

UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

**PROCEDIMIENTO PARA EL DESARROLLO DE UN
SISTEMA DE INTELIGENCIA DE NEGOCIOS EN LOS
ENSAYOS CLÍNICOS QUE SE GESTIONAN EN EL CENTRO
DE INMUNOLOGÍA MOLECULAR**

TESIS PRESENTADA EN OPCIÓN AL TÍTULO DE MÁSTER EN INFORMÁTICA
APLICADA

Autora: Ing. Martha Denia Hernández Ramírez

Tutores: MSc. Yaikiel Hernández Díaz

MSc. Maypher Román Durán

Ciudad de La Habana, diciembre de 2010

DECLARACIÓN JURADA DE AUTORÍA Y AGRADECIMIENTOS

Yo Martha Denia Hernández Ramírez, con carné de identidad 85111621712, declaro que soy la autora principal del resultado que expongo en la presente tesis titulada "Procedimiento para el desarrollo de un sistema de inteligencia de negocios para los ensayos clínicos que se gestionan en el Centro de Inmunología Molecular", para optar por el título de Máster en Informática Aplicada.

Este trabajo fue desarrollado durante el período 2009 - 2010 en colaboración con mis compañeros de equipo: Ing. Anthony Rafael Sotolongo, Themis P. Díaz Morales, José S. Bermúdez Rodríguez, Javier Rodríguez Sotolongo, Yohan O. Peralta, Yailín Simón Mir y Yoander Iñiguez Bermúdez; quienes me reconocen la autoría principal del resultado expuesto en esta investigación.

A todos mis amigos les estoy muy agradecida. En especial deseo agradecer a mis tutores por ayudarme en mi formación como máster. Además, deseo agradecer a todas las personas que de una forma u otra contribuyeron a mi crecimiento profesional y humano en general. A todos ellos, así como a otros colegas y amigos que no he mencionado por razones de espacio, les doy las más sinceras gracias.

Finalmente declaro que todo lo anteriormente expuesto se ajusta a la verdad, y asumo la responsabilidad moral y jurídica que se derive de este juramento profesional.

Y para que así conste, firmo la presente declaración jurada de autoría en Ciudad de La Habana a los ____ días del mes de _____ del año 2010.

Ing. Martha Denia Hernández Ramírez

RESUMEN

La presente investigación surge como parte de la colaboración que existe entre la Universidad de las Ciencias Informáticas (UCI) y el Centro de Inmunología Molecular (CIM). Uno de los principales objetivos de este centro es gestionar la información que se obtiene a partir de la conducción de los Ensayos Clínicos (EC) una vez que se aplica un producto determinado a pacientes que padecen de cáncer; por lo que existe la necesidad de almacenar y analizar dicha información. En la presente investigación se realiza una propuesta de procedimiento para conducir el desarrollo de soluciones de inteligencia de negocio en el CIM. La implementación de este procedimiento permitirá almacenar toda la información que se gestiona, de manera íntegra y estándar; logrando viabilizar los análisis estadísticos que se necesitan realizar por parte de los especialistas de la institución. El procedimiento es evaluado a partir del método de expertos Delphi, obteniéndose el resultado de "Muy adecuado". Se cuenta además de un aval del centro cliente, donde se valora de satisfactorio el trabajo realizado.

PALABRAS CLAVES:

Business Intelligence (BI), Centro de Inmunología Molecular, datawarehouse (DWH), diseño de un DWH, ensayos clínicos, ETL, inteligencia de negocio, procedimiento.

Introducción	1
Capítulo 1: Fundamentación Teórica	6
1.1. Inteligencia de negocios	6
1.1.1. Fases y etapas de desarrollo BI	7
1.2. Gestión de los ensayos clínicos del CIM	9
1.3. Metodologías para el desarrollo BI	13
1.3.1. Metodologías de arquitectura	13
1.3.2. Metodologías de desarrollo	14
1.4. Lista de chequeo	16
1.5. Método para evaluar el procedimiento BI para los EC del CIM	16
1.5.1. El método Delphi	16
1.6. Conclusiones parciales	17
Capítulo 2: Desarrollo del procedimiento BI para los EC del CIM	18
2.1. Etapa de diseño del DWH	18
2.1.1. Análisis del negocio	19
2.1.2. Especificación de las necesidades	21
2.1.3. Desarrollo del modelo conceptual	24
2.1.4. Desarrollo del modelo lógico	28
2.1.5. Evaluación del diseño con lista de chequeo	31
2.2. Proceso ETL	35
2.2.1. Análisis de las fuentes de datos	36
2.2.2. Diseño de la arquitectura	36
2.2.3. Selección de las herramientas informáticas para implementar el proceso de ETL	37
2.2.4. Desarrollo del modelo físico	38
2.2.5. Extracción, transformación y carga	39
2.2.6. Definición del período de actualización del datawarehouse	41
2.2.7. Evaluación del proceso ETL con lista de chequeo	41
2.3. Análisis de datos, con técnicas OLAP	43
2.3.1. Definición de las restricciones del diseño	43
2.3.2. Selección de las herramientas informáticas para implementar análisis OLAP	45
2.3.3. Desarrollo de los cubos OLAP	46
2.3.4. Diseño de las consultas MDX	46
2.3.5. Construcción de una interfaz de usuario	47
2.3.6. Evaluación del análisis con lista de chequeo	48
2.4. Conclusiones parciales	50
Capítulo 3: Evaluación del procedimiento BI para los EC del CIM	51
3.1. Aplicación del método Delphi	51
3.1.1. Elección de los expertos	51
3.1.2. Elaboración del cuestionario para la validación de la propuesta	54

3.1.3. Establecimiento de la concordancia de los expertos	55
3.1.4. Desarrollo práctico y explotación de los resultados	55
3.2. Conclusiones parciales	58
Conclusiones	59
Recomendaciones	60
Referencias Bibliográficas	61
Bibliografía	65
Anexos.....	67
Glosario de Términos	72

Introducción

Las organizaciones modernas disponen cada vez de más datos sobre sus negocios. De estos datos se puede obtener información relevante para mejorar el desempeño e innovar en sus procesos de manera que se transforme en una organización competitiva y exitosa. Esto se puede lograr con el uso de la inteligencia de negocios (en inglés se conoce como *Business Intelligence* o BI), término que en 1989 Howard Dresner definió como “... un conjunto de conceptos y métodos para mejorar el proceso de decisión utilizando un sistema de soporte basado en hechos...” (1). Mediante el uso de BI se logra una unión entre el mundo de los datos y el de los negocios, con una solución basada en almacenes de datos o Datawarehouse (DWH por sus siglas en inglés) (2).

Existen muchas empresas u organizaciones que implementan DWH para lograr el éxito en sus negocios de mercado (3). Ejemplo en Venezuela, en el área de la seguridad ciudadana (4); en Perú, en hospitales para la sectorización de pacientes en el consumo de medicamentos (5). También en Cuba para la gestión de inventario (6) y recursos humanos (7), y en la Universidad de las Ciencias Informáticas para la toma de decisiones en cuanto al consumo energético (8) (9). Todo esto presupone que el uso de esta tecnología es de gran utilidad para apoyar la toma de decisiones en diferentes tipos de organizaciones. En el sector de la Salud Pública Cubana, específicamente en los centros investigativos, implementar una solución de este tipo es conveniente debido al gran número de información que se gestiona. De ahí que el desarrollo de una solución BI en el Centro de Inmunología Molecular (CIM) puede apoyar en la toma de decisiones de los investigadores clínicos.

El CIM es un centro destinado a la creación de biofármacos para el tratamiento del cáncer y otras enfermedades crónicas no transmisibles. Para la aprobación, prueba e introducción de dichos productos en el mercado se realizan Ensayos Clínicos (EC); que permiten evaluar nuevos fármacos o tratamientos médicos a través de su aplicación a seres humanos (10). Actualmente el CIM se encuentra trabajando con 28 productos diferentes que abarcan 62 EC; donde se encuentran involucrados más de 500 hospitales de todo el país (11).

La gestión y el diseño de los Cuadernos de Recogida de Datos (CRD) es realizado por diferentes especialistas mediante el sistema EpiData; generándose un gran volumen de documentación que puede encontrarse en diferentes formatos. Esto provoca que la información se encuentre

descentralizada y de manera diferente para cada ensayo; dificultando la realización de análisis estadísticos, ya que solo es posible realizarlos por EC y no sobre los resultados de varios EC. El uso apropiado de las herramientas de BI puede contribuir con la realización de mejores análisis estadísticos, facilitando una adecuada toma de decisiones a los especialistas.

Para implementar una solución BI se deben desarrollar las tres etapas que esta comprende: el diseño del DWH, el proceso ETL y el análisis de datos; por lo que se han propuesto metodologías de desarrollo para una o varias etapas. Entre las más modernas se destacan Hefesto, la propuesta de Trujillo y el ciclo de vida Kimball. Optar por una de estas metodologías para que los especialistas del CIM guíen este proceso, no resulta conveniente debido a que cada una de ellas presenta problemas que no facilitan el desarrollo de soluciones de este tipo en dicha institución:

- Hefesto: no contempla componentes importantes de la arquitectura del DWH como son el área temporal y los metadatos (12).
- Trujillo: no detalla la manera en que se deben diseñar los modelos de datos ni la forma de obtener las variables para lograr la correspondencia con los datos fuentes. Además no es contemplada como un ciclo completo de desarrollo de BI, sino que abarca solo la etapa de diseño del DWH (13).
- Kimball: es muy amplia la manera de abordar los elementos para las etapas de desarrollo de BI, dejando claro el qué se debe hacer pero no el cómo lograrlo; provocando demoras en los resultados. Además que no contempla una forma de evaluación del proceso (14).

Por lo que se plantea como **problema científico**: *la necesidad de ajustar las metodologías de inteligencia de negocios para desarrollar soluciones BI, que contribuyan al almacenamiento y análisis de los ensayos clínicos en el Centro de Inmunología Molecular, por parte de los especialistas de la institución.*

La investigación tiene como **objeto de estudio** *el proceso de inteligencia de negocios, enmarcado en el campo de acción el empleo de la inteligencia de negocios en los ensayos clínicos que se gestionan en el Centro de Inmunología Molecular.*

La **hipótesis científica** que se plantea es:

- ✓ *La elaboración de un procedimiento para desarrollar soluciones BI en los ensayos clínicos que se gestionan en el Centro de Inmunología Molecular, a partir de un ajuste de metodologías de inteligencia de negocio, facilitará la implementación de soluciones BI en el centro.*

El **objetivo general** de la investigación es:

Desarrollar un procedimiento que facilite la implementación del proceso de inteligencia de negocios en los ensayos clínicos, que se gestionan en el Centro de Inmunología Molecular; a partir de lo mejor y más aplicable de las metodologías de desarrollo BI que existen.

