajo se centra en la metodología de proyecto del *Project Management Institute* (PMI) y en el modelo *Capability Maturity Model Integration* (CMMI) de calidad de procesos.

Muchos trabajos se han realizado procurando superar lo que muchos consideran la principal debilidad de la industria del software buscando determinar los factores de éxito/fracaso de los proyectos, pero pocos utilizan la enorme cantidad de documentos almacenados para generar datos que puedan analizarse desde la óptica de las técnicas de Business Intelligence. Es bien conocida la denominada paradoja de Cobb, "Sabemos por qué fallan los proyectos, sabemos cómo prevenir las fallas, entonces ¿porqué fallan todavía?"

Por consiguiente el objetivo del proyecto se centra en generar un marco conceptual para responder la pregunta. La metodología se basa en el desarrollo de un marco conceptual (*framework*), que abarca las siguientes etapas: 1) determinación de requerimientos, 2) definición de las fuentes de datos, 3) diseño e implementación de los cubos del Data Mart, 4) determinación de patrones mediante consultas OLAP y técnicas de Data Mining, 5) evaluación de patrones. En todas las etapas se utiliza software de uso público, en particular MySQL, Pentaho, Weka, R y Tanagra.

Los resultados preliminares permiten identificar las principales características que permiten anticipar el éxito/fracaso de los proyectos informáticos. Los mismos son un insumo considerado relevante en la toma de decisiones al brindar información que permite tomar acciones correctivas en menor tiempo (que si no se utilizaran). [Lipke and Chapter, 2013]

2. Marco Teórico

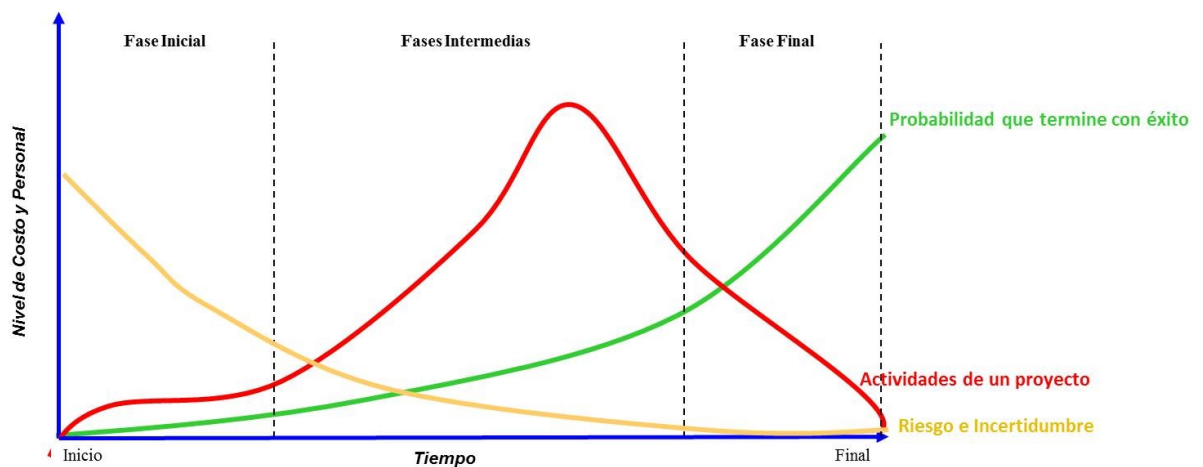
En esta sección se presentan el conjunto de buenas prácticas actualmente más difundidas para la gestión de proyectos. Estas recomendaciones son las propuestas por PMI y por CMMI.

2.1. *Project Management Institute (PMI)*

PMI es una organización Internacional sin fines de lucro fundada en 1969 en USA, cuyo objetivo es la profesionalización del gerenciamiento de proyectos. Cuenta con más de 300.000 miembros, profesionales de diferentes disciplinas, representando más de 125 países.

