



Autores: Campanaro, Rosa, Ingrassia, Ramiro

Instituto de Investigaciones y Asistencia Tecnológica en Administración

Propuesta de Datamart para el análisis de las consecuencias de la dinámica empresarial.

1. Datawarehouse – Datamining y la Responsabilidad Social Empresaria

1.1- Sistemas de Control

A medida que la dinámica del mercado y la competencia fueron incrementándose, la Administración necesitó de sistemas de información con mayor capacidad para trabajar la complejidad impuesta por los aquellos. Por tal motivo, no sólo se requirió de mejores recursos de hardware y software sino de nuevos paradigmas en el campo del 'modelaje lógico'. En Marchese, A., Gaibazzi, M., Pérez, N., Ingrassia, R. (2009), se planteó la evolución de las corrientes de Administración, las cuales permitieron la incorporación y apropiación de nuevas herramientas para la gestión empresarial.

La problemática del ambiente organizacional puede dividirse en dos momentos (Olcese, A., Rodríguez, M. y Alfaro, J., 2008): la gestión de productos tangibles, y por otro, la de intangibles. Específicamente, los autores citados se centran en la evolución del tablero de comando como herramienta de gestión empresarial.

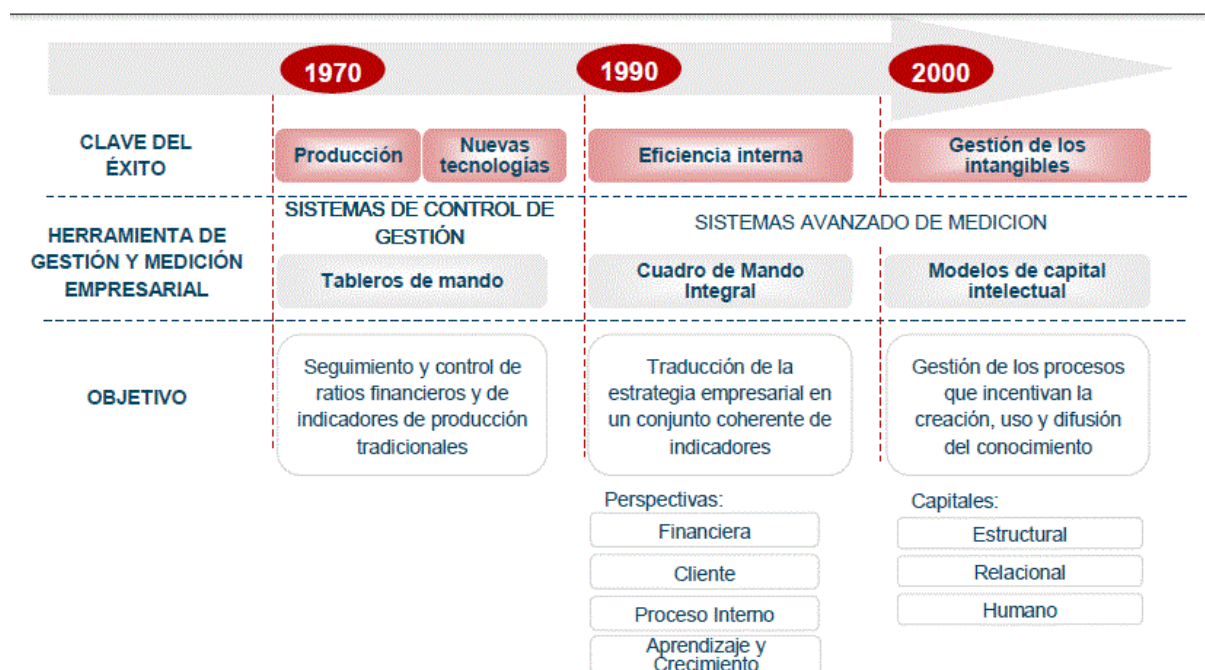


Ilustración 1: Evolución histórica de los sistemas de control de la gestión empresarial. En la presente línea histórica se puede visualizar la evolución de los sistemas de información en función de las claves de éxito imperantes en cada período. *Fuente:* Olcese, A. y otros, 2008, p. 121.



La gestión de productos tangibles corresponde a la concepción de Estrategia “como un medio de establecer el propósito de la organización en términos de sus objetivos a largo plazo, sus programas de acción y sus prioridades en materia de asignación de recursos.” (Hax, A., N., Majluf, 2012, p. 24) durante la hegemonía de la Escuela Neoclásica y posteriormente, la Escuela Burocrática. Los sistemas de información computacionales de aquella época eran los denominados **‘Sistemas de Procesamiento de Transacciones’**.

Hacia la década de 1970, comienza el incremento del dinamismo en la competencia empresarial y la herramienta de gestión fue el cuadro (o tablero) de mando.

Esta herramienta incorporaba diversos ratios para el control económico-financiero y de la calidad en los resultados de los procesos del negocio. Mediante este dispositivo el equipo directivo podía evaluar la evolución de esos campos de la gestión y tomar de una forma rápida las medidas correctivas necesarias. (Olcese, A. y otros, 2008, p. 117)

El cual es producto del surgimiento de los **‘Sistemas de Información Gerencial’**. Y desde el punto de vista de la Administración, las escuelas imperantes en dicha época fueron la ‘Teoría de las Organizaciones’ y ‘Teoría de los Sistemas’.

A partir de la década de 1990, la gestión comienza a incorporar algunas variables intangibles pero sólo con el objetivo de medir su eficiencia interna.

Una de las características más destacadas del cuadro de mando integral es la combinación de indicadores financieros y no financieros... los indicadores que incorpora el cuadro de mando integral de Norton y Kaplan se clasifican en cuatro perspectivas de valor: financiera, clientes, procesos internos y personas o crecimiento y aprendizaje. Los indicadores de la perspectiva financiera se deben explicar por los resultados de la perspectiva de clientes con los que tiene relación. Los indicadores de clientes por los resultados obtenidos en los de la perspectiva interna o procesos con los que están relacionados. Finalmente estos últimos están organizados y dispuestos en red con los existentes en la perspectiva de personas o de crecimiento y aprendizaje. (Olcese, A. y otros, 2008, p. 118)

El surgimiento de esta herramienta de gestión es el producto de la maduración, en la metodología científica, de la selección del universo de indicadores (expresivos y predictivos) y su posterior combinación para establecer conceptos clasificatorios de una variable. En el caso del cuadro de mando integral, se define a partir de qué conceptos (financiera, clientes, procesos internos y aprendizaje/crecimiento) se puede medir la eficiencia de una organización y, en consecuencia, clasificarla con la etiqueta ‘organización eficiente’ (o no eficiente).