Este objetivo se desglosa en los siguientes **objetivos específicos**:

- ✓ *Definir un procedimiento por etapas de desarrollo, para facilitar el proceso de inteligencia de negocios de los ensayos clínicos que se gestionan en el Centro de Inmunología Molecular; a partir de lo mejor y más aplicable de las metodologías de desarrollo BI.*
- ✓ *Aplicar el procedimiento de inteligencia de negocios desarrollado, a los datos de los ensayos clínicos del producto Nimotuzumab (hR3), que se gestiona en el Centro de Inmunología Molecular.*
- ✓ *Evaluar el procedimiento propuesto utilizando criterio de expertos.*

Para dar cumplimiento a los objetivos específicos se trazaron las siguientes **tareas investigativas**:

- ✓ *Revisión bibliográfica sobre el término inteligencia de negocios y los conceptos relacionados; así como del proceso de gestión de EC en el CIM:*
- ✓ *Estudio, análisis y selección de una de las metodologías de arquitectura existentes para el desarrollo de un DWH.*
- ✓ *Estudio, análisis y selección de lo mejor y más aplicable de las metodologías de desarrollo de inteligencia de negocios que existen.*
- ✓ *Elaboración de un procedimiento para facilitar la etapa de diseño de un datawarehouse para los ensayos clínicos que se gestionan en el Centro de Inmunología Molecular.*
- ✓ *Elaboración de un procedimiento para facilitar el proceso de extracción, transformación y carga de un datawarehouse para los ensayos clínicos del Centro de Inmunología Molecular.*

- ✓ *Elaboración de un procedimiento para facilitar el proceso de análisis de los datos recopilados en un datawarehouse de los ensayos clínicos del Centro de Inmunología Molecular.*
- ✓ *Integración de los procedimientos definidos para cada una de las etapas comprendidas en el proceso de inteligencia de negocios.*
- ✓ *Aplicación del procedimiento general para la implementación y análisis de un datawarehouse a partir de los datos de los ensayos clínicos del producto hR3 que se gestiona en el Centro de Inmunología Molecular.*
- ✓ *Estudio del método Delphi para evaluar los procedimientos propuesto, mediante el criterio de expertos.*

Para realizar la presente investigación se tuvo en cuenta los métodos siguientes:

Métodos teóricos:

Método Analítico - Sintético: durante todo el proceso investigativo se hace un estudio de la bibliografía relacionada con el tema y a partir del análisis realizado se selecciona una síntesis de lo estudiado. Se utiliza además para el procesamiento de la información y arribar a las conclusiones de la investigación.

Método Histórico - Lógico: para la realización de la investigación se hace necesario estudiar la evolución del problema y la existencia de metodologías y procedimientos similares al que se pretende elaborar; determinando cuáles son las tendencias actuales para el desarrollo de soluciones BI.

Métodos empíricos:

Método Delphi: consiste en la utilización sistemática del juicio intuitivo de un grupo de expertos para obtener un consenso de opiniones informadas. Se utiliza durante la investigación para evaluar la propuesta de procedimiento.

Técnicas de recopilación de información:

Entrevista: se utiliza para obtener información en forma verbal, a través de preguntas que se proponen para conocer en detalle sobre el tema que se investiga.

Encuesta: durante la realización de la investigación fue necesario realizar encuestas en formas de cuestionarios para la validación de la solución propuesta.

Novedad y aporte teórico:

- ✓ *Un procedimiento para facilitar el proceso de inteligencia de negocios en la gestión de los ensayos clínicos del Centro de Inmunología Molecular, a partir de lo mejor y más aplicable de las metodologías de desarrollo BI que existen.*

Novedad y aporte práctico:

- ✓ *La implementación de un datawarehouse para los ensayos clínicos del producto hR3, el cual contará con información histórica, íntegra y estándar de todos los ensayos conducidos.*
- ✓ *Análisis estadísticos sobre eficacia y seguridad del producto hR3 desarrollado en el CIM, para el tratamiento del cáncer en diferentes localizaciones.*

El trabajo que a continuación se presenta se estructura en tres capítulos fundamentales:

El **Capítulo 1** abarca temas relacionados con el concepto de inteligencia de negocios, sus áreas de aplicación y el datawarehousing como proceso que lo apoya; describiéndose las etapas y fases que comprende. Luego se aborda sobre la gestión de los ensayos clínicos en el CIM y las principales características que lo distinguen; punto de partida que se utiliza para el análisis y selección de una de las metodologías de arquitectura que existen para soluciones BI; así como para el estudio de las metodologías de desarrollo. Además se fundamenta sobre el método utilizado durante la investigación como técnica de evaluación.

El **Capítulo 2** describe el procedimiento propuesto para desarrollar soluciones BI en los EC que se gestionan en el CIM. Se describen las actividades que se realizan y los entregables que se generan. Se describen los principales resultados de la aplicación del procedimiento a través del desarrollo de una solución BI con análisis OLAP de los datos de EC del producto hR3 que se gestiona en el CIM; donde se obtienen reportes dimensionales que evalúan la eficacia y seguridad de este producto.

En el **Capítulo 3** se aplica el método seleccionado para la evaluación del procedimiento BI para los EC del CIM, a través de la técnica Delphi mediante criterio de expertos; detallándose los resultados arrojados y los cuestionarios utilizados para obtener los criterios de los encuestados.

Capítulo 1: Fundamentación Teórica

En este capítulo se abarcan temas relacionados con el concepto de inteligencia de negocios, sus áreas de aplicación y el datawarehousing como proceso que lo apoya; describiéndose las etapas y fases que comprende. Luego se aborda sobre la gestión de los ensayos clínicos en el CIM y las principales características que lo distinguen; punto de partida que se utiliza para el análisis y selección de una de las metodologías de arquitectura que existen para soluciones BI; así como para el estudio de las metodologías de desarrollo. Además se fundamenta sobre el método utilizado durante la investigación como técnica de evaluación.

1.1. Inteligencia de negocios

La inteligencia de negocios permite que el proceso de toma de decisiones esté fundamentado sobre un amplio conocimiento de sí mismo y del entorno, minimizando de esta manera el riesgo y la incertidumbre. Además, propicia que las organizaciones puedan traducir sus objetivos en indicadores de estudio y que estos puedan ser analizados desde diferentes perspectivas (12, p.5) (15) (16). Se puede describir BI como un proceso dividido en fases, que se unen etapas, que integran el almacenamiento y el procesamiento de grandes cantidades de datos (17) (18) (19). El principal objetivo de este proceso es transformarlos en conocimiento y decisiones, en tiempo real, a través del análisis y la exploración (20).

Entre los beneficios más importantes que una solución BI proporciona a las organizaciones se encuentran: (15, p. 32-37) (21) (22)

- ✓ Reducir el tiempo que se requiere para recoger toda la información relevante de un tema en particular; partiendo del hecho de que la misma se encontrará integrada en una única fuente de fácil acceso.
- ✓ Automatizar la asimilación de la información, debido a que la extracción y carga de los datos necesarios se realizará a través de procesos ya definidos.
- ✓ Proporcionar herramientas de análisis para establecer comparaciones y tomar decisiones.
- ✓ Permitir a los usuarios no depender de reportes o informes programados, porque los mismos serán generados de manera dinámica.
- ✓ Permitir el acceso y análisis directo de los indicadores de éxito.

- ✓ Permitir al usuario consultar y analizar los datos de manera sencilla e intuitiva.

1.1.1. Fases y etapas de desarrollo BI

Según la metodología Hefesto (12, p. 6-7) la inteligencia de negocios se divide en fases las cuales se explican a continuación (ver figura 1):

- ✓ Fase 1 - *Dirigir y planear*: recolectar los requerimientos de información de los diferentes usuarios; así como entender sus diversas necesidades. Generando las preguntas que les ayudarán a alcanzar sus objetivos.
- ✓ Fase 2 - *Recolección de información*: realizar el proceso de extraer desde las diferentes fuentes de información de la empresa.
- ✓ Fase 3 - *Procesamiento de datos*: integrar y cargar los datos desde las fuentes de origen en un formato utilizable para el análisis.
- ✓ Fase 4 - *Análisis y producción*: trabajar sobre los datos extraídos e integrados, utilizando herramientas y técnicas propias de la tecnología BI, para crear inteligencia.
- ✓ Fase 5 - *Difusión*: entregar a los usuarios que lo requieran, las herramientas necesarias que les permitan explorar los datos de manera sencilla e intuitiva.

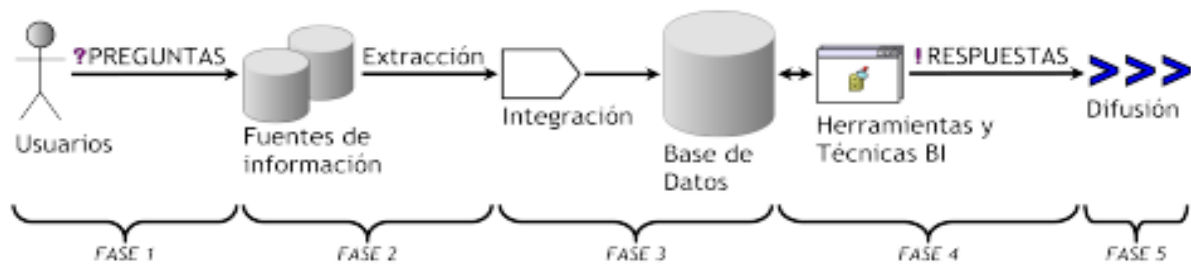


Figura 1 Fases del proceso BI

Como se muestra en la figura 1 para implementar soluciones BI es necesario gestionar datos que se encuentran guardados en diversos formatos, luego depurarlos e integrarlos y finalmente almacenarlos en un solo destino que permita su exploración. Esto se logra con la implementación del proceso Datawarehousing; donde se realiza la extracción, transformación, integración y centralización de los datos de una organización (15, p. 46-47) (23).

Para cumplir con los objetivos del proceso Datawarehousing es necesario que la información sea almacenada en una base de datos, con estructura multidimensional, denominada Datawarehouse (DWH o almacenes de datos). De ahí que las cinco fases presentadas anteriormente suelen unirse en muchas bibliografías en tres etapas que organizan todo el proceso de una solución BI (15, p.49), (24): Diseño del DWH, Proceso de extracción, transformación y carga de los datos (conocido por ETL de sus términos en inglés *Extraction, Transform, Load*) y finalmente el Análisis de datos.