Difunde la profesión mediante publicaciones, congresos, capítulos locales y otras actividades de desarrollo profesional. Elabora y mantiene, entre otras normas, un compendio de conocimientos, teorías y prácticas más aceptadas en dirección de proyectos: el PMBOK® [Rose, 2013]

Según se define en [Rose, 2013] el ciclo de vida de un proyecto es una colección de fases generalmente secuenciales y a veces superpuestas, cuyos nombres y cantidades dependerán de las necesidades de dirección y control de la organización u organizaciones involucradas en el proyecto, la naturaleza del proyecto en sí misma, y sus áreas de aplicación. Esto puede verse en la figura 1



Notas:

- Aún cuando muchos ciclos de vida tienen nombres de fases similares, y requieren entregables similares, muy pocos son idénticos
- Los interesados tienen más poder de influir en las características y en el costo del producto (entregable del proyecto)

Figura 1: Ciclo de vida de los proyectos

Los problemas que contribuyen a la falla de proyectos pueden verse en la imagen ya clásica de (<http://www.alfonsovillar.com/wp-content/uploads/2011/11/projects.jpg>) que puede verse en la figura 2. En dicha imagen se aprecian gráficamente las visiones de los diferentes involucrados en los proyectos:

1. Cómo el cliente explicó sus requerimientos
2. Cómo el líder del proyecto lo entendió
3. Cómo lo diseñó el analista
4. Cómo lo describe el consultor de negocio
5. Cómo está documentado el proyecto
6. Qué se instala

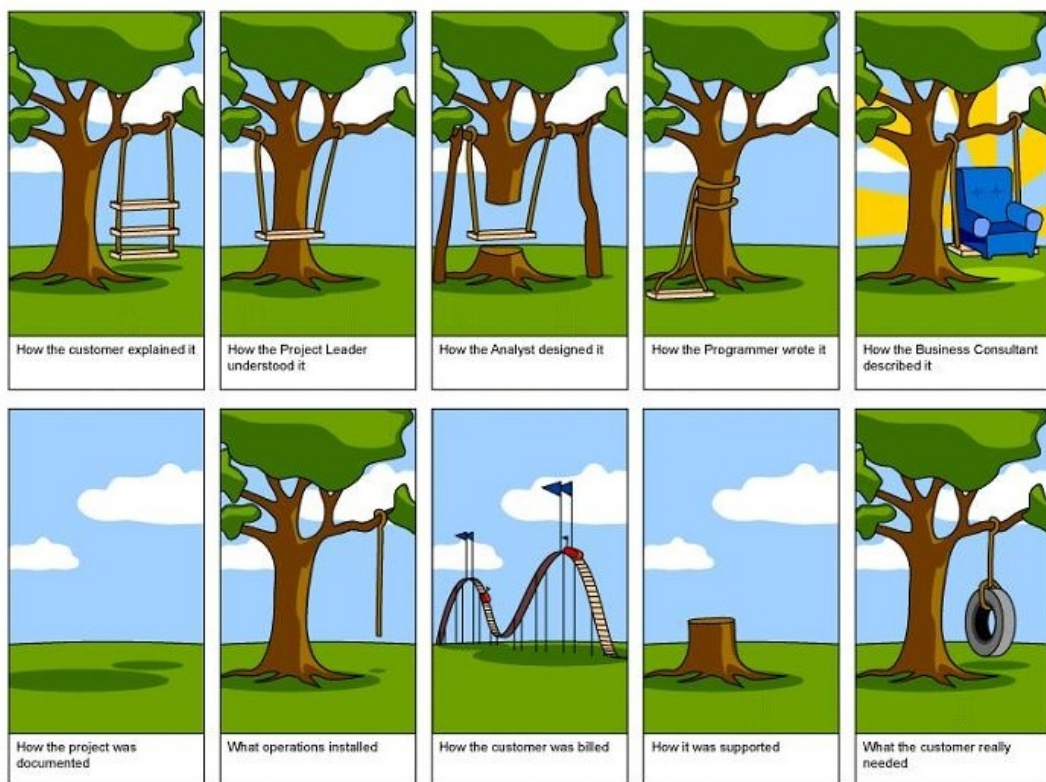


Figura 2: Distintas visiones de los proyectos

7. Qué pagó el cliente
8. Qué soporte se obtiene
9. Qué es lo que el cliente realmente necesitaba

La misión de un Gerente de Proyectos es justamente articular todas estas distintas visiones que se ha caricaturizado en la figura de forma que todas fluyan en un mismo sentido y que se pueda concluir con un proyecto exitoso.

Como dice [Verma, 1996] las personas son la columna vertebral y el recurso más valioso para ejecutar con éxito cualquier proyecto. Para sobrevivir y crecer en el siglo 21, los profesionales de gestión de proyectos deben aprender y utilizar habilidades interpersonales apropiadas que inspiran a todos los involucrados en un proyecto. Existen pautas prácticas que pueden ser utilizados para desarrollar e implementar las prácticas de comunicación, motivación, negociación, resolución de conflictos, los conflictos y el manejo del estrés y el liderazgo.