Distinguiremos de dos tipos de indicadores: expresivos y predictivos...La distinción, como toda dicotomía de este tipo, sólo se justifica por razones de conveniencia, y se centra en los extremos de una serie. Se refiere al lugar que los indicadores ocupan en los procesos hipotéticos que median entre la observación originante y el conjunto de imágenes conceptuales que se desarrollan para organizarla. (Lazarsfeld, P., 1973, p. 27)

Cuando los indicadores están más correlacionados con la variable bajo análisis, se los denomina indicadores expresivos y tienden a estar más cerca de la observación originante. El autor antes citado, entiende como observación originante al proceso subyacente que permite explicar las diferencias y variaciones en una variable. Por otro lado, cuando los indicadores están próximos al conjunto de imágenes conceptuales, se los denomina indicadores predictivo porque “indican” el efecto o causa del valor de la variable.

En el ejemplo del cuadro de mando integral, las perspectivas están correlacionadas entre sí y como se expuso anteriormente, cada perspectiva está relacionada con la anterior. Es



decir, las sucesivas perspectivas explican en última instancia la situación financiera pero no se puede afirmar que son causas o consecuencias de aquellas. Por tal motivo, los indicadores utilizados, en dicha herramienta de gestión, están más próximos a la observación originante y se los podría clasificar como indicadores expresivos.

A finales de la década de los '90, la ventaja competitiva se centralizó en la Gestión de los productos 'intangibles' como es el capital intelectual. El motivo central de la importancia de esta última es que la concepción del conocimiento implica abarcar todos los aspectos de un fenómeno. De esta manera, el administrador no sólo tiene que saber si una 'organización es eficiente o no' (fenómeno) sino cuáles son las causas que permiten lograr la eficiencia en las organizaciones. En consecuencia, continuando con las herramientas de gestión orientadas a indicadores, la Administración necesita un conjunto combinado de indicadores expresivos y predictivos.

En éste último, radica la problemática de la gestión de productos intangibles. Por ejemplo, la experiencia permite dilucidar aquellas dimensiones de las variables que permiten 'predecir' el comportamiento de una variable. ¿Qué herramienta se tiene disponible para aumentar la experiencia?

Los sistemas informáticos permiten soportar esta tarea con los denominados **Sistemas de Apoyo a las Decisiones controlados por modelos** (Laudon y Laudon, 2002). Dicho tipo de sistema permite modelar matemáticamente las variables (a través de las dimensiones pertinentes) y crear distintos escenarios. Para poder establecer las ecuaciones matemáticas o seleccionar las dimensiones oportunas, están los **Sistemas de Apoyo a las Decisiones controlados por datos** (Laudon y Laudon, 2002). En estos sistemas de información, el usuario puede extraer información y definir la multidimensional (establecer dimensiones – data warehouse-), y luego, analizarla de tal manera que pueda encontrar patrones y relaciones ocultas (datamining). Y por último, dichas relaciones incorporarlas a la simulación.

De esta manera, no se tiene límites en la incorporación de dimensiones de una variable ni tampoco al número de variables bajo análisis. En consecuencia, esta tecnología permite integrar a la Responsabilidad Social Empresaria a la Administración, específicamente a la gestión de productos intangibles. Actualmente, a la gestión que utiliza este tipo de tecnología se la denomina "Business Intelligence".

Desde esta última perspectiva, el valor agregado de estas tecnologías radica en la posibilidad de visualizar y modelar las externalidades que generan las empresas durante su actividad económica. Y desde la Administración Pública, incorporar todas aquellas variables que integran el entramado social de la interacción "sociedad-empresa-estado".

1.2- Data warehouse - Datamining

Qué entendemos por un "data warehouse", para entenderlo tomamos la definición de W.H.Inmon (Dyché, 2001, p. 34) "Un Data Warehouse es una colección de datos orientados a temas, integrados, variantes en el tiempo y no volátiles, que tiene por objetivo dar soporte a la toma de decisiones". De esta definición surgen no solo las características del Data Warehouse, sino las diferencias con las bases de datos transaccionales u operacionales, que volcamos en el siguiente cuadro:



Tabla 1: Cuadro comparativo de la Base de datos transaccionales y Data Warehouse.

Base de Datos Transaccionales	Data Warehouse
Datos para las transacciones	Datos para la información
Orientados a la aplicación	Orientados al tomador de decisión
Originados por la aplicación	Integrados: múltiples orígenes: base de datos transaccionales – planillas de cálculos – archivos internos como externos
Volátil: los datos varían permanentemente, son actualizados para reflejar la situación a un momento. Ej. el stock de un producto está actualizado al momento	No Volátil: los datos no se modifican, guardan su situación vinculados a períodos. Ej. el stock se tendrá para cada fecha que contenga la historia de ese producto.
Temporales, responden a períodos de tiempo ligados a la operación.	Histórico, se guardan altos períodos de tiempo para comparaciones, tendencias y proyecciones.
Consultas predefinidas y actualizables	Consultas complejas, no anticipadas
Pensados para soportar muchos usuarios concurrentes	Pocos usuarios concurrentes
Muchos accesos que implican pocos registros a procesar -- Datos Normalizados (sin redundancias perjudiciales)	Pocos accesos que implican procesamiento de alto volumen de registros – Se admiten redundancias, guardar datos calculados, etc. para mejorar la performance
Estructura Relacional	Estructura Multidimensional - Cubos

Fuente: Elaboración propia.

Las diferencias entre el *Procesamiento en Línea de Transacciones* (OLTP), propio de las based de datos transaccionales, y el *Procesamiento Analítico en Línea* (OLAP), relacionado al datawarehouse, pueden resumirse en:

a) OLTP:

- Sistemas transaccionales, enfocados a gestionar un gran número de transacciones concurrentes.
- Permiten insertar, actualizar, borrar y consultar una pequeña cantidad de registros.
- Base de Datos Transaccionales

b) OLAP:

- Enfocados al análisis de grandes cantidades de datos.
- Proporcionan respuestas rápidas y complejas.
- Data Warehouse. Bases de Datos Multidimensionales.
- El modelado dimensional, permitirá ver los datos como aristas de un cubo.