- La *etapa de diseño* comprende el análisis del negocio en cuestión, para así poder identificar las necesidades o requerimientos del negocio. De esta forma definir un modelo conceptual que debe traducirse más tarde en el modelo lógico del DWH, basado en una tipología de esquema multidimensional (16, p.73) (18) (25). Existen tres tipos de esquemas las cuales son: esquemas en estrella, esquema en bola de nieve o copo de nieve y constelaciones de hechos. En el esquema en estrella, la base de datos relacional consiste en una tabla simple de hechos y una tabla para cada dimensión. Cada tupla de la tabla de hechos incluye las medidas consideradas y una referencia a cada dimensión. Por su parte, los esquemas en bola de nieve o copo de nieve son un refinamiento del esquema en estrella que soporta jerarquías manteniendo normalizadas las tablas. Los esquemas en estrella y bola de nieve pueden generalizarse con la inclusión de distintas tablas de hechos que comparten todas o algunas de las dimensiones, a estos se les conocen como constelaciones de hechos (12, p. 37-40).
- La *etapa de ETL* se encarga de extraer los datos desde las diversas fuentes que se requieran, transformarlos para resolver posibles problemas de inconsistencias entre los mismos y finalmente, después de haberlos depurado, cargarlos en el depósito de datos o almacén. Para esto se puede crear un área temporal o staging area y los metadatos. En el caso del staging area es un área temporal donde se recogen los datos que se necesitan de los sistemas origen. Se recogen los datos estrictamente necesarios para las cargas y se aplica el mínimo de transformaciones a los mismos. No se aplican restricciones de integridad ni se utilizan claves, los datos se tratan como si las tablas fueran ficheros planos; de esta manera se minimiza la afectación a los sistemas origen, haciendo la carga lo más rápida posible y se reduce también al mínimo la posibilidad de error. Por su parte los metadatos son datos que describen o dan información de otros datos, que en este caso, existen en la arquitectura del DWH. Brindan información de localización, estructura y significado de los datos, básicamente mapean los

mismos. El concepto de metadatos es análogo al uso de índices para localizar objetos en lugar de datos. La gran ventaja que trae aparejada el DWH en relación con los metadatos es que el usuario puede gestionarlos, exportarlos, importarlos, realizarles mantenimiento e interactuar con ellos, ya sea manual o automáticamente (16, p. 87) (26) (27). La etapa ETL tiene como precedente la etapa de diseño que servirá de punto de partida para su implementación.

- La *etapa de análisis* es donde se les da el valor real a los datos al extraer de ellos la información requerida para auxiliar la toma de decisiones y mostrarla de manera entendible a los usuarios finales. Existen diferentes técnicas de análisis como: los Reportes y consultas, la Minería de datos y el Procesamiento analítico en línea (conocido como OLAP por sus términos en inglés *Online Analytical Processing*). En el caso de la minería de datos es considerada como el proceso de descubrir conocimientos interesantes: patrones, asociaciones, cambios, anomalías y estructuras significativas, a partir de grandes cantidades de datos (28). Por su parte la técnica reportes y consultas se refiere al uso de herramientas destinadas a la producción de consultas y reportes. Estas les ofrecen a los usuarios la posibilidad de generar informes del área de interés del negocio que se esté analizando, a través de pantallas gráficas intuitivas. El usuario solo debe seleccionar las opciones que le brindan para especificar los elementos de datos, sus condiciones, criterios de agrupación y demás atributos que se consideren significativos (12). Y en el caso de OLAP es la tecnología que permite que la información sea vista multidimensionalmente, a través de cubos con categorías descriptivas y valores cuantitativos. El uso de cubos le da a los especialistas la ventaja de poder configurarlos, filtrarlos y reestructurarlos en tiempo real, actualizando inmediatamente el reporte. Además este modelo de datos concibe una única consulta MDX sobre el cubo OLAP de datos, previamente diseñado, permitiendo obtener reportes donde la sencillez y el nivel de detalles sean elevados. Otra ventaja que ofrece el modelo dimensional es poder cambiar de datos resumidos a datos detallados y filtrar los datos en subconjuntos significativos (28).

1.2. Gestión de los ensayos clínicos del CIM

Dependiendo de la finalidad y utilidad que se desee con la solución BI que se implementa, se escogerán las etapas a desarrollar. De ahí que sea importante analizar el negocio particular donde se desee implementar la solución, identificando las características que lo distinguen. Estas serán el punto de partida para el estudio y selección de las metodologías a utilizar, tanto para la arquitectura como para el

desarrollo de la solución; así como para la selección de la técnica de análisis. De ahí que a continuación se presente un estudio sobre el Centro de Inmunología Molecular ya que es el negocio de la presente investigación.

El Centro de Inmunología Molecular tiene como principal misión la búsqueda de nuevos productos para el diagnóstico y tratamiento del cáncer y enfermedades relacionadas con el sistema inmune. Las líneas básicas de su investigación están concentradas en la inmunoterapia del cáncer, especialmente en el desarrollo de vacunas moleculares, ingeniería de anticuerpos, ingeniería celular, bioinformática y regulación de la respuesta inmune (29).

Para llevar el control de los estudios clínicos se crean los CRD; donde se recoge toda la información relacionada con el paciente durante su tratamiento. Estos datos deben mantenerse almacenados durante 15 años posteriores a la realización del ensayo de forma segura, con el objetivo de cumplir con las buenas prácticas clínicas exigidas por las Agencias Regulatorias a nivel mundial. Los CRD se digitalizan y se guardan en bases de datos electrónicas para facilitar los resultados estadísticos que de ellos se derivan. Actualmente este proceso se realiza a través del sistema EpiData, el cual no presenta un diseño relacional de bases de datos, y del cual se generan reportes en diferentes formatos como: SPSS, Excel, SASS y Text. Para la gestión de los EC acceden a este programa diferentes especialistas, lo que provoca que la información esté disgregada en diferentes modelos y que estos no presenten igual estructura de diseño. De ahí que la información no se encuentra integrada ni estandarizada, dificultándose la realización de análisis estadísticos complejos entre diferentes EC. Actualmente se trabaja en el desarrollo de alasClínicas, un sistema de gestión de EC en el CIM, el cual permite crear CRD dinámicos a partir de hojas de cálculos Excel, por lo que una vez desplegada esta solución en el CIM no se podrá visualizar el diseño de base de datos para los CRD, sino que se logrará obtener un Excel como reporte del CRD electrónico.

El programa de desarrollo clínico de un fármaco, incluye varios ensayos clínicos que permiten investigar la seguridad y eficacia de los productos. Para poder determinar la eficacia del producto se debe tener en cuenta la relación que existe entre la evaluación de la respuesta al medicamento y la supervivencia de los pacientes, con relación al tratamiento y al número de dosis recibida. En el caso de la seguridad del producto se deben analizar los eventos adversos más frecuente en relación con el tratamiento y número

de dosis recibida, así como la cantidad de pacientes que lo presentan, identificándose la intensidad y causalidad de los eventos adversos presentados.

Uno de los productos que actualmente desarrolla el CIM es el Nimotuzumab, anticuerpo monoclonal recombinante contra el receptor del factor de crecimiento epidérmico. Este es un medicamento que se utiliza como monoterapia o en combinación con radioterapia y/o quimioterapia para el tratamiento de cáncer de origen epitelial y gliomas. El Nimotuzumab recibe alternativamente las denominaciones comerciales de CIMAher®, TheraCIM®-hR3, Theraloc®, BIOMAb®-EGFR, YMB 1000 y VECTHIX®. Hasta la fecha el hR3 ha sido administrado a más de 1000 pacientes en EC terminados y en más de 5000 pacientes en estudios en curso. El programa de desarrollo clínico de este fármaco, incluye 33 ensayos clínicos adicionales que están en curso y 3 programas de acceso expandido, para investigar la seguridad y eficacia de Nimotuzumab en cáncer de mama, esófago, cáncer de cabeza y cuello, próstata, cáncer de hígado, páncreas, pulmón (células no pequeñas), cáncer de cuello de útero y glioma en pacientes pediátricos y adultos (30). A continuación se reportan algunos de los estudios clínicos finalizados para este producto, que será objeto de análisis en esta investigación. En el caso de los estudios 069 y 079 son dos ensayos en curso con 80 y 30 número de pacientes a incluir respectivamente (ver tabla 1).

Tabla 1 Ensayos clínicos del producto hR3

Número del estudio (País)	Indicación	Fase	Dosis de Nimotuzumab	Fármaco (s)	No de pacientes tratados/Dosis
IIC RD EC040 (Cuba)	Cáncer de Cabeza y Cuello	I/II	50, 100, 200 o 400 mg	nimotuzumab + RT	3/50 mg 4/100 mg 3/200 mg 4/400 mg
IIC RD EC046 (Cuba)	Cáncer de Cabeza y Cuello	I/II	200 o 400 mg	nimotuzumab + RT	5/200 mg 5/400 mg
IIC RD EC055 (Cuba)	Cáncer de Cabeza y Cuello	II/III	200 mg	nimotuzumab ± RT	112
IIC RD EC076 (Cuba)	Cáncer de Cabeza y Cuello	I	200 mg	nimotuzumab ± RT	10
IIC RD EC075 (Cuba)	Cáncer de Esófago avanzado	II	200 mg	nimotuzumab ±RT/QT	68
IIC RD EC053 (Cuba)	GBM o AA adultos	I/II	200 mg	nimotuzumab + RT	29

Número del estudio (País)	Indicación	Fase	Dosis de Nimotuzumab	Fármaco (s)	No de pacientes tratados/Dosis
IIC RD EC069 (Cuba)	Glioma – adultos	III	200 mg	nimotuzumab ± RT	31
IIC-RD-EC070 (Cuba)	Cáncer de Mama localmente avanzado	I/II	50 a 400 mg	nimotuzumab +QT	15
IIC RD EC079 (Cuba)	Metástasis cerebrales de cáncer de Pulmón células no pequeñas	II	200 mg	nimotuzumab ± RT	27
IIC RD EC035 (Cuba)	Tumores Sólidos Epiteliales	I	50, 100, 200 o 400 mg	nimotuzumab	3/50 mg 3/100 mg 3/200 mg 3/400 mg

A partir de lo anteriormente expuesto a continuación se resumen las características que se identificaron en el negocio de EC en el CIM:

- Para la gestión de datos de los ensayos clínicos se utiliza el sistema EpiData y se trabaja en la creación de alasClínicas, sistemas que no presentan un diseño de base de datos relacional para los CRD electrónicos.
- La información de los ensayos no está integrada ni estandarizada, por lo que existen dificultades para realizar análisis estadísticos entre diferentes ensayos clínicos de un mismo producto.
- Necesidad de almacenar durante 15 años los datos de un ensayo, logrando seguridad de la información.
- Se necesita conocer la seguridad y eficacia de los productos que se crean en el CIM para el tratamiento del cáncer, teniendo en cuenta las diferentes localizaciones de la enfermedad: mama, esófago, próstata, pulmón, glioma, entre otras.

Debido a la situación que actualmente presentan los especialistas del CIM para la realización de los análisis estadísticos de los productos que se desarrollan en el CIM y a partir de las ventajas que presenta la implementación de soluciones BI, se decide implementar las tres etapas que comprende el BI en el CIM. Se hará uso de la técnica OLAP para realizar el análisis del almacén de datos, porque esta permite analizar el negocio desde diferentes escenarios históricos, y proyectar cómo se ha venido comportando y evolucionando en un ambiente multidimensional, o sea, mediante la combinación de

diferentes perspectivas, temas de interés o dimensiones, que es lo que se necesita para el análisis de la seguridad y eficacia de los productos. Para guiar el proceso BI que se desea implementar es necesario analizar las propuestas metodológicas que existen, teniendo en cuenta las características del negocio para seleccionar la más adecuada.