Para lograr el éxito en la dirección del proyecto, el gerente debe:

1. Seleccionar los procesos apropiados que sean requeridos para alcanzar los objetivos del proyecto;
2. Adoptar un enfoque claro y definido para identificar y cumplir con los requisitos del proyecto;
3. Cumplir con los requisitos de acuerdo a las expectativas y necesidades de los actores involucrados en el proyecto;
4. Balancear adecuadamente las demandas de alcance, tiempo, costos, calidad, recursos y riesgos para alcanzar el producto, resultado o servicio especificado para el proyecto.

En [Rose, 2013] se explica que es fundamental la triple restricción que debe cumplirse, como se ve en la figura 3 que aparece en (<http://mabsistemas.com/img/graficas/1.png>). Esta restricción implica que no se puede modificar una de las dimensiones sin afectar a las otras dos. Si se modifica el alcance obviamente se modificará el tiempo, el costo o ambos. De la misma forma, si se atrasan los tiempos, esto implicará más costo, o reducir el alcance si no hay fondos disponibles. Lo mismo ocurre si por algún motivo se incrementan los costos.

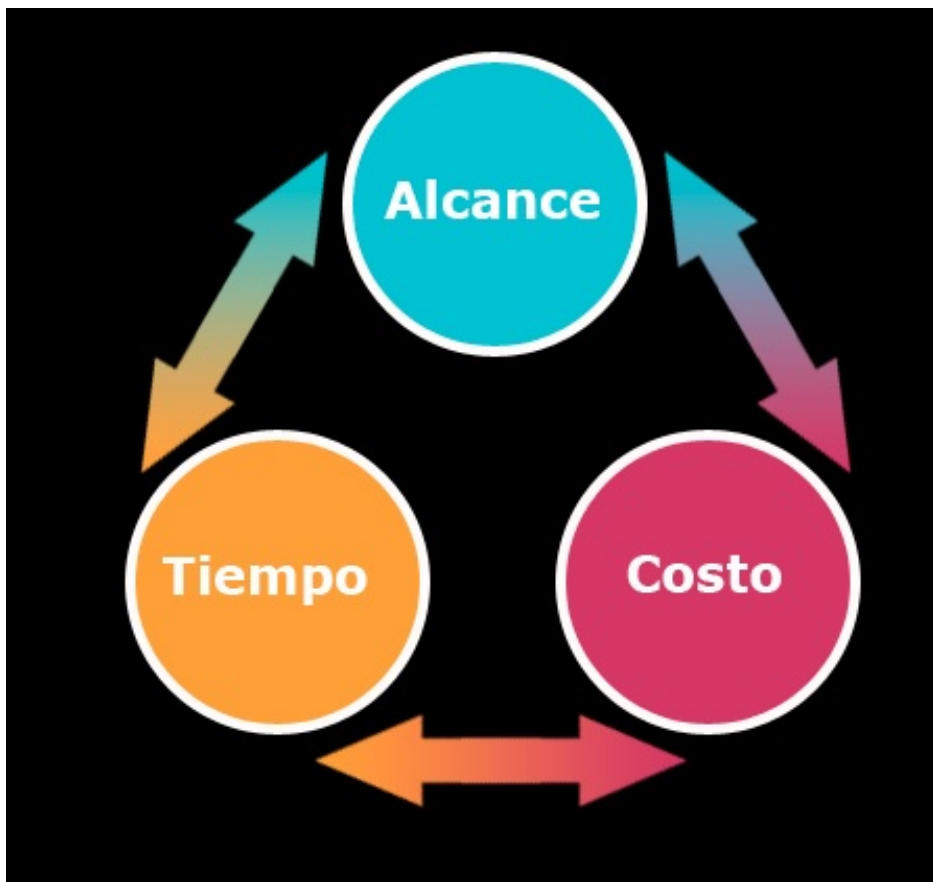


Figura 3: Triple restricción en los proyectos

Uno de los procesos fundamentales en el ciclo de un proyecto es el cierre del mismo. El propósito de dicho proceso es proveer un punto fijo en el cual se pueda: confirmar la aceptación del resultado del proyecto; verificar que los objetivos trazados al inicio del proyecto, e incluidos en el alcance del proyecto han sido alcanzados; reconocer que el proyecto no tiene nada más que contribuir.

Un proyecto concluye o es finalizado cuando: sus objetivos se han alcanzado; sus objetivos no podrán ser alcanzados en el tiempo y con el costo proyectado; o cuando desaparece la necesidad del proyecto.