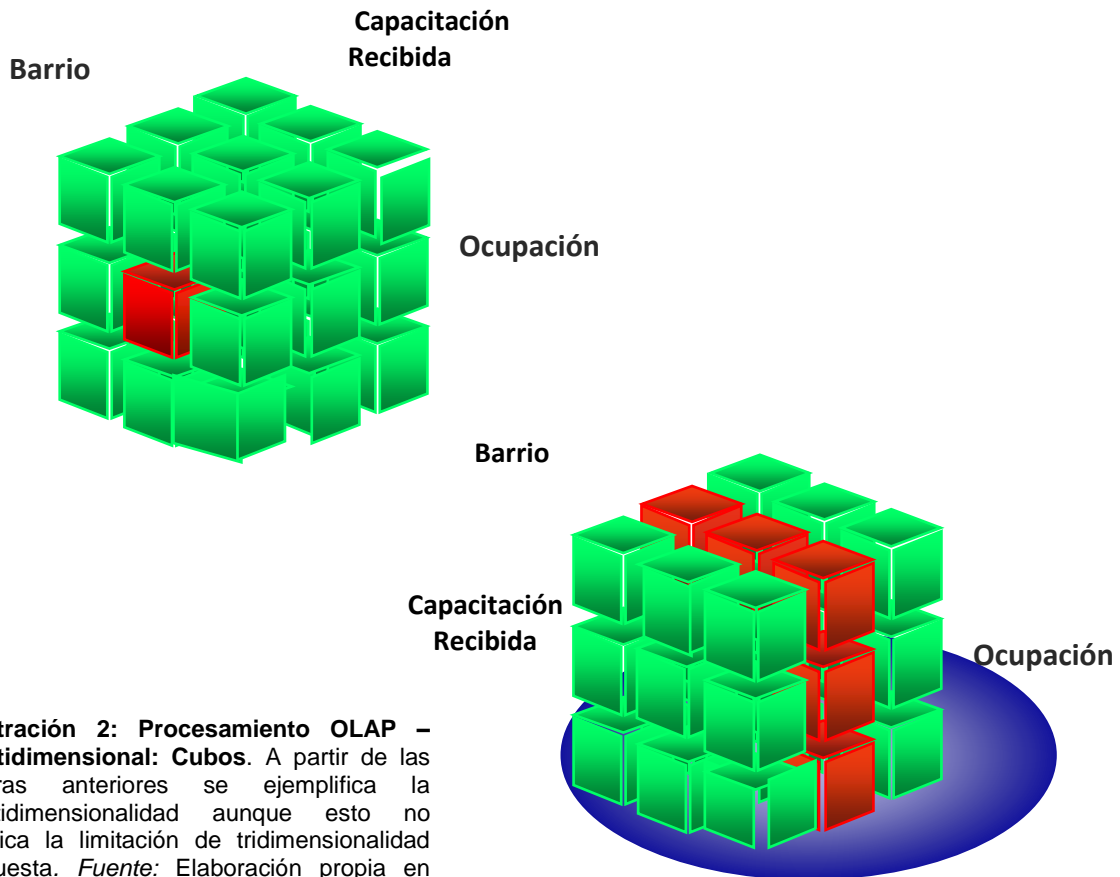


Ilustración 2: Procesamiento OLAP – Multidimensional: Cubos. A partir de las figuras anteriores se ejemplifica la multidimensionalidad aunque esto no implica la limitación de tridimensionalidad expuesta. *Fuente:* Elaboración propia en base al esquema propuesto por Bertello (2003).

Que es un “Data-Mart”, es la construcción de pequeños almacenamientos de datos (pequeños Data Warehouse) que llevan menor tiempo de construcción y permiten rápidamente obtener resultados, esto también permite flexibilidad si se quisiese completar o ampliar alguna visión de la información.

1.3- Metodología para la construcción del Data Warehouse/Data Mart

Se decidió adoptar la metodología HEFESTO (Bernabeu, 2007) ya que sus características principales, coincidían con las necesidades que teníamos para llevar a cabo el proyecto:

- ✓ Que fuese clara y sencilla.
- ✓ Independiente de las herramientas que se utilicen en su implementación.
- ✓ Independiente de las estructuras físicas del Data Warehouse.
- ✓ Se aplique tanto para Data Mart como para Data Warehouse.

Para que esto no sea un proyecto caótico o desintegrado, hay dos estrategias para su construcción:

- **Estrategia Top-Down:** que consiste en diseñar primero el “Data Warehouse” y a partir de él se construirán los distintos Data-Marts. Cuál es el inconveniente, lleva un elevado tiempo de análisis y hay que conocer en detalle todos los datos que lo integrarán y sus interrelaciones, cualquier cambio en una aplicación impactará en todo el Data Warehouse.



- **Estrategia Botton Up:** por el contrario, se construirán Data-Marts para cada módulo y aspecto a relevar para luego integrarlo en un gran Data Warehouse.

1.4- Trabajos anteriores

El actual grupo de investigación comenzó a estudiar la temática de Responsabilidad Social Empresaria en la ciudad de San Lorenzo como nuevo paradigma de gestión empresaria en el proyecto de *investigación "El proceso de conocimiento en las organizaciones mediante la aplicación de herramientas informáticas"* (2007 – 2010), y como continuación, *"Aplicación de herramientas informáticas de conocimiento a la gestión de responsabilidad social"* (2011-2014). Los trabajos presentados por los miembros de los respectivos proyectos pueden resumirse en:

Marchese, A., et. al. (2009), se planteó las disciplinas que se vinculan a partir de la temática de la Responsabilidad Social Empresaria y un resumen de la evolución de cada disciplina. Luego, se identificó la problemática urbanística de la ciudad de San Lorenzo y los distintos actores sociales involucrados en esta. Finalmente, se propuso un esquema de data warehouse para entrecruzar las distintas fuentes de datos de tales actores sociales.

Marchese, A., Ingrassia, R. (2010), se expuso el modelo de la triple hélice de Etzkowits para enmarcar la implementación del modelo de data warehouse formulado anteriormente. Por tal motivo, se presentó la encuesta de opinión sobre responsabilidad social realizada a los jefes de hogar de los barrios de la ciudad de San Lorenzo junto los datos provisorios sobre el relevamiento de las asociaciones civiles vecinales de la ciudad de San Lorenzo.