1.3. Metodologías para el desarrollo BI

Teniendo en cuenta las características identificadas en el negocio de EC en el CIM se presenta a continuación un análisis de las metodologías de desarrollo y de arquitectura que existen para la implementación de una solución BI. Se estudian las opciones que se tienen en cuanto a las metodologías de arquitectura de un DWH (Kimball e Inmon) y las metodologías de desarrollo existentes para cada una de las etapas comprendidas en BI.

1.3.1. Metodologías de arquitectura

En el caso de las propuestas que se tienen para optar por una arquitectura de DWH se encuentran la de Ralph Kimball y Bill Inmon. Kimball es el principal promotor del enfoque dimensional para el diseño de almacenes de datos y propone una metodología ascendente (bottom-up); partiendo de la idea de que un DWH es la unión de todos los Data Marts (datos de un área de negocio específica, áreas departamentales) de una entidad. También se encuentra la visión arquitectónica propuesta por Bill Inmon, conocido por muchos como el padre del datawarehouse. Este propone una metodología descendente (top-down); partiendo de la premisa de que los Data Marts se deben crear después de haber terminado el DWH completo de la organización.

La propuesta de Kimball es la más aceptada en todo el mundo por ser efectiva para desarrollar una solución de construcción de DWH. Apunta a una solución completa que se puede implementar en poco tiempo, cumpliendo con lo que se necesita en una metodología de construcción de DWH: dar resultados rápidos para demostrar el valor de la solución al negocio y no perder la confianza de los involucrados. Sin embargo la propuesta de Inmon por el contrario, puede tener una implementación mucho más tardada, y es recomendada cuando se hace difícil representar el modelo a través de dimensiones y la complejidad de la solución se hace demasiado grande. Ambas metodologías pueden implementarse en un mismo DWH; aunque se recomienda utilizar la metodología Inmon en iteraciones ya avanzadas y siempre empezar con Kimball (24, p. 280-282) (31) (32).

De las dos metodologías de arquitectura analizadas, se decide utilizar Kimball por ser la más aceptada en todo y por ser la que más se ajusta a las condiciones del tratamiento de la información del CIM, donde se analizan los resultados de eficacia y seguridad de cada uno de los productos como áreas departamentales.

1.3.2. Metodologías de desarrollo

En el caso de las metodologías que existen para guiar el desarrollo de una solución de inteligencia de negocios se analizaron las propuestas más utilizadas: Hefesto, el ciclo de vida Kimball, la propuesta de Trujillo para el desarrollo de almacenes de datos dirigidos por modelos, Data Warehouse Engineering Process (DWEPE) y Rapid Warehousing Methodology (RWM). A continuación se describen de forma resumida cada una de estas metodologías analizadas:

- Hefesto es una metodología propia, cuya propuesta está fundamentada en una amplia investigación, comparación de metodologías existentes y experiencias propias en procesos de confección de almacenes de datos. Se enmarca en la etapa de diseño de un DWH, identificando las perspectivas de análisis a partir de diseños de bases de datos relacionales. Las otras dos etapas BI solo son mencionadas, obviando componentes importantes de la arquitectura de un DWH (área temporal y los metadatos) y la seguridad de la información. Esta metodología presenta un caso de estudio que permite entender mejor la propuesta (12).
- El ciclo de vida Kimball comienza con una planificación de proyecto, en la cual se define el alcance, se identifican y programan las tareas, se planifica el uso de los recursos, conformado así el plan de proyecto. En la segunda etapa de este ciclo se definen los requerimientos del negocio y luego el proyecto se enfoca en tres líneas concurrentes: tecnología, datos y aplicaciones de la inteligencia de negocios. Esta metodología aborda los elementos de forma amplia, dejando claro el qué se debe hacer pero no cómo hacerlo, pudiendo provocar demoras en la obtención del resultado final (13).
- La propuesta de Trujillo abarca sobre el uso de una arquitectura dirigida por modelos para el desarrollo multidimensional del DWH, facilitando el diseño mediante la automatización del proceso y la obtención de un modelo lógico a partir de un modelo multidimensional conceptual derivado tanto de los requisitos de información, como de las fuentes de datos. Esta no detalla la manera en que se deben diseñar los modelos de datos, ni la forma de obtener las variables para

lograr la correspondencia con los datos fuentes. Además no es contemplada como un ciclo completo de desarrollo de BI, sino que abarca solo la etapa de diseño del DWH (14).

- La metodología DWEP permite modelar todos los aspectos fundamentales en los modelos de datos lógico, conceptual y físico de los DWH y está basada en el proceso unificado manteniendo las cuatro fases del: inicio, elaboración, construcción y transición (33).
- RWM es una metodología iterativa que está basada en el desarrollo incremental del proyecto de un DWH dividido en cinco fases: definición de los objetivos, definición de los requerimientos de información, diseño y modelización, implementación y revisión (34).

Teniendo en cuenta el estudio de las metodologías de desarrollo que plantea anteriormente, se evidencia que optar por una de ellas para guiar el proceso en el CIM no facilitará la implementación de la solución. Esto se debe a que algunas no son contempladas como ciclo completo BI, por lo que se deberían combinar para su utilización. Otras no tiene en cuenta elementos arquitectónicos importantes como el área temporal y los metadatos. No se crean artefactos que especifiquen el cómo implementar la solución y no se presenta una forma para evaluar el proceso realizado. Por todo esto se decide elaborar un procedimiento, por etapas de desarrollo BI, que permita facilitar la implementación de la solución en el CIM.

A continuación se detallan las ventajas que se tendrán en cuenta de las metodologías consultadas para la elaboración del procedimiento, descartándose las metodologías DWEP y RWM porque lo encontrado en ambas, son elementos que se abordan de forma amplia en las demás propuestas metodológicas revisadas:

- La metodología basada en modelos propuesta por Trujillo presenta una base consistente en la especificación de los requisitos a través de la identificación de objetivos (13, p.4).
- En el caso de la metodología Hefesto posee una explicación detallada de cómo analizar las necesidades de la organización identificadas, y cómo realizar el diseño de los modelos de datos (12, p.82-85).
- Por otra parte, el ciclo de vida Kimball, posee una estructura bien organizada para el desarrollo de un DWH (14, p. 8.1-8.10) y aborda sobre el acceso a la información analizada a través de aplicaciones Web (14, p. 10.18). Tiene en cuenta la seguridad de los datos designando roles de

acceso para los análisis (14, p.3.15) y el tipo de gráficas de los reportes según las necesidades e información del cliente (14, p.1.6).

1.4. Lista de chequeo

Teniendo en cuenta que las metodologías de desarrollo BI consultadas no presentan un artefacto para evaluar el proceso realizado se analizaron los estándares que el Centro Nacional de Software CALISOFT ha establecido para definir su propio modelo de aseguramiento de la calidad de software. Uno de los mecanismos más utilizados por este centro para la evaluación y control de los productos informáticos es la lista de chequeo; documento que tiene un conjunto de parámetros a medir sobre un aspecto determinado para verificar el grado de cumplimiento de determinadas reglas. Consiste en un formulario de preguntas referentes al atributo de calidad que se está probando y de las características del documento en el caso de la documentación. Cada pregunta tiene asociada una evaluación que da una medida del grado de cumplimiento y disponibilidad del indicador evaluado.

1.5. Método para evaluar el procedimiento BI para los EC del CIM

Cuando se realiza una investigación de este tipo, uno de los principales inconvenientes que tiene asociado es la posibilidad de verificar y demostrar la confiabilidad de su propuesta resultante. Con el objetivo de erradicar este problema se crearon los métodos de expertos; los cuales utilizan como fuente de información y validación un grupo de personas. Teniendo en cuenta las ventajas y desventajas (36) que presentan los métodos expertos, lo ideal es utilizar un método que extraiga los beneficios de la interacción directa de estos y elimine sus inconvenientes. Esta intenta ser la filosofía del método Delphi, el cual se utilizará para la evaluación del procedimiento propuesto.

1.5.1. El método Delphi

El Delphi (35) (36) (37) es uno de los métodos de pronóstico más confiables a través de la elaboración estadística de las opiniones de un grupo de expertos en el tema tratado. Permite alcanzar una imagen integral de un fenómeno determinado, reflejando las valoraciones individuales de los expertos. El método se basa en la organización de un diálogo anónimo entre los expertos consultados de modo individual, a partir de la aplicación de un cuestionario. Las respuestas se procesan estadísticamente para conocer la coincidencia o discrepancia que estos presentan. Las características fundamentales son:

- Anonimato: los expertos contestan las preguntas sin consultarse mutuamente.

- Retroalimentación controlada: después de cada ronda de preguntas, se tabulan las respuestas y se procesan antes de la siguiente ronda, para que los participantes puedan evaluar los resultados de la ronda anterior, así como las razones dadas para cada respuesta y su dispersión del promedio. Esto permite que aumente el acuerdo al transcurrir varias rondas del proceso.
- Respuesta estadística del grupo: el procesamiento de cada ronda se realiza con métodos estadísticos. Esta es la característica más importante que diferencia a este método de otros existentes.

Por lo expresado anteriormente, es que se decide usar este método, en este caso la variante propuesta por Silvia Colunga y Georgina Amayuela (38) a su vez empleada por el Licenciado Carlos Álvarez Martínez de Santelices en su tesis de maestría (39). Dichos investigadores no utilizan el método clásico Delphi para la valoración de expertos, sino una variante para propiciar mayor objetividad a los criterios de los especialistas a partir de la introducción de escalas valorativas.

1.6. Conclusiones parciales

Después del análisis de la bibliografía sobre inteligencia de negocios y las características principales de la gestión de los EC en el CIM se arriban a las siguientes conclusiones:

- ✓ Se decide utilizar como metodología de arquitectura la propuesta por Kimball, siendo la más aceptada en todo el mundo como una metodología efectiva para desarrollar una solución de construcción de un DWH.
- ✓ Se propone tomar lo mejor y más aplicable de cada metodología analizada con el objetivo de elaborar un procedimiento para cada una de las tres etapas de BI; que permita integrar y analizar los datos de los EC que se gestionan en el CIM.
- ✓ Se propone crear tres listas de chequeo para evaluar el proceso BI en cada una de sus etapas, e incluirlas como artefactos del procedimiento.
- ✓ Se decide aplicar el procedimiento de BI que se obtenga a los resultados de algunos de los EC del producto Nimotuzumab (hR3).
- ✓ Se decide utilizar el método Delphi para la evaluación del procedimiento, haciendo uso principalmente de cuestionarios como técnica de recopilación de información de los expertos.