Decimos que un proyecto será exitoso sólo en el primero de estos casos (se alcanzan los objetivos dentro del plazo y presupuesto proyectado). En todos los demás será necesario ver cuales son las causas del fracaso, para prevenirlas en futuros proyectos. En este sentido este artículo provee mecanismos para realizar dicho estudio.

2.2. *Capability Maturity Model Integration (CMMI)*

CMMI es un método de probada eficacia para la gestión del rendimiento, con décadas de resultados que muestran que funciona. Las organizaciones que utilizan CMMI tienen coste predecible, calendario y resultados de calidad de negocios que sirven como discriminadores entre sus competidores.

CMMI se construye [Technology, 2013] con las prácticas y los objetivos vistos en miles de organizaciones en el mundo. Pueden utilizarse estas prácticas y metas para evaluar el desempeño propio y decidir qué mejorar para sus propias razones de negocios. Las dimensiones críticas de una empresa son: la gente, los procedimientos y métodos, y las herramientas y equipo. Los procesos son los encargados de unir tales dimensiones con el propósito de alcanzar los objetivos del negocio. El enfoque en los procesos ayuda a construir una plataforma de mejora continua, ya que se está de acuerdo en que la gente y la tecnología cambian y son sólo los procesos los que trascienden en el tiempo, adaptándose a nuevas personas y tecnologías.

En el mercado actual existen modelos de madurez, estándares, metodologías y guías que pueden ayudar a una organización a mejorar la forma de hacer su negocio. Sin embargo, la mayoría de los enfoques de mejora existentes se centran en una parte específica de su actividad y no tienen un enfoque sistemático de los problemas a los que se enfrentan la mayoría de las organizaciones. Desafortunadamente, al centrarse en mejorar un área de negocio, estos modelos han hecho que persistan los nichos y las barreras

existentes en el seno de las organizaciones.

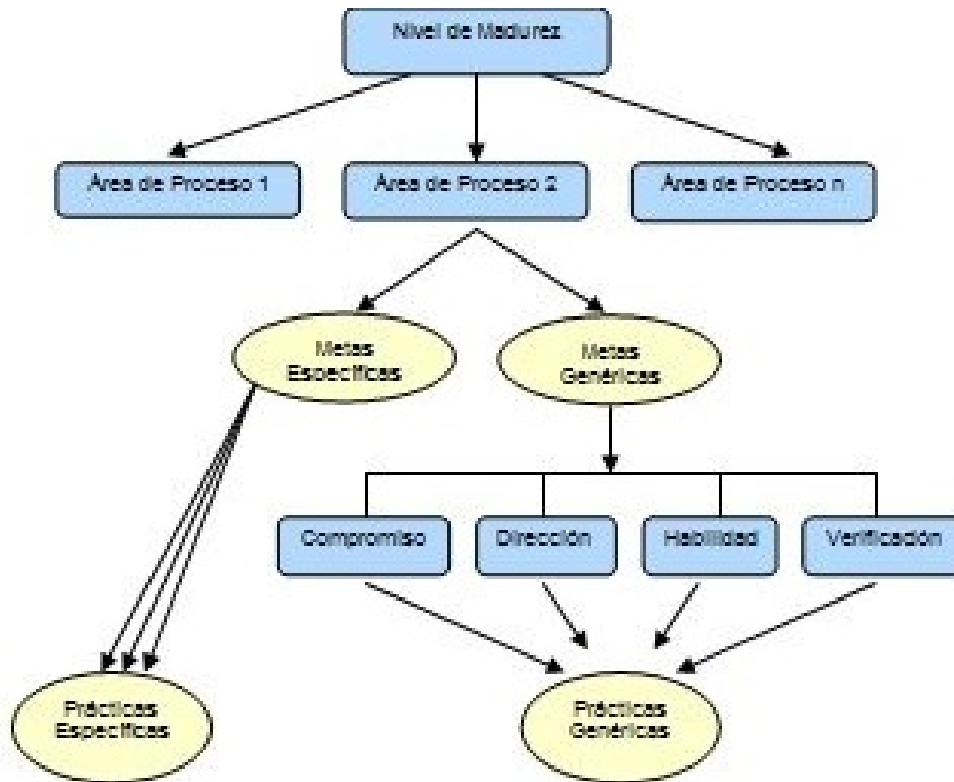


Figura 4: Modelo CMMI

El Software Engineering Institute (SEI) de la Carnegie Mellon University de los Estados Unidos, creador del modelo CMMI y de la mayoría de sus predecesores, ha elaborado sus modelos bajo la premisa que la calidad de un producto o servicio está altamente influenciada por la calidad de los procesos que los producen y los mantienen [Team, 2006]. Es por ello que la mejora continua de los procesos debiese ir paulatinamente incrementando el nivel de capacidad y madurez de una organización, esto se esquematiza en la figura 4

Los procesos en conjunto transitan desde procesos no definidos, es decir, procesos cuya organización cuenta con poca capacidad y con inmadurez para realizarlos, a procesos disciplinados cuya organización cuenta con la capacidad y madurez suficiente para desarrollarlos con calidad probada.