Marchese, A., Campanaro, R., Ingrassia, R., Miller, E., Díaz Toledo, S., Pérez, N., (2011), se presentó una experiencia de aplicación de herramientas de minería de datos y como resultado, se expuso los clusters (agrupamiento o conglomerados) de familias encontrados a partir de las variables: sexo, edad, enfermedades, edad del jefe de hogar, ocupación y subsidios del estado que recibe la familia.

2. Planeamiento de Objetivos

2.1- Objetivo General.

La aplicación de una de las herramientas de business intelligence para el análisis de la "Descripción socio-laboral del Jefe de Hogar" en el ámbito de la Responsabilidad Social Empresaria.

2.2- Objetivos Específicos

- Depurar los datos obtenidos del relevamiento a los jefes de hogar de los barrios aledaños a las principales industrias, en la ciudad de San Lorenzo;
- Definir las variables a utilizar, los indicadores (resultados) y dimensiones (perspectivas) a analizar en el Data Mart;

3. Metodología

La estructura del presente trabajo puede resumirse de la siguiente manera:

Introducción

De este modo, la introducción requiere que el autor establezca:



- (i) el marco contextual en el que se inserta el problema que se va a resolver: **Datawarehouse – Datamart. Y sistemas de información para la Responsabilidad Social Empresaria.**
- (ii) qué es lo que se sabe acerca del asunto en cuestión: el **crecimiento económico y poblacional,**
- (iii) que es lo que no se sabe: : **las interrelaciones de la dinámica empresarial con la sociedad circundante en la ciudad de San Lorenzo,** y
- (iv) qué representaría (económica, social, tecnológica, y/o científicamente) que sepamos lo que no se sabe: **la explicitación de las interrelaciones anteriores permiten un mejoramiento en las decisiones de carácter público-privado.**

Esta construcción de lo general a lo particular concluye naturalmente con el objetivo y/o la hipótesis del trabajo con lo que debería cerrarse la introducción.

Materiales y Métodos (¿Cómo se estudia el problema?)

- i. Tipo de estudio: **transversal.**
- ii. Área de estudio: **familias de la ciudad de San Lorenzo.**
- iii. Métodos y recolección de datos:
 - a. Fuentes primarias: **Encuesta Opinión de Responsabilidad Social (EORS),**
- iv. Tabulación y análisis de los datos: **datamart.**

El presente trabajo pretende sentar las bases para el análisis de los datos recopilados en la Encuesta de Responsabilidad Social, a través de un Datamart. Este tipo de análisis se puede enmarcar, dentro de las *Estrategias Generales* expuestas por Bryman y Burgess (Scribano, 2008), en la "Teoría Generativa" o "Estrategia de comparación constante". La cual se describe en los siguientes pasos:

- 1) Consiste en descubrir la teoría desde los datos.
- 2) Después de algunas recolecciones y puesta en relación con el problema, se procede a construir categorías.
- 3) Luego, se analiza su saturación, esto asegura su significado e importancia,
- 4) Posteriormente, se establecen relaciones y conexiones entre las categorías, proponiendo algunas hipótesis sobre el problema.
- 5) Luego, se vuelve al campo y se inicia un análisis de los esquemas generados. Finalmente, las interconexiones encontradas entre esquemas teóricos y datos vuelven a ser testadas. (Scribano, 2008, p. 139)

Y dentro del segundo enfoque, en que puede abordarse el análisis de datos, el denominado *Procesos Generales*, se adopta la estrategia de "Generación de Conceptos". En el cual las etapas son:

- 1) Inmersión en los datos y luego búsqueda de patrones,
- 2) Identificar posibles fenómenos sorpresas y,
- 3) Ser sensitivo hacia las inconsistencias.



4. Propuesta de Data Mart Lógico

El objetivo del "Business Intelligence" es optimizar el activo de datos para transformarlos en conocimiento, a través de distintas herramientas. De esta manera, si se aplica a nuestro caso ejemplificador se obtiene:

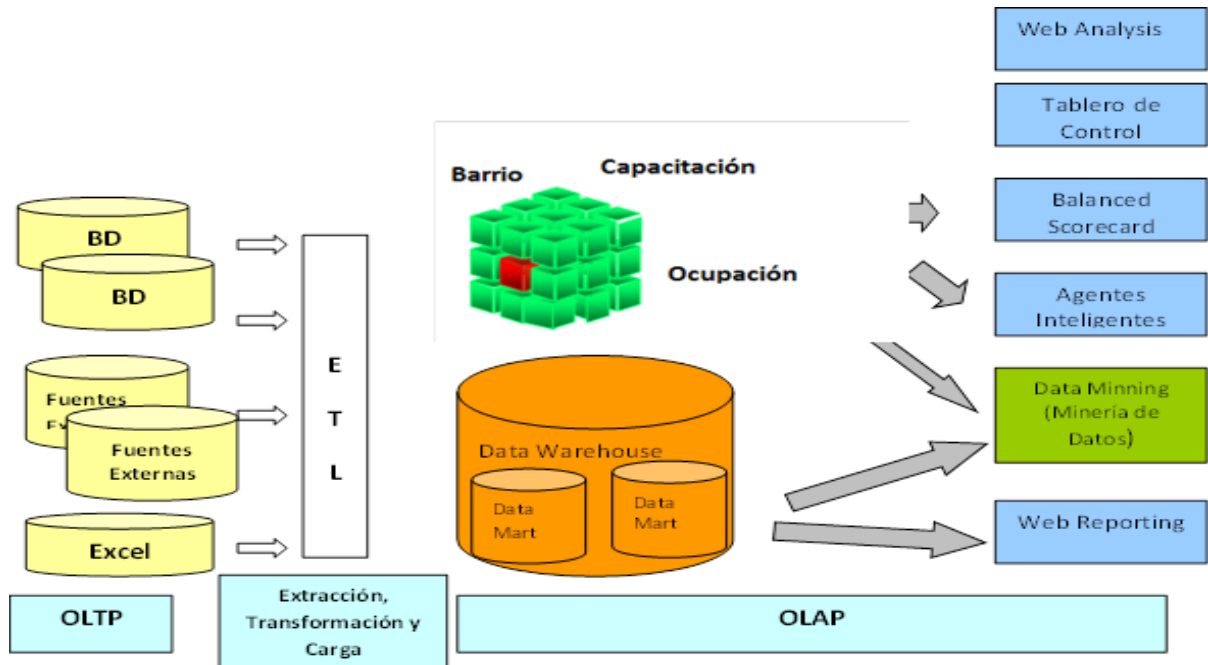


Ilustración 3: Arquitectura de Business Intelligence. En esta ilustración se expone el proceso de datawarehouse y datamining junto con la relación con otras herramientas de gestión. *Fuente:* Elaboración propia en base al esquema propuesto por Bertello (2003).