Capítulo 2: Desarrollo del procedimiento BI para los EC del CIM

En este capítulo se expone la propuesta de procedimiento contemplado como un ciclo completo de desarrollo BI, dividido por las tres etapas: diseño del DWH proceso ETL, análisis OLAP. A continuación se especifican cada una de las actividades comprendidas dentro de cada paso, así como los entregables que se generan; describiéndose los resultados más significativos de la aplicación del procedimiento a los EC del producto hR3 (40) (41) (42), para lograr un mayor entendimiento por parte del lector.

2.1. Etapa de diseño del DWH

La etapa de diseño del DWH (ver figura 2) consta de cinco pasos, de los cuales los cuatro primeros fueron definidos según la metodología Hefesto, incluyéndosele la identificación de las reglas del negocio, por su importancia para la especificación de variables en el CIM, la construcción de un diccionario de datos, para la identificación de variables que se relacionan con cada perspectiva e indicador y la evaluación del modelo conceptual y el diccionario de datos por parte de los especialistas del negocio. En el caso de las reglas fueron tomadas de la metodología para el desarrollo de software: Rational Unified Process y en el caso del diccionario de datos es un artefacto que se ajusta a la situación real del centro, el cual no cuenta con un diseño de base de datos relacional que permita hacer uso de las formas tradicionales. Se propone además un quinto paso que permita la evaluación del diseño a través de una lista de chequeo y la identificación de no conformidades.

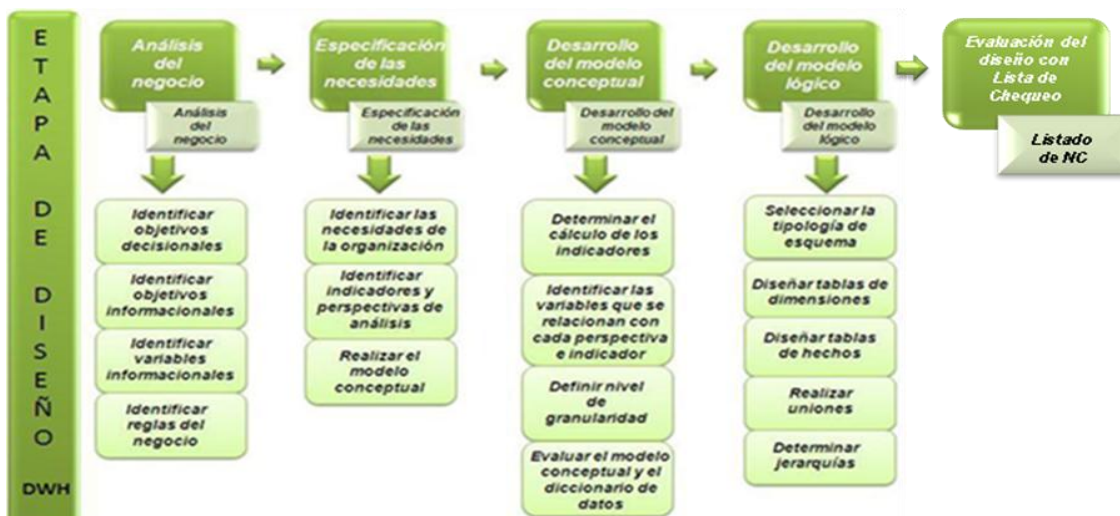


Figura 2 Procedimiento para la etapa de diseño del DWH

2.1.1. Análisis del negocio

Para desarrollar un DWH que cumpla con las necesidades reales de los usuarios y conduzca al éxito del proyecto, es muy importante la especificación de los requisitos. Esta tarea no es trivial ya que el análisis de requisitos en almacenes de datos está relacionado con la comprensión, por parte del diseñador, de los procesos de toma de decisiones y los objetivos que el usuario pretende alcanzar con la información suministrada por el DWH. Para identificar las necesidades de la organización, pueden llevarse a cabo diferentes técnicas como por ejemplo: entrevistas, cuestionarios y observaciones. Obtener directamente del usuario los requisitos es una tarea compleja, por lo que se propone para definirlos, luego de haber aplicado una de las técnicas de captura de requisitos, una aproximación basada en objetivos.

Entregable: Análisis del negocio.

Actividad 1: Identificar objetivos decisionales

Los principales objetivos del proceso de negocio de una empresa se definen por sus objetivos estratégicos, los cuales son definidos por los directivos de la institución. Estos objetivos corresponden con cambios desde una situación actual hacia otra mejor. Basados en los objetivos estratégicos, se deben especificar los objetivos decisionales, que permiten definir las acciones que se deben tomar para cumplir con un objetivo estratégico. El cumplimiento de un objetivo decisional solo es beneficioso para la organización si ayuda a alcanzar un objetivo estratégico (13, p. 5).

Caso práctico: A partir de los objetivos estratégicos del CIM, los cuales están relacionados con el análisis de la seguridad y eficacia de los productos que allí se desarrollan, se identificaron dos objetivos decisionales:

- *Verificar la eficacia del producto hR3 en los pacientes a los cuales se aplica.*
- *Verificar la seguridad del producto hR3 en los pacientes a los cuales se aplica.*

Actividad 2: Identificar objetivos informacionales

Los objetivos informacionales se obtienen de la información que se necesita para evaluar el cumplimiento de un objetivo decisional (13, p. 5).

Caso práctico: Se identificaron 14 objetivos informacionales relacionados con la información necesaria para alcanzar los objetivos decisionales. Para cada objetivo identificado se debe tener en cuenta el tiempo, la localización, el número del ensayo y los datos demográficos. Algunos de ellos son:

Para verificar la eficacia del producto hR3, es necesario conocer:

- *Relación entre la evaluación de la respuesta y el tratamiento recibido.*

- *Relación entre la evaluación de la respuesta y el número de dosis recibidas.*
- *Relación entre el tiempo de supervivencia y el tratamiento recibido.*
- *Relación entre el tiempo de supervivencia y el número de dosis recibidas.*
- *Relación entre la supervivencia y el tratamiento recibido.*
- *Relación entre la supervivencia y el número de dosis recibidas.*

Para verificar la seguridad del producto hR3, es necesario conocer:

- *Cuáles son los tipos de eventos adversos más frecuentes en cada tratamiento.*
- *Cuáles son los tipos de eventos adversos más frecuentes por número de dosis recibidas.*
- *Cuántos pacientes presentaron eventos adversos por tratamiento.*
- *Cuántos pacientes presentaron eventos adversos por número de dosis recibidas.*
- *Relación entre intensidad de los eventos adversos y tratamiento.*
- *Relación entre intensidad de los eventos adversos y número de dosis recibidas.*
- *Relación entre causalidad de los eventos adversos y tratamiento.*
- *Relación entre causalidad de los eventos adversos y número de dosis recibidas.*

Actividad 3: Identificar variables informacionales

En este paso se identifican las variables informacionales que se relacionan con los objetivos informacionales y se establece la correspondencia de estas variables con los datos fuentes. Se debe especificar cómo se van a encontrar las variables y en caso de ser calculables especificar la forma de calcularlas (13, p. 5).

Caso práctico: Alguna de las variables informacionales que se identificaron a partir de los objetivos informacionales fueron:

- *Tratamiento: es el nivel de dosis de determinado producto (hR3) o el placebo (0 mg).*
- *Número de dosis: es la cantidad de veces que se le suministra al paciente un determinado tratamiento.*
- *Tiempo de supervivencia: esta variable es calculable (hay que tomar la fecha de la última consulta realizada o la última visita o el último contacto con el paciente o la fecha de fallecido y restarla con la fecha de inclusión).*
- *Eventos adversos: son las reacciones negativas que manifiestan los pacientes durante el ensayo. Algunos tipos de eventos adversos son: cefalea, dolor de cabeza o migraña, fiebre, diarrea y vómitos.*

Actividad 4: Identificar reglas del negocio

Las reglas de negocio describen políticas que deben cumplirse o condiciones que deben satisfacerse, por lo que regulan algún aspecto del negocio, puede utilizarse para definir el valor de algunos datos en

blancos que son de importancia para el análisis. El proceso de especificación implica que hay que identificarlas dentro del negocio, evaluar si son relevantes dentro del campo de acción que se está modelando e implementarlas en la propuesta de solución. Es en este momento dónde se comienzan a identificar; pero durante toda la etapa de diseño se pueden estar identificando nuevas reglas del negocio, que puedan surgir de la situación que exista en los datos fuentes.

Caso práctico: En esta primera etapa se comenzaron a identificar las reglas del negocio claras para el cliente que están en correspondencia con las condiciones que debe cumplirse, a continuación se relacionan cinco de las 22 reglas identificadas:

- *Si un paciente está en un grupo de tratamiento donde se definió en el protocolo que recibe placebo (ningún producto) o es control, se define que el nivel de dosis es 0.*
- *La variable que identifica a un paciente en las bases de datos es una concatenación entre sus iniciales y el número de inclusión en el ensayo.*
- *Cuando la localización del tumor es mama (base de datos Mama 070) se asume que el sexo del paciente es femenino.*
- *Si la intensidad del evento adverso es normal entonces no se incluye como evento adverso.*
- *Para los tratamientos previos se determinó que si el paciente tiene al menos uno de los tres (quimioterapia, radioterapia o cirugía) ya se asume que el paciente tuvo algún tratamiento previo.*

2.1.2. Especificación de las necesidades

En esta etapa se realiza un trabajo de mesa con todo el equipo que desarrolla el DWH para realizar la especificación de las necesidades de la organización. Se identifican las necesidades de la organización a través del análisis de los objetivos y de las variables informacionales identificadas en el paso anterior y se redactan a manera de requisitos. Luego estas necesidades se analizarán para identificar cuáles serán los indicadores y perspectivas que serán tomadas en cuenta para la construcción del DWH. Finalmente se confeccionará un modelo conceptual donde se podrá visualizar el resultado obtenido en este segundo paso.

Entregable: Especificación de las necesidades.

Actividad 1: Identificar las necesidades de la organización

Se analizan cada uno de los objetivos y variables informacionales identificando cuáles son los elementos principales de cálculo y a través de cuáles aristas se analizarán los mismos. Se debe tener en cuenta que

cada una de estas necesidades deben ser: completas, consistentes, no ambiguas, precisas y verificables. (12, p. 82).

Caso práctico: A partir de los objetivos y variables informacionales se identificaron 10 necesidades de información agrupadas según los objetivos estratégicos que las identifican. Para cada una de estas necesidades se debe tener en cuenta el tiempo, la localización, el número del ensayo y los datos demográficos de los pacientes (sexo, edad, raza, estadio, clasificación anatomopatológica, grado de diferenciación y tratamientos previos).