Luego una organización es capaz de definir su calidad total por medio del nivel de madurez de capacidades en que se encuentre de acuerdo a sus procesos.

El SEI, en sus investigaciones para ayudar a las organizaciones a desarrollar y mantener productos y servicios de calidad, ha identificado tres dimensiones en las que una organización puede centrarse para mejorar su actividad:

1. las personas,
2. los métodos y procedimientos, y
3. el equipamiento y herramientas.

Para mantener todo unido y aceitado es necesario mejorar en primer lugar los procesos utilizados en su organización. Éstos le permiten alinear su modo de trabajar, abordar la escalabilidad y proporcionan una forma para incorporar el conocimiento de cómo hacer las cosas mejor. Los procesos le permiten explotar mejor sus recursos y analizar las tendencias de su actividad.

3. Project Management Framework (PMF)

El gerenciamiento de proyectos es una disciplina que, no cabe duda, en muchas ocasiones resulta ser casi o tanto más crítica que la solución 'técnica' de los proyectos informáticos. Si bien tiene ya varios años la bien conocida paradoja de Cobb [Cobb et al., 1989] "Sabemos por qué fallan los proyectos; sabemos cómo prevenir sus fallos – así es que ¿por qué siguen fallando?", aún se continúa enfrentando el problema señalado.

En esta propuesta se busca una solución a la línea de pensamiento planteada [Hillson, 2012], en el sentido de que podría ser imposible resolver la Paradoja de Cobb, por los siguientes motivos:

- Todos los proyectos son arriesgados
- La mayoría de los proyectos incluyen riesgos inmanejables
- La gestión de riesgos no siempre se hace bien
- Las actas de constitución a menudo omiten los umbrales de riesgo
- Los proyectos deberían existir en un portafolio balanceado de riesgo
- La innovación se genera en el fallo

- Fallo al aprender

En busca de brindar elementos que permitan superar los problemas planteados, se presenta en este artículo un *Framework* para que a través del conocimiento y análisis del riesgo, sea posible aprender de los errores, innovar y contribuir llegar a un mundo en que sea más probable que los proyectos informáticos sean exitosos (en palabras de Sir Peter Gershon "Los proyectos y programas deberían entregarse en coste, en tiempo, cumpliendo con los beneficios anticipados").

En la Figura 5 se presentan los diversos niveles que se tienen en cuenta el el proceso, partiendo del inferior de la pirámide en niveles transaccionales hasta llegar a los niveles superiores, al aumentar el potencial de toma de decisiones.

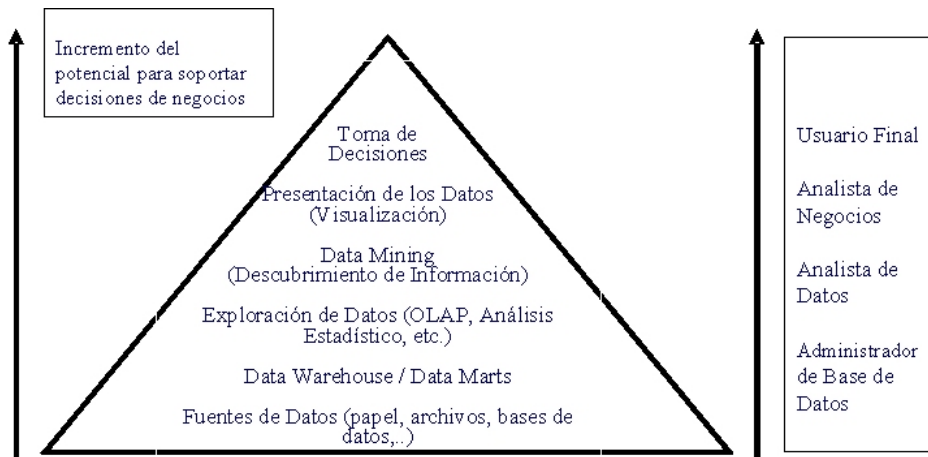


Figura 5: Niveles en la aplicación de técnicas de *Business Intelligence*

En la Figura 6 presentan las etapas del *Framework*. Dado que el proceso es iterativo, las flechas siempre se plantean en ambos sentidos al conectar las etapas. Sin embargo, lo indispensable y más frecuente son los vínculos en sentido horario. Sin embargo, puede ocurrir que sea necesario retornar a la etapa anterior, por lo cual se presentan las flechas en sentido antihorario.