Para llegar a esta arquitectura, nos basamos en los distintos actores intervinientes (Marchese y otros, 2009, p. 9) sobre los que se propuso la realización de un relevamiento para construir un almacén de datos ("cubo") con la siguiente información, tal como se muestra en la Ilustración 4:

- De la municipalidad: información en relación a las inspecciones que realice tanto de índole medio ambiental como catastral para detectar los controles existentes sobre la información presentada por las empresas y/o proponer la recolección de información, si resultare necesario.
- De las empresas: la presentación del Balance Social, para visualizar su desempeño en la "interiorización de las externalidades".
- De las Cámaras Empresariales: Indicadores de mercado para monitorear el costo de eliminar las externalidades, por industria.
- De los centros educativos: Informes del desempeño de los estudiantes, para monitorear los efectos de las externalidades.
- De los barrios (o centros vecinales): las inquietudes propias de los ciudadanos.
- De los sindicatos: las prestaciones que se autorizan por medio de su obra social, para monitorear los costos de las externalidades.
- De los centros médicos: las prestaciones médicas realizadas, para monitorear los efectos de las externalidades.

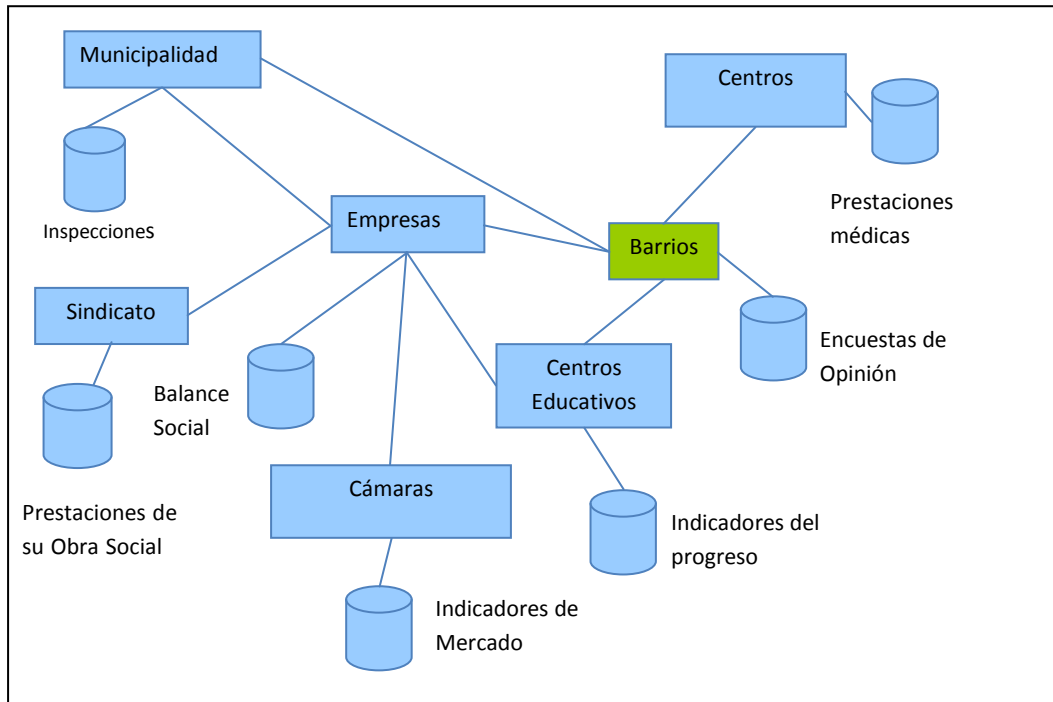


Ilustración 4: Esquema Informático. En la presente ilustración se visualizan los distintos actores identificados en la problemática social junto con las "bases de datos" de cada una de ellas. *Fuente:* Marchese, A., Navarro, D., Ingrassia, R., Gaibazzi, M., 2010.

En la siguiente ilustración se muestra la integración de los cubos en un "Data Warehouse" que sirva de base para realizar la Minería de Datos (Marchese y otros, 2009, p. 12):

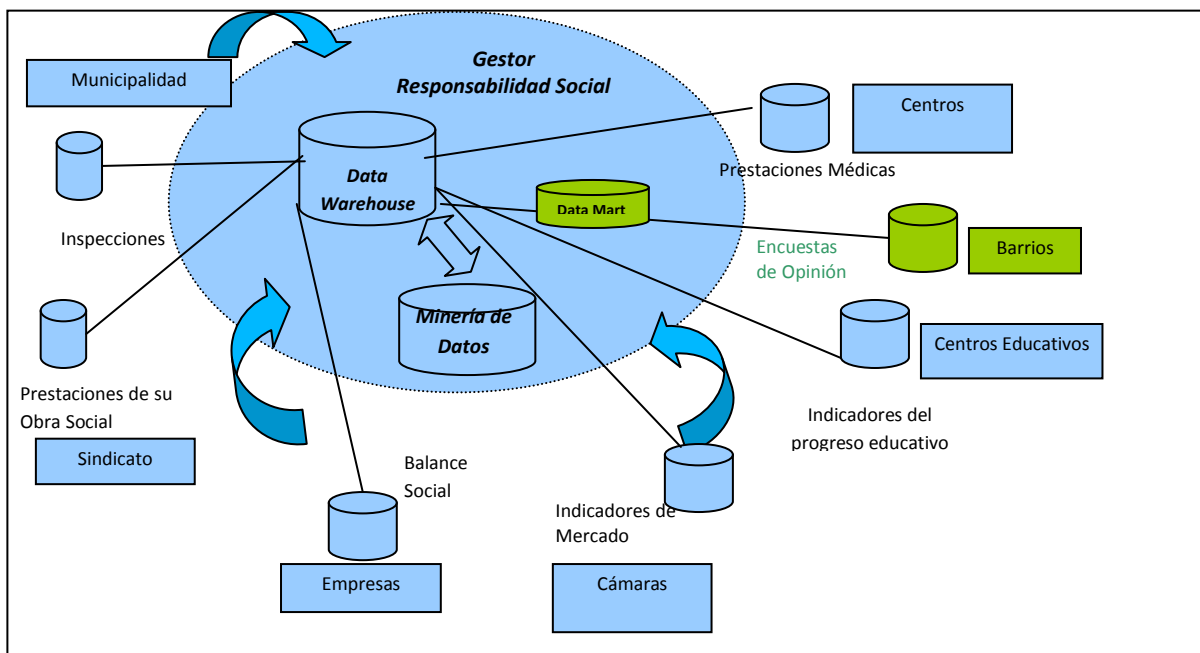


Ilustración 5: Integración Data Marts – Datawarehouse como base para la Minería de Datos. Se presenta el esquema de integración de las distintas bases de datos. *Fuente:* Marchese, A., Gaibazzi, M., Pérez, N., Ingrassia, R. 2009, p.12.