Para el proceso de eficacia:

- *Cantidad de respuestas por tipos de respuestas para un tratamiento.*
- *Cantidad de respuestas por tipos de respuestas para una cantidad determinada de número de dosis.*
- *Tiempo de supervivencia para un tratamiento.*
- *Tiempo de supervivencia para una cantidad determinada de número de dosis.*
- *Cantidad de supervivencias para un tratamiento.*
- *Cantidad de supervivencias para una cantidad determinada de número de dosis.*

Para el proceso de seguridad:

- *Cantidad de eventos adversos por tipos, por intensidad y por causalidad de los eventos adversos según el tratamiento.*
- *Cantidad de eventos adversos por tipos, por intensidad y por causalidad de los eventos adversos según el número de dosis.*
- *Cantidad de pacientes con eventos adversos según tratamiento.*
- *Cantidad de pacientes con eventos adversos para una cantidad determinada de número de dosis.*

Actividad 2: Identificar indicadores y perspectivas de análisis

En este paso se debe proceder a la descomposición de las necesidades para identificar los indicadores que se utilizarán y las perspectivas de análisis que intervendrán. Para ello, se debe tener en cuenta que los indicadores son valores numéricos y representan lo que se desea analizar concretamente. En cambio, las perspectivas se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de resolver las necesidades planteadas. Cabe destacar, que la variable tiempo siempre se debe tener en cuenta como una perspectiva más de análisis, para cumplir con la característica del DWH de variable en el tiempo (12, p. 83).

Caso práctico: Los indicadores y perspectivas identificados en las necesidades de la organización se detallan a continuación en la tabla 2.

Tabla 2 Especificación de indicadores y perspectivas

Objetivo estratégico	Indicadores	Perspectivas
Seguridad	<ul style="list-style-type: none"> • Cantidad de respuestas • Tiempo de supervivencia • Cantidad de supervivencias 	<ul style="list-style-type: none"> • Tipos de respuestas • Tratamiento • Número de dosis • Tiempo • Localización • Número del ensayo • Datos demográficos
Eficacia	<ul style="list-style-type: none"> • Cantidad de eventos adversos • Cantidad de pacientes 	<ul style="list-style-type: none"> • Tipos de eventos adversos • Intensidad de los eventos adversos • Causalidad de los eventos adversos • Tratamiento • Número de dosis • Tiempo • Localización • Número del ensayo • Datos demográficos

Actividad 3: Realizar el modelo conceptual

En este paso se construirá un modelo conceptual a partir de los indicadores y perspectivas obtenidas en el paso anterior. El modelo se confecciona colocando en la parte izquierda las perspectivas seleccionadas, que serán unidas a un óvalo central que representa la relación que existe entre ellas, es decir los principales objetivos estratégicos de la institución. De dicha relación y entrelazados con flechas, se desprenden los indicadores, que se ubican en la parte derecha del esquema (12, p. 84-85).

Caso práctico: Se realizaron dos modelos conceptuales uno para cada objetivo estratégico. A continuación, en la figura 3 se presenta el de eficacia.

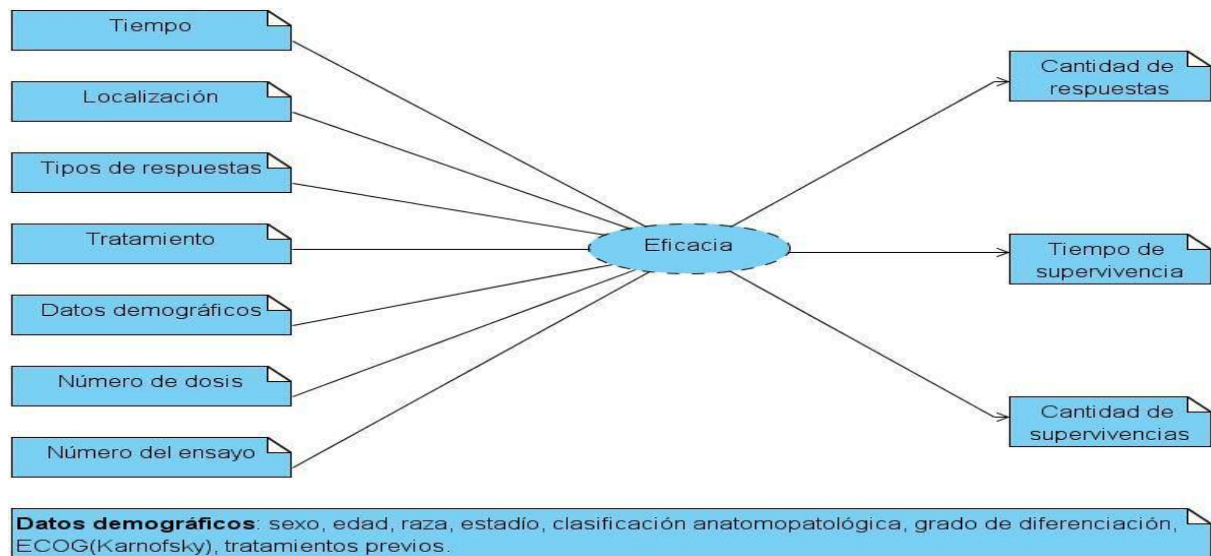


Figura 3 Modelo conceptual para eficacia

2.1.3. Desarrollo del modelo conceptual

Seguidamente se desarrollará el modelo conceptual para determinar cómo serán calculados los indicadores y para establecer las respectivas correspondencias entre las perspectivas del modelo conceptual y las variables que se relacionan, quedando explícita la relación en el diccionario de datos. Luego se define el nivel de granularidad, actualizándose el diccionario de datos al determinar la variable que integrará cada perspectiva. Finalmente, se evaluarán tanto el modelo conceptual como el diccionario de datos para pasar a la etapa del desarrollo del modelo lógico.

Entregable: Desarrollo del modelo conceptual.

Actividad 1: Determinar el cálculo de los indicadores

En este paso deberá quedar explícito cómo se calcularán los indicadores, definiendo su respectiva fórmula de cálculo, por ejemplo: suma de valores, el porciento de un indicador, contar las perspectivas (12, p. 85).

Caso práctico: A continuación se presentan los cálculos definidos para alguno de los indicadores:

- *Cantidad de respuesta: se obtiene al sumar cada una de las respuestas que cumplan con los criterios de análisis.*
- *Tiempo de supervivencia: para cada paciente se resta la fecha de inclusión de un paciente en el ensayo a la última fecha de contacto. El valor total se obtiene al calcular la mediana del tiempo de supervivencia de cada uno de los pacientes que cumplen con los criterios de análisis.*

Actividad 2: Identificar las variables que se relacionan con cada perspectiva

En este paso se revisarán todas aquellas fuentes de información con las que se cuentan para poblar el DWH. Para ello se realiza un análisis detallado de las mismas definiendo la correspondencia que existe entre las perspectivas con las variables de las tablas, bases de datos o modelos determinados. Se propone identificar estas correspondencias a través de un diccionario de datos.

Caso práctico: Se revisaron todas las fuentes de información proporcionadas por el CIM, realizándose un análisis exhaustivo y detallado de cada una de estas. Se encontraron las correspondencias existentes entre las variables de los modelos con las perspectivas identificadas; confeccionándose 10 diccionarios de datos, uno para cada base de datos. A continuación se presenta un ejemplo de uno de los diccionarios para el caso de la base de datos Cabeza y Cuello 040 (BD C y C 040), donde se especifican las perspectivas analizadas, su significado en el negocio, el modelo donde se puede encontrar en los datos fuentes, la variable dentro del modelo y los posibles valores que esta toma (ver tabla 3). A raíz de la correspondencia con las variables de los datos fuentes de algunas perspectivas como la edad, el sexo, entre otras se definieron nuevas reglas de negocio.

Tabla 3 Diccionario de datos - BD: hR3 C y C 040

<i>Perspectiva</i>	<i>Significado en el negocio</i>	<i>Modelo (s) donde se encuentra</i>	<i>Correspondencia con variable (s) de los datos fuentes</i>	<i>Posibles valores</i>
<i>Datos demográficos</i>				
<i>Sexo</i>	<i>Nos dice el sexo del paciente</i>	<i>Modelo 1. Datos Demográficos</i>	<i>Sexo (v11)</i>	<i>1:Femenino 2:Masculino</i>
<i>Edad</i>	<i>Nos dice la edad del paciente</i>	<i>Modelo 1. Datos Demográficos</i>	<i>Edad (v6)</i>	<i>Entre 18 y 80</i>
<i>Raza</i>	<i>Nos dice la raza del paciente</i>	<i>Modelo 1. Datos Demográficos</i>	<i>Raza (v10)</i>	<i>1: Blanca 2: Negra 3: Mestiza</i>
<i>Estadío</i>	<i>Nos dice el estadío en que se encuentra el paciente</i>	<i>Modelo1. Datos Demográficos</i>	<i>Estadío (v18)</i>	<i>1: I 2: II 3: III</i>
<i>Clasificación anatomopatológica</i>	<i>Nos dice la clasificación del tumor al diagnosticar al paciente</i>	<i>Modelo1. Datos Demográficos</i>	<i>V21a V21b</i>	<i>Texto hasta 80 caracteres. Estas variables deben concatenarse</i>
<i>Grado de diferenciación</i>	<i>Forma parte de la clasificación anatomopatológica</i>	<i>Modelo1. Datos Demográficos</i>	<i>V21a V21b</i>	<i>1.Bien Diferenciado 2.Moderado 3.Poco Diferenciado o Indiferenciado</i>
<i>ECOG (Karnofsky)</i>	<i>Nos dice el estado general del paciente cuando se diagnostica</i>	<i>Modelo1. Datos Demográficos</i>	<i>Estado (OMS) (v20)</i>	<i>De 0-2 según la escala de ECOG</i>

CAPÍTULO 2: DESARROLLO DEL PROCEDIMIENTO BI PARA LOS EC DEL CIM

<i>Perspectiva</i>	<i>Significado en el negocio</i>	<i>Modelo (s) donde se encuentra</i>	<i>Correspondencia con variable (s) de los datos fuentes</i>	<i>Posibles valores</i>
<i>Tratamientos previos</i>	<i>Nos dice el tratamiento que ha recibido el paciente antes del ensayo</i>	<i>Modelo1. Datos Demográficos</i>	<i>Cirugía (v22) Radioterapia (v24) Quimioterapia (v26)</i>	<i>1: True 2: False</i>
Tiempo de supervivencia				
<i>Fecha de inclusión en el ensayo</i>	<i>Nos dice la fecha de inclusión del paciente en el ensayo</i>	<i>Modelo 1. Datos Demográficos</i>	<i>Fecha (v3)</i>	<i>El valor de esta variable en los datos fuentes es una fecha</i>
<i>Fecha de último contacto con el paciente</i>	<i>Nos dice la última vez que se tuvo contacto con el paciente</i>	<i>Modelo 12. Salida del ensayo</i>	<i>Fecha de Fallecimiento (V10). Depende de V9 (Falleció, puede ser Y,N; si es Y se coge V10) Si no hay Fecha de Fallecimiento se busca Fecha de Salida (V7). Depende de V6 (Salió del ensayo, puede ser Y,N; si es Y se coge V7)</i>	<i>El valor de esta variable en los datos fuentes es una fecha</i>
Tratamiento				
<i>Nivel de dosis</i>	<i>Nivel de dosis aplicada al paciente</i>	<i>Modelo5. Monitoreo Clínico del Paciente.</i>	<i>(V5)</i>	<i>1: 50 mg 2: 100 mg 3: 200 mg 4: 400 mg</i>
<i>Número de dosis</i>	<i>Cantidad de dosis aplicadas al paciente</i>	<i>Modelo5. Monitoreo Clínico del Paciente.</i>	<i>(V6)</i>	<i>1: Primera Administración 2: Segunda Administración 3: Tercera Administración 4: Cuarta Administración 5: Quinta Administración 6: Sexta Administración</i>
Respuesta				
<i>Tipos de respuestas</i>	<i>Tipos de respuestas del paciente</i>	<i>Modelo 10. Evaluación de la Respuesta Clínica</i>	<i>(V12)</i>	<i>1: Remisión Completa 2: Remisión Parcial 3:</i>