A continuación se detalla cada una de las fases anteriormente indicadas.

1. **Determinación de requerimientos:** En esta fase es necesario identificar los diferentes actores en el escenario de los proyectos y los requerimientos específicos planteados. Se sigue la propuesta planteada por [Xu et al., 2012], en el sentido de utilizar un método incremental iterativo. Este método permite descomponer, refinar y especificar los requisitos, aclarando y comprendiendo los mismos. Algunos ejemplos de

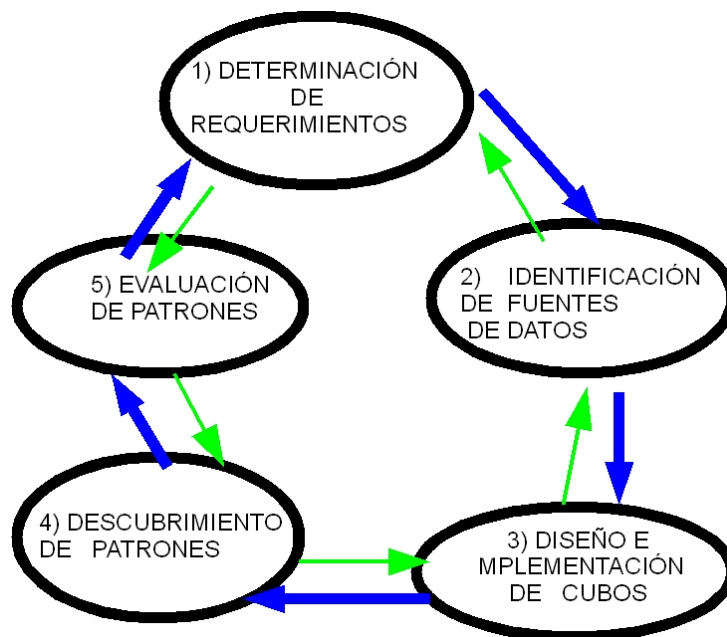


Figura 6: Estructura del *Project Management Framework*

requisitos fueron planteados anteriormente, identificar los factores de contribuyen al éxito/fracaso de los proyectos, contar con información precisa, relevante y fácil de consultar por los diversos involucrados, construir modelos que permitan explotar los datos almacenados, entre otros.

2. **Definición de fuentes de datos:** El origen de datos es muy variado. Las principales fuentes de datos utilizadas contienen datos de :

- a) características del proyecto (área, recursos disponibles, grado de involucramiento de los diversos involucrados, entre otros)
- b) herramientas de gestión de proyectos utilizadas
- c) habilidades de los gerentes (en base a las 12 señaladas por [Chang and Torkzadeh, 2013])
- d) aspectos socioculturales (siguiendo las 7 líneas planteadas por [Avison and Torkzadeh, 2013])
- e) entorno nacional y regional entorno nacional, regional, mundial
- f) modalidades de comunicación utilizadas y herramientas que las soportan

Las fuentes de información para obtener los datos anteriormente mencionadas son: bases de datos disponibles, información documental, creación de formularios autoadministrados, entrevistas y observación, esencialmente.

3. **Diseño e implementación de cubos:** Los DataWarehouses están tomando cada vez más importancia en todo tipo de organizaciones, dado que se está pasando de centrar de el énfaseis en solamente la recolección de datos, al énfasis en brindar métodos cada vez más sencillos de análisis de los mismos. Según la definición del 'padre' del *Data Warehouse* [Inmon, 2000] consiste en una colección de datos orientada a un determinado ámbito (orientada al tema: empresa, organización, etc.), integrado, no volátil, historiado y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza. Consiste en un repositorio de de datos que brinda historiales muy diferentes de la habitual información transaccional y operacional. Además está almacenado en una base de datos diseñada para favorecer el análisis y la divulgación eficiente de datos, favoreciendo el hallazgo y comunicación de la información contenida en los datos (especialmente con herramientas OLAP, *On Line Analytical Processing*). En la Figura 7 se presenta la estructura de un *Data Warehouse* y los diversos niveles de agregación de los datos.