Dado lo disímil de cada módulo relevado y/a relevar, el análisis total del diseño del Data Warehouse llevaría mucho tiempo, lo que implica imposibilidad para ver resultados. Por este motivo adoptamos el criterio de construir Data-Marts para módulos independientes. Es decir, se selecciona la estrategia **Bottom up** en contraposición a la **Top-down**, para obviar los inconvenientes propias de esta última y obtener resultados más rápido y poder aprender y volcarlo en el diseño de los Data Marts restantes y en consecuencia, en el Data Warehouse.

En función del modelado de la *Base de Datos Multidimensionales* (cubos), de las tres maneras de modelado de los datos multidimensionales: Esquema Estrella (Star Scheme), Esquema Copo de nieve (Snowflake Scheme) y Esquema Constelación (Starflake Scheme); se eligió el Esquema Estrella dado que:

- I. Las dimensiones y jerarquías son pocas.
- II. Simplifica la visión y uso por parte del usuario de los datos. Esto es importante ya que el modelo se dejará a los implicados (Ilustración 9: "Uniones y Jerarquías – Situación socio-laboral de los Jefes de Hogar").
- III. Está orientado hacia el usuario, por lo tanto facilita la interacción con herramientas de análisis.
- IV. Su diseño es fácilmente modificable, lo que permite adecuarlo a medida que incorporamos experiencia en la implementación.

A continuación, se describe el procedimiento para el diseño lógico del Data Mart de acuerdo a las etapas identificadas en la metodología Hefestos.

1) Análisis de Requerimientos.

- a. Identificar preguntas: Se tuvo en cuenta los aspectos que se querían conocer de los Jefes de Hogar para luego integrarlo con los indicadores de las GRI (Global Reporting Initiative).
- b. Identificación de Indicadores y Perspectivas del análisis: Cantidad de Jefes de Hogar discriminados por Barrio, Sexo, Edad, Ocupación, Nivel de Estudios: al Inicio de su actividad laboral y Actual

Para clarificar los puntos anteriores se construye la siguiente tabla:

**Tabla 2:** Indicadores y Perspectivas – Situación Socio-Laboral de los Jefes de Hogar

N° Indicador y descripción	Perspectiva	Perspectiva	Perspectiva
1- Cantidad Jefes de Hogar (J.H.) discriminados: Edad y Sexo Ocupación	Barrio	Jefes de Hogar, discriminados: Edad, Sexo y Ocupación	
2- Cantidad de J.H. por Nivel de Estudios iniciar su activ. laboral	Barrio	Jefes de Hogar, discriminados: Edad, Sexo y Ocupación	Rama de Actividad de la organización en que trabaja
3- Cantidad de J.H. por Nivel de Estudios Actual	Barrio	Jefes de Hogar, discriminados: Edad, Sexo y Ocupación	Rama de Actividad de la organización en que trabaja
4-Comparativo Nivel de Estudios al Inicio vs Actual	Barrio	Jefes de Hogar, discriminados: Edad, Sexo y Ocupación	Rama de Actividad de la organización en que trabaja
5-Discriminación del Medio de Traslado al lugar de Trabajo	Barrio	Jefes de Hogar, discriminados: Edad, Sexo, Residencia en el Barrio del Lugar de trabajo	Rama actividad de la organización en que trabajan
6- Cantidad de Cuadras que camina hasta la Parada de Colectivo	Barrio	Jefe de Hogar Medio de traslado Colectivo	
7- Total Empleados discriminados: Mujeres Varones	Barrio	Rama actividad de la organización en que trabajan	
8- Total Empleados discriminados: Mujeres Varones	Barrio	Rama actividad de la organización en que trabajan	Jefe de Hogar discriminado por Sexo y Edad
9- Total Empleados que recibieron capacitación en: AL, SSL, DDHH y Otros	Barrio	Rama actividad de la organización en que trabajan	
10- Cantidad de hs. anuales de capacitación recibidas	Barrio	Jefes de Hogar que trabajan	Rama de actividad
11- Cantidad de J.H. que trabajan por Rama de Actividad	Barrio	Rama actividad de la organización en que trabajan los Jefes de Hogar	

Fuente: Elaboración propia



- c. Modelo Conceptual: Se describe gráficamente y en un alto nivel los Resultados a obtener (Indicadores), las Variables que se utilizarán (Perspectivas) y la Relación que hay entre ellos (Situación Socio-Laboral).