<i>Perspectiva</i>	<i>Significado en el negocio</i>	<i>Modelo (s) donde se encuentra</i>	<i>Correspondencia con variable (s) de los datos fuentes</i>	<i>Posibles valores</i>
				<i>Estabilización 4: Progresión 5: Muerte Precoz</i>
Eventos Adversos				
<i>Tipos de eventos adversos</i>	<i>Son las reacciones negativas que manifiestan los pacientes durante el ensayo</i>	<i>Modelo 6. Reacciones Adversas</i>	<i>VARIABLES 6, 13, 20, 27, 34, 41, 48, 55, 62, 69 y la de otras (76). Todas estas variables están relacionadas con las de la columna de al lado que dice si el evento adverso se produjo o no.</i>	<i>Listado de eventos adversos</i>
<i>Causalidad de los eventos adversos</i>	<i>Especifica si el EA está relacionado o no con el tratamiento</i>	<i>No se recogió en ningún modelo. El cliente determinó llenar los campos de esta variable como missing.</i>		
<i>Tipos de eventos adversos</i>	<i>Son las reacciones negativas que manifiestan los pacientes durante el ensayo</i>	<i>Modelo 6. Reacciones Adversas</i>	<i>VARIABLES 6, 13, 20, 27, 34, 41, 48, 55, 62, 69 y la de otras (76). Todas estas variables están relacionadas con las de la columna de al lado que dice si el evento adverso se produjo o no.</i>	<i>Listado de eventos adversos</i>
<i>Intensidad de los eventos adversos</i>	<i>Grado de intensidad del EA</i>	<i>Modelo 6. Reacciones Adversas</i>	<i>VARIABLES 11, 18, 25, 32, 39, 46, 53, 60, 67, 74.</i>	<i>1. Leve: No requiere medicación. 2. Moderado: Responde a la medicación. 3. Severa: No respuesta a la medicación. 4. Muy Severa: Puede poner en peligro la vida del paciente.</i>

Actividad 3: Definir nivel de granularidad

La granularidad representa el nivel de detalle con el que se desea almacenar la información sobre el negocio analizado. Mientras mayor sea el nivel de detalle de los datos, se tendrán mayores posibilidades analíticas; ya que los mismos podrán ser resumidos. Es decir, los datos que posean granularidad fina podrán ser resumidos hasta obtener una granularidad media o gruesa. No sucede lo mismo en sentido

contrario, ya que por ejemplo, los datos almacenados con granularidad media podrán resumirse, pero no tendrán la facultad de ser analizados a nivel de detalle, es decir hasta una granularidad fina. En este paso se determina la variable que integrará cada perspectiva, actualizándose el diccionario de datos. Debe prestarse atención, ya que esta acción determinará la granularidad de la información encontrada en el DWH (12, p. 88).

Caso práctico: Sobre la base del negocio analizado se definió la granularidad a nivel de paciente pues es así como se desea almacenar la información en el DW; actualizándose el diccionario de datos, especificando cuál variable de las encontradas, integrará cada perspectiva.

Actividad 4: Evaluar el modelo conceptual y el diccionario de datos

En este paso se realiza una evaluación del diccionario de datos, así como del modelo conceptual para pasar a la próxima etapa del procedimiento. Esta evaluación se realiza en forma de revisión con el cliente con el objetivo de verificar si tanto los modelos conceptuales como los diccionarios de datos están en correspondencia con las necesidades de la institución. De ser satisfactoria esta evaluación se procede con el desarrollo del modelo lógico, de lo contrario se analiza cuál fue el error encontrado y se retorna al paso correspondiente para dar solución al problema.

Caso práctico: Se realizó la evaluación de los dos modelos conceptuales y de los diez diccionarios de datos por parte de los especialistas del CIM; resultando satisfactoria teniendo como constancia un aval de aceptación.

2.1.4. Desarrollo del modelo lógico

A continuación se debe confeccionar el modelo lógico de la estructura del DWH, teniendo como base el modelo conceptual que ya ha sido creado. Para ello, primero se define el tipo de modelo que se utilizará (estrella, constelación o copo de nieve) y luego se llevarán a cabo las acciones propias al caso, para diseñar las tablas de dimensiones y de hechos. Finalmente, se realizarán las uniones pertinentes entre estas tablas y se determinarán las jerarquías (14, p. 8.6).

Entregable: Desarrollo del modelo lógico.

Actividad 1: Seleccionar la tipología de esquema

Se debe seleccionar cuál será el tipo de esquema que se utilizará para contener la estructura del depósito de datos, que se adapte mejor a los requerimientos y necesidades del usuario. Es muy importante definir

objetivamente si se empleará un esquema en estrella, constelación de hechos o copo de nieve; ya que esta decisión afectará considerablemente la elaboración del modelo lógico (12, p. 91).

Caso práctico: Teniendo en cuenta que existían perspectivas comunes para los procesos que se están analizando, se decide emplear como tipología de esquema constelación de hechos.

Actividad 2: Diseñar tablas de dimensiones

En este paso se deben diseñar las tablas de dimensiones que formarán parte del DWH. Para los tres tipos de esquemas, cada perspectiva definida en el modelo conceptual constituirá una tabla de dimensión. Para ello deberá tomarse cada perspectiva con sus campos relacionados y realizarse el siguiente proceso:

- ✓ Se elegirá un nombre que identifique la tabla de dimensión.
- ✓ Se añadirá un campo que represente su clave principal.
- ✓ Se redefinirán los nombres de los campos si es que no son lo suficientemente intuitivos.

Para los esquemas copo de nieve, cuando existan jerarquías dentro de una tabla de dimensión, esta tabla deberá ser normalizada (12, p. 91-93)

Caso práctico: Se realizó el diseño de diez tablas de dimensiones teniendo en cuenta que cada perspectiva definida en modelo conceptual constituye una tabla de dimensión.

Actividad 3: Diseñar tablas de hechos

En este paso se definirán las tablas de hechos, que son las que contendrán los hechos (procesos) a través de los cuales se construirán los indicadores de estudio (12, p. 93-96) (40).

Para los esquemas en estrella y copo de nieve, se realizará lo siguiente:

- ✓ Se le deberá asignar un nombre a la tabla de hechos que represente el o los objetivos decisionales.
- ✓ Se definirá su clave primaria, que se compone de la combinación de las claves primarias de cada tabla de dimensión relacionada.
- ✓ Se crearán tantos campos de hechos como indicadores se hayan definido en el modelo conceptual y se les asignará los mismos nombres que estos. En caso que se prefiera, podrán ser nombrados de cualquier otro modo.

Para los esquemas constelación se realizará lo siguiente:

- ✓ Las tablas de hechos se deben confeccionar teniendo en cuenta el análisis de las necesidades de la organización obtenidas en pasos anteriores y sus respectivos indicadores y perspectivas.

- ✓ Cada tabla de hechos debe poseer un nombre que la identifique, contener sus hechos correspondientes y su clave debe estar formada por la combinación de las claves de las tablas de dimensiones relacionadas.
- ✓ Al diseñar las tablas de hechos, se deberá tener en cuenta:
 - Caso 1:* Si en dos o más necesidades figuran los mismos indicadores pero con diferentes perspectivas de análisis, existirán tantas tablas de hechos como necesidades cumplan esta condición.
 - Caso 2:* Si en dos o más necesidades figuran diferentes indicadores con diferentes perspectivas de análisis, existirán tantas tablas de hechos como necesidades cumplan esta condición.
 - Caso 3:* Si el conjunto de necesidades cumplen con las condiciones de los dos puntos anteriores se deberán unificar aquellas necesidades que posean diferentes indicadores pero iguales perspectivas de análisis, para luego reanudar el estudio de las necesidades.

Caso práctico: Se realizó el diseño de dos tablas de hechos teniendo en cuenta que cada proceso o relación entre perspectivas e indicadores que se definió en el modelo conceptual constituye una tabla de hecho.

Actividad 4: Realizar uniones

Luego de haber diseñado las tablas de dimensiones y las tablas de hechos a partir de los indicadores y las perspectivas identificados en el modelo conceptual, se realiza un modelo lógico previo que expresa las correspondencias que existen entre dichas tablas. Una vez que se obtiene este modelo se analizan las tablas de dimensiones para verificar que no existan datos redundantes. De ser necesario se pueden unir tablas de dimensiones con el objetivo de eliminar la redundancia de datos y obtener un modelo lógico final más eficiente (12, p. 96-97).

Caso práctico: Se realizó la unión de las tablas de dimensiones y de hechos para obtener el modelo lógico. En aras de eliminar la redundancia en los datos se desarrollaron dos modelos lógicos haciendo uso de la tipología constelación de hechos, quedando finalmente como se muestra en la figura 4.