De las metodologías utilizadas para el diseño y construcción de un *Data Warehouse* se considera que la más versátil es la propuesta por [Kimball, 1998]. La misma se ha mantenido en permanente evolución, con continuas reediciones ampliadas hasta el año actual [Kimball and Ross, 2013].

Los cubos (más formalmente hipercubos) permiten caracterizar la información de los proyectos, y así identificar en grandes líneas interesantes patrones.

4. **Descubrimiento de patrones:** El descubrimiento de patrones se realiza en base a los datos almacenados en el *Data Warehouse*, aunque también se analiza la información documental disponible y se trabaja en base a datos transaccionales y operativos. En particular se centra en tres tipos de grandes líneas de análisis:

- Consultas OLAP (*On Line Analytical Processing*)

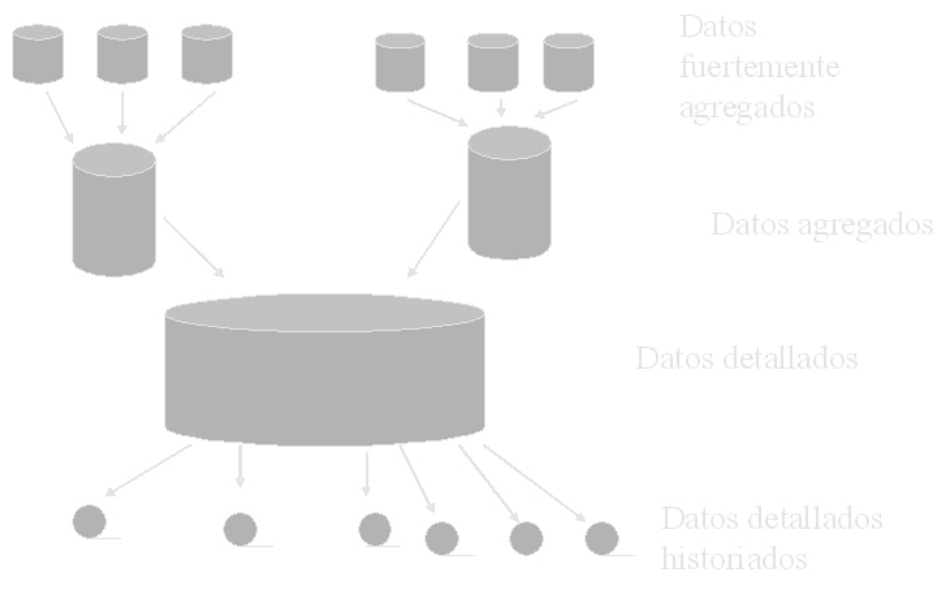


Figura 7: Estructura de *Data Warehouse*

- Técnicas de *Data Mining* [Fayyad et al., 1996], [Chatzikonstantinou et al., 2013]
- Técnicas de *Text Mining* [Tan et al., 1999], [Yu and Hsu, 2013], [Jiang et al., 2013]

La Minería de Datos y de Textos es un área interdisciplinaria en la que confluyen múltiples áreas de conocimiento, con el objetivo de 'descubrir' el conocimiento contenido en los datos. Requiere de un proceso que se presenta en la Figura 8.

5. **Evaluación de patrones:** El uso de los patrones descubiertos se espera que sea útil para contribuir a la toma de mejores decisiones, reduciendo la incertidumbre y contribuyendo tanto a identificar proyectos exitosos como a identificar automáticamente lo antes posible desvíos que se consideren 'peligrosos'. De todos los patrones obtenidos en la anterior etapa, es necesario seleccionar aquellos que realmente constituyen las 'pepitas de conocimiento' buscadas.

4. Conclusiones y Futuros Trabajos

Actualmente el proyecto cuenta con un prototipo en etapa experimental, al que se han cargado datos de un conjunto de proyectos, en variadas áreas,