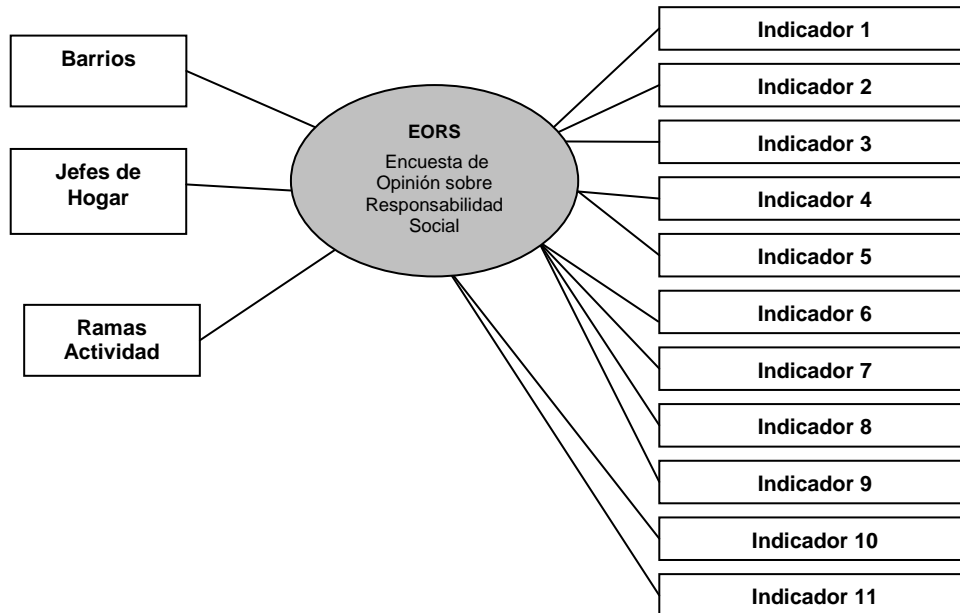


Ilustración 6: “Modelo Conceptual - Descripción Socio-Laboral de los Jefes de Hogar”. Se visualizan los resultados a obtener (Indicadores), las Variables que se utilizarán (Perspectivas) y la Relación que hay entre ellos (Situación Socio-Laboral). *Fuente:* Elaboración propia en base a la metodología Hefesto.

2) Análisis de los OLTP

- a. Establecer correspondencia con los requerimientos: como se vio al describir al Data Warehouse, los datos que lo alimentan pueden provenir de distintas fuentes. En este caso, no proviene de un OLTP (Proceso de Transacciones en Línea) sino de una Planilla de Cálculo.
- b. Seleccionar los criterios que integrarán cada perspectiva. Nivel de granularidad: Se entiende por granularidad, el nivel de detalle que tendrán las perspectivas: Barrios, Jefes de Hogar y Ramas de Actividad.

Barrios: Id_Barrio, Nom_Barrio

Jefes_Hogar (JH): Id_JH, Edad_JH, Rango_Edad_JH, Sexo_JH, Educación_Inicio_JH, Educación_Actual_JH, Ocupación_JH, Situación_Laboral_JH, Sindicato_JH, Categoría_JH, Antigüedad_JH, Lugar_Trab_Barrio_JH, Forma_Traslado_JH, Cuadras_Camina_JH

Ramas_Actividad: Id_Rama, Total_Empleados, Tot_Mujeres_Trabajo, Tot_Hombres, Capacit_AL, Capacit_SSL, Capacit_DDHH, Capacit_Otros, Tot_Hs_Año_Capac, Frec_Acc_Laborales, Frec_Ausentismo, Frec_Enferm_Lab, Frec_Discriminac, Casos_Trab_Infant, Invers_Donac, Detalle_Donac.



3) Elaboración del Modelo Lógico de la Estructura del Data Warehouse.

a. Diseñar tablas de dimensiones.

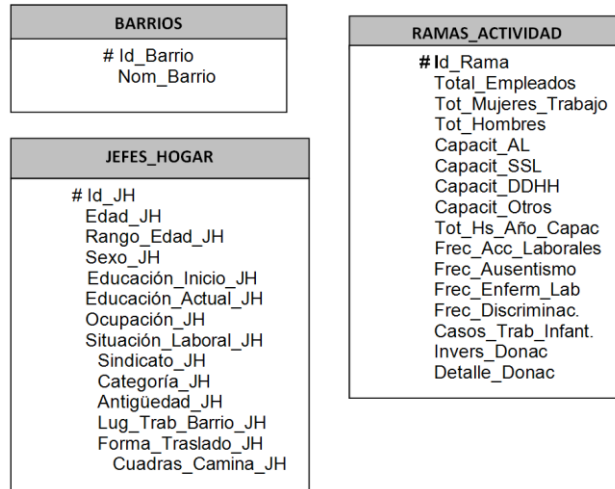


Ilustración 7: “Dimensiones – Situación Socio-Laboral de los Jefes de Hogar”. Se describe la estructura de datos de las tablas dimensiones. Fuente: Elaboración propia.

b. Diseñar tablas de hechos: La tabla de hechos contiene las claves externas y los indicadores:

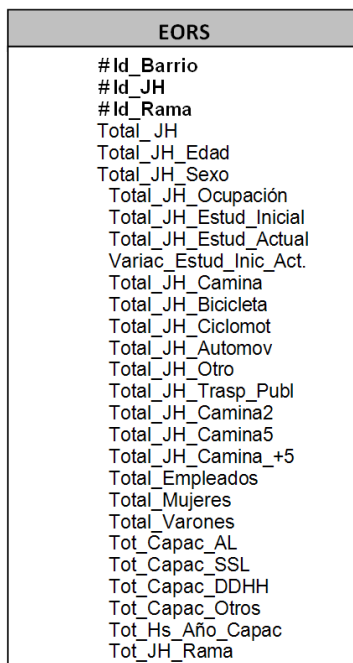


Ilustración 8: “Diseño de la tabla de hechos – Situación Socio-Laboral de los Jefes de Hogar”. Se describe la estructura de datos de la tabla de hechos. Fuente: Elaboración propia.

- c. Realizar uniones, y,
- d. Determinar jerarquías,

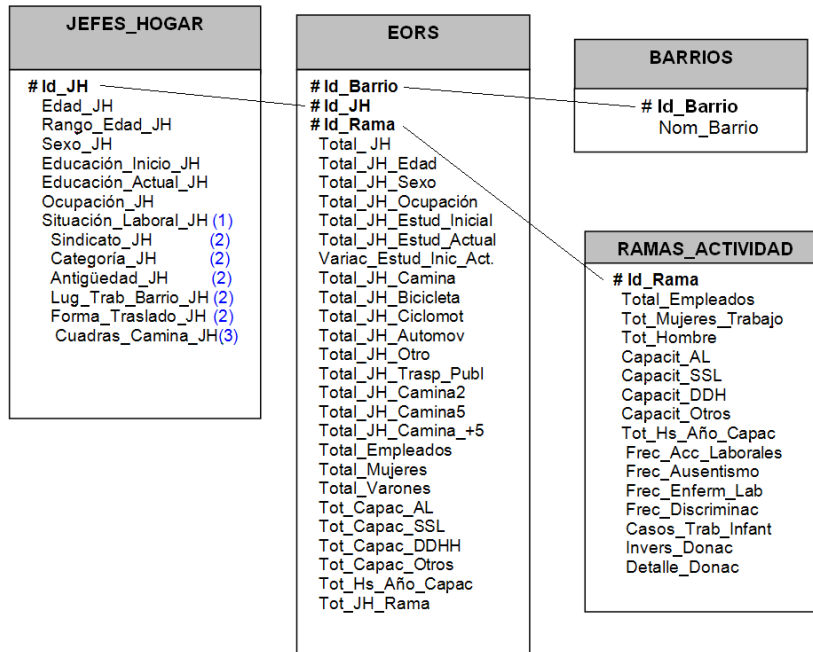


Ilustración 9: “Uniones y Jerarquías – Situación Socio-Laboral de los Jefes de Hogar”. Se exponen la estructura de datos de las tablas de hecho y dimensiones, y, los claves que vinculan cada tabla. *Fuente:* Elaboración propia.