El proceso de Data/Text Mining

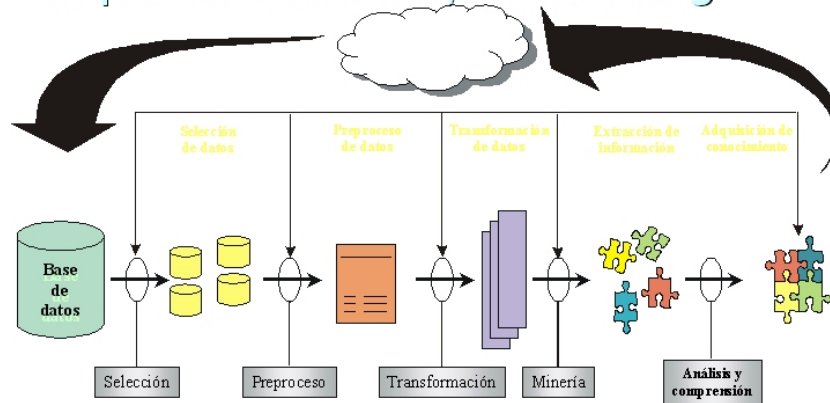


Figura 8: Disciplinas de confluyen en Data Mining

con muy disímiles presupuestos y perfil tanto de gerentes como de usuarios. En base al estudio se han identificado algunos factores que preliminarmente pueden ser considerados considerados críticos para el logro de los objetivos de los proyectos. Entre los mismos se ubican:

- Importancia de la capacidad de liderazgo de los gerentes de proyecto
- Importancia del *timing* en los proyectos, no es bueno que los hitos del proyecto coincidan con fechas importantes de la organización
- Importancia de la comunicación dentro del proyecto
- Importancia del registro de incidencias y la adecuada comunicación del mismo
- Aprender de los errores. Importancia del documento de lecciones aprendidas
- Disponer de mecanismos de control adecuados (por ejemplo contenido y calidad de los archivos cargados).
- Contar con métricas apropiadas para cuantificar todos los aspectos considerados relevantes

Este estudio resulta la primera etapa de un proyecto que permita que los diversos usuarios puedan cargar datos on-line y conformar una gran base que pueda usarse *on line* en principio a nivel nacional.

Bibliografía

- [Avison and Torkzadeh, 2013] Avison, D. and Torkzadeh, R. (2013). Getting the right balance in is project management: Seven guidelines for an equal emphasis on the sociocultural aspects.
- [Chang and Torkzadeh, 2013] Chang, J. C.-J. and Torkzadeh, G. (2013). Perceived required skills and abilities in information systems project management. *International Journal of Information Technology Project Management (IJITPM)*, 4(1):1–12.
- [Chatzikonstantinou et al., 2013] Chatzikonstantinou, G., Kontogiannis, K., and Attarian, I.-M. (2013). A goal driven framework for software project data analytics. In *Advanced Information Systems Engineering*, pages 546–561. Springer.
- [Cobb et al., 1989] Cobb, D. F., Smith, B., and Yocom, J. (1989). *Douglas Cobb's Paradox 3 Handbook*. Bantam Books, Inc., New York, NY, USA, 2nd edition.
- [Fayyad et al., 1996] Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (1996). Advances in knowledge discovery and data mining.
- [Hillson, 2012] Hillson, D. (2012). Resolución de la paradoja de cobb. *PMI. Capítulo Madrid-España*. <http://www.pmi-mad.org/>.
- [Inmon, 2000] Inmon, W. (2000). What is a data warehouse?
- [Jiang et al., 2013] Jiang, S., Zhang, H., and Zhang, J. (2013). Research on bim-based construction domain text information management. *Journal of Networks*, 8(6):1455–1464.
- [Kimball, 1998] Kimball, R. (1998). *The data warehouse lifecycle toolkit: expert methods for designing, developing, and deploying data warehouses*. Wiley. com.
- [Kimball and Ross, 2013] Kimball, R. and Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. Wiley.
- [Lipke and Chapter, 2013] Lipke, W. and Chapter, P.-O. C. (2013). Is something missing from project management? *CrossTalk*, page 17.
- [Rose, 2013] Rose, K. H. (2013). A guide to the project management body of knowledge (pmbok® guide)—fifth edition. *Project Management Journal*, 44(3):e1–e1.

- [Tan et al., 1999] Tan, A.-H. et al. (1999). Text mining: The state of the art and the challenges. In *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*, pages 65–70.
- [Team, 2006] Team, C. P. (2006). Cmmi for development, version 1.2.
- [Technology, 2013] Technology, S. A. C. (2013). Security by design with cmmi for development, version 1.3: An application guide for improving processes for secure products. CMMI Institute.
- [Verma, 1996] Verma, V. (1996). The human aspects of project management: human resource skills for the project manager, volume two. Project Management Institute.
- [Xu et al., 2012] Xu, X., Zhang, B.-f., and Lin, J. (2012). Management information system requirements analysis model based on the agile development. In *Control Engineering and Communication Technology (ICCECT), 2012 International Conference on*, pages 986–990. IEEE.
- [Yu and Hsu, 2013] Yu, W.-d. and Hsu, J.-y. (2013). Content-based text mining technique for retrieval of cad documents. *Automation in Construction*, 31:65–74.