4) Procesos ETL, Limpieza de Datos y Sentencias SQL.

Terminado el Modelo deberán cargarse los datos. A este proceso se lo conoce con el nombre ETL (Extracción, Transformación y Carga):

Extracción: determinar los datos que se extraerán de las distintas fuentes. Cada una de ellas requerirá su análisis particular.

Los datos que se extraen se colocan en un archivo intermedio, con un formato especial para ser utilizado en el proceso siguiente. Los registros rechazados deben guardarse en un archivo o log de rechazos para luego analizarlos y poder cargarlos al data warehouse (Ej. datos duplicados, datos vacíos, mal codificados, etc.).

Transformación: como su nombre indica consiste en transformar los datos extraídos al formato que requiere el data warehouse. Esto puede consistir en separar un campo fecha en día, mes y año, combinar campos en un único código, unificar codificaciones, tratamiento a dar a datos faltantes, etc.

Carga: consiste en tomar los datos extraídos y transformados y volcarlos al data warehouse.

En el caso planteado de la EORS, existen programas utilitarios para estas tareas que no exigen preparar programas específicos.

5. Conclusiones y Discusiones.

Se logró consensuar en la Estructura y Modelo del Data Mart. Nos encontramos con un modelo distinto al que estábamos acostumbrados dado que no teníamos la dimensión tiempo.



Queda ahora la ardua tarea de lograr ejecutar las etapas del ETL, en especial la etapa de Transformación, para cada dato faltantes o inconsistente hay que estipular una regla para su tratamiento de manera que pueda cargarse en el data mart.

En cuanto a la obtención de los Indicadores a través de distintos análisis en línea es el siguiente desafío, porque abrirá las puertas para nuevos indicadores y su interrelación con los indicadores de los restantes data marts.

Concluyendo, estamos seguros que entramos a una nueva etapa del proyecto donde se abren muchísimas posibilidades de integración y comparación como paso previo de la búsqueda de esa información oculta que trataremos obtener a través de la minería de datos.



REFERENCIA BIBLIOGRAFICA

- Bernabeu, R., (2007). *Data Warehousing: Investigación y Sistematización de Conceptos – HEFESTO: Metodología Propia para la Construcción de un Data Warehouse*. Córdoba: Autor.
- Bertello, E., (2003). *Seminario de Business Intelligence*. Buenos Aires: Centro de Capacitación y Formación Gerencial.
- Connolly, T., Begg, C., (2005). *Sistemas de base de datos: un enfoque práctico para diseño, implementación y gestión*. Buenos Aires: Pearson Education.
- Datawarehouse, (2003). Rosario: Crystal Solutions.
- Dyché, J. (2001). *E-data: Transformando datos en información con Data Warehousing*. Buenos Aires: Pearson Education.
- Hax, A., Majluf, N. (1997/2012). *Estrategias para el liderazgo competitivo: de la visión a los resultados*. (2ª reimpr.). Buenos Aires: Granica.
- La nueva organización de la información: Datawarehousing, (1996). Buenos Aires: Institute for International Reserch S.A.
- Laudon, K., Laudon, J. (2002). Mejoramiento de la toma de decisiones gerenciales. En su: Sistemas de información gerencial. Organización y tecnología de la empresa conectada en red (pp. 466-495). México: Pearson Educación.
- Lazarsfeld, P., (1973). Nacimiento y desarrollo de las variables. En: F. Korn. *Conceptos y variables en la investigación social* (pp. 23-47). Buenos Aires: Ediciones Nueva Visión.
- Marchese, A., Ingrassia, R., (Noviembre, 2010). *Sistemas de Conocimiento en Organismos Públicos*. Trabajo presentado en las 15ª Jornadas "Investigaciones en la Facultad" de Ciencias Económicas y Estadísticas, Rosario. Recuperado de: <http://www.fcecon.unr.edu.ar/web-nueva/?q=investigacion/actas-de-las-jornadas-anales>.
- Marchese, A., Campanaro, R., Ingrassia, R., Miller, E., Díaz Toledo, S., Pérez, N., (Noviembre, 2011). *Tecnologías de conocimiento en responsabilidad social empresarial. Caso San Lorenzo*. Trabajo presentado en las 16ª Jornadas "Investigaciones en la Facultad" de Ciencias Económicas y Estadísticas, Rosario. Recuperado de: <http://www.fcecon.unr.edu.ar/web-nueva/?q=investigacion/actas-de-las-jornadas-anales>.
- Marchese, A., Gaibazzi, M., Pérez, N., Ingrassia, R., (Noviembre, 2009). *Sistemas de Conocimiento en Organismos Públicos*. Trabajo presentado en las 14ª Jornadas "Investigaciones en la Facultad" de Ciencias Económicas y Estadísticas, Rosario. Recuperado de: <http://www.fcecon.unr.edu.ar/web-nueva/?q=investigacion/actas-de-las-jornadas-anales>.
- Marchese, A., Navarro, D., Ingrassia, R., Gaibazzi, M. (Agosto, 2010). *Indicadores para la Gestión del Conocimiento. Experiencias Convergentes*. Trabajo presentado en las 39ª Jornadas Argentinas de Informática e Investigación Operativa, CABA.
- Olcese, A., Rodríguez, M. y Alfaro, J. (2008). Gobierno y dirección de la empresa responsable y sostenible. En su: *Manual de la Empresa Responsable y Sostenible. Conceptos y herramientas de la Responsabilidad Social Corporativa o de la Empresa* (pp. 67-150). Madrid: McGraw-Hill Interamericana de España.
- Scribano, A. (2008). *El proceso de investigación social cualitativa*. Buenos Aires: Prometeo.