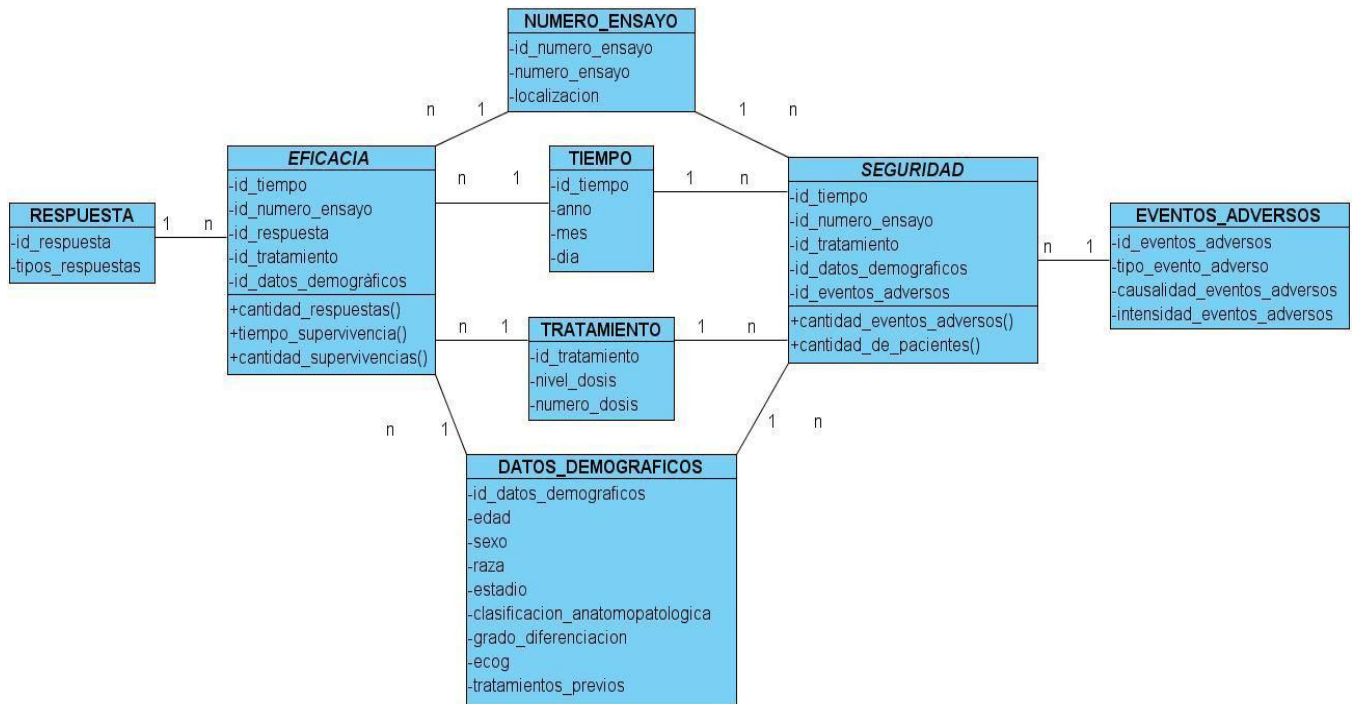


Figura 4 Modelo lógico final

Actividad 5: Determinar jerarquías

Una jerarquía representa una relación lógica entre dos o más atributos pertenecientes a una tabla de dimensión; siempre y cuando posean su correspondiente relación padre-hijo.

Las jerarquías poseen las siguientes características:

- ✓ Pueden existir varias en una misma tabla de dimensión.
- ✓ Están compuestas por dos o más niveles.
- ✓ Se tiene una relación 1-n o padre-hijo entre atributos consecutivos de un nivel superior y uno inferior.

La principal ventaja de manejar jerarquías, reside en poder analizar los datos desde su nivel más general al más detallado y viceversa, al desplazarse por los diferentes niveles (43).

Caso práctico: No se determinó ninguna jerarquía.

2.1.5. Evaluación del diseño con lista de chequeo

La lista de chequeo que a continuación se presenta (ver tabla 4), se utiliza para la evaluación de los entregables de la etapa de diseño del DWH. Esta fue confeccionada a partir de elementos importantes

que no se pueden obviar a la hora de realizar el diseño (12) (13) (14) y utilizando una estructura similar a las listas de chequeo que presenta el centro nacional de calidad de software CALISOFT que se componen de tres partes:

- ✓ **Estructura del documento:** abarca todos los aspectos definidos por el expediente de proyecto o el formato establecido por el proyecto.
- ✓ **Indicadores definidos por la etapa:** abarca todos los indicadores a evaluar durante la etapa de BI.
- ✓ **Semántica del documento:** contempla todos los indicadores a evaluar respecto a la ortografía, redacción y otros aspectos de forma y estilo.

Los elementos que forman parte de la estructura de la lista de chequeo son:

- ✓ **Peso:** define si el indicador a evaluar es crítico o no.
- ✓ **Indicadores a evaluar:** son los indicadores a evaluar en las secciones *Estructura del documento*, *Semántica del documento* e *Indicadores definidos por la etapa*.
- ✓ **Evaluación (Eval):** es la forma de evaluar el indicador en cuestión. El mismo se evalúa de 1 en caso de que exista alguna dificultad sobre el indicador y 0 en caso de que el indicador revisado no presente problemas.
- ✓ **NP (No Procede):** se usa para especificar que el indicador no es necesario evaluarlo en ese caso.
- ✓ **Cantidad de elementos afectados:** especifica la cantidad de errores encontrados sobre el mismo indicador.
- ✓ **Comentario:** especifica los señalamientos o sugerencias que quiera incluir la persona que aplica la lista de chequeo. Pueden o no existir señalamientos o sugerencias.

Una vez aplicada la lista de chequeo se detectan los indicadores evaluados de mal y con el objetivo de darles solución se especifican en una tabla de No Conformidades (NC), la cual presenta la siguiente estructura:

- ✓ **No.:** es un número consecutivo e indica la cantidad de no conformidades identificadas.
- ✓ **No Conformidad:** especifica la NC a la que se refiere.
- ✓ **Fase correspondiente:** especifica la fase o paso del procedimiento a la que corresponde la NC encontrada.
- ✓ **Significación:** especifica si la NC es o no significativa, dependiendo si el indicador es o no crítico.
- ✓ **Recomendación:** especifica si la NC es una recomendación, es decir que no es de obligatorio cumplimiento que se solucione por parte de los especialistas técnicos.

- ✓ **Estado NC:** especifica el estado de solución en que se encuentra la NC que puede ser pendiente o solucionada.
- ✓ **Respuesta del equipo de desarrollo:** si es necesario se especifica la respuesta dada, por el equipo de desarrollo, a la NC.

Para poder determinar la evaluación de la etapa revisada se utilizan los siguientes parámetros:

Es abortada la revisión si:

- ✓ Existen al menos dos indicadores críticos evaluados de mal en la sección **Indicadores evaluados por la etapa** que posee la lista de chequeo.
- ✓ Más del 50 % de los indicadores a evaluar están evaluados de mal.
- ✓ Se mantienen las NC de una revisión a otra.

Es evaluada de regular si no cumple los criterios para ser abortado y además:

- ✓ Incumple con los indicadores críticos a evaluar de las secciones **Estructura del documento** y **Semántica del documento** que posee la lista de chequeo.
- ✓ Existe al menos un indicador crítico evaluado de mal.
- ✓ Existen al menos cinco indicadores no críticos evaluados de mal de la sección **Indicadores evaluados por la etapa** que posee la lista de chequeo.

Es evaluado de bien si no cumple ninguno de los criterios anteriores y:

- ✓ No existe ningún indicador crítico evaluado de mal.
- ✓ Si la cantidad de indicadores no críticos evaluados de mal de la sección **Indicadores evaluados por la etapa** que posee la lista de chequeo no es mayor que cuatro.

Tabla 4 Lista de chequeo para evaluar el diseño

Peso	Indicadores a evaluar	Eval	NP	Cantidad de elementos afectados	Comentarios
Estructura del documento					
crítico	1. ¿Los entregables contienen las secciones obligatorias de la plantilla estándar definidas para un expediente de proyecto? (portada, control de versiones, reglas de confidencialidad, tabla de contenidos y contenido). (ver expediente de proyecto)				
Indicadores definidos por la etapa					
	1. ¿Se realizó un análisis profundo del área de la organización, responsable de la utilización del DWH?				

Peso	Indicadores a evaluar	Eval	NP	Cantidad de elementos afectados	Comentarios
	2. ¿Se realizó un análisis profundo de las fuentes de datos que poblarán el DWH?				
	3. ¿Las personas que realizarán la estructuración del DWH están capacitadas para eso?				
crítico	4. ¿Existe una total correspondencia entre el análisis del negocio realizado y los objetivos estratégicos de la institución?				
	5. ¿Las necesidades de la organización se encuentran correctamente especificadas; es decir: completas, consistentes, no ambiguas, precisas, verificables, modificables y que puedan ser leídas?				
crítico	6. ¿Las reglas del negocio identificadas regulan algún aspecto del negocio?				
	7. ¿Los indicadores son valores numéricos y representan lo que se desea analizar concretamente?				
	8. ¿Quedó explícito cómo se calculará cada indicador?				
	9. ¿Las perspectivas se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de resolver las necesidades planteadas?				
	10. ¿Se definieron las correspondencias que existen entre las variables de las tablas, bases de datos o modelos determinados con sus perspectivas?				
crítico	11. ¿Posee el modelo conceptual un alto nivel de definición de los datos y permite que su explicación sea fácil de entender por los usuarios?				
crítico	12. ¿La granularidad definida representa el nivel de detalle con el que se desea almacenar la información sobre el negocio analizado?				
crítico	13. ¿La granularidad definida permite tener mayores posibilidades analíticas?				
	14. ¿La tipología de esquema seleccionada es la que mejor se adapta a los requerimientos y necesidades del usuario?				
	15. ¿Se diseñaron correctamente las tablas de dimensiones?				
	16. ¿Se diseñaron correctamente las tablas de hechos?				
	17. ¿Se realizaron las uniones correspondientes				

Peso	Indicadores a evaluar	Eval	NP	Cantidad de elementos afectados	Comentarios
	entre las tablas de dimensiones y las tablas de hechos?				
	18. ¿La jerarquía planteada tiene una relación lógica “padre-hijo”?				
crítico	19. ¿Permite el modelo lógico tener una primera representación de la estructura de los datos?				
Semántica del documento					
crítico	1. ¿Se han identificado errores ortográficos en los entregables o en los modelos diseñados?				
crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?				
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?				

Caso práctico: Por último se procedió a evaluar el diseño realizado utilizando la lista de chequeo, donde se identificó un indicador con dificultad, del cual se generó una NC como se muestra en la tabla 5. Esta fue tratada durante la etapa de desarrollo del modelo conceptual por parte de los especialistas técnicos, a través de la identificación de las reglas del negocio. De manera general el diseño es evaluado de “Bien” ya que no hubo ningún indicador crítico evaluado de mal y no existieron problemas con los formatos de las plantillas. Además no se encontraron errores ortográficos en los documentos revisados y la NC generada fue solucionada.

Tabla 5 No conformidades

No.	NC	Fase correspondiente	Significación	Recomendación	Estado NC	Respuesta del equipo de desarrollo
1	No se definieron algunas (13) de las correspondencias que deben existir entre las variables de las tablas, bases de datos o modelos determinados con sus perspectivas.	Desarrollo del modelo conceptual	No significativa		Resuelta	No se encontraron los valores de las perspectivas en los modelos debido a que cuando se condujo el ensayo no se recogieron estos datos. No obstante, se decidió para cada caso darle solución a través de las reglas del negocio.

2.2. Proceso ETL

El proceso ETL (ver figura 5) consta de siete pasos de los cuales cinco fueron definidos según el ciclo de vida Kimball, estos son: diseño de la arquitectura, selección de herramientas informáticas, desarrollo del modelo físico, ETL y definición del período de actualización del DWH. Se agregó el análisis de las fuentes

de datos, a partir del diccionario de datos que se construye en el diseño del DWH, y la evaluación del proceso a través de una lista de chequeo y la identificación de no conformidades. En el caso del paso ETL se propone la identificación de las reglas de transformación a partir de las reglas de negocio identificadas en el diseño del DWH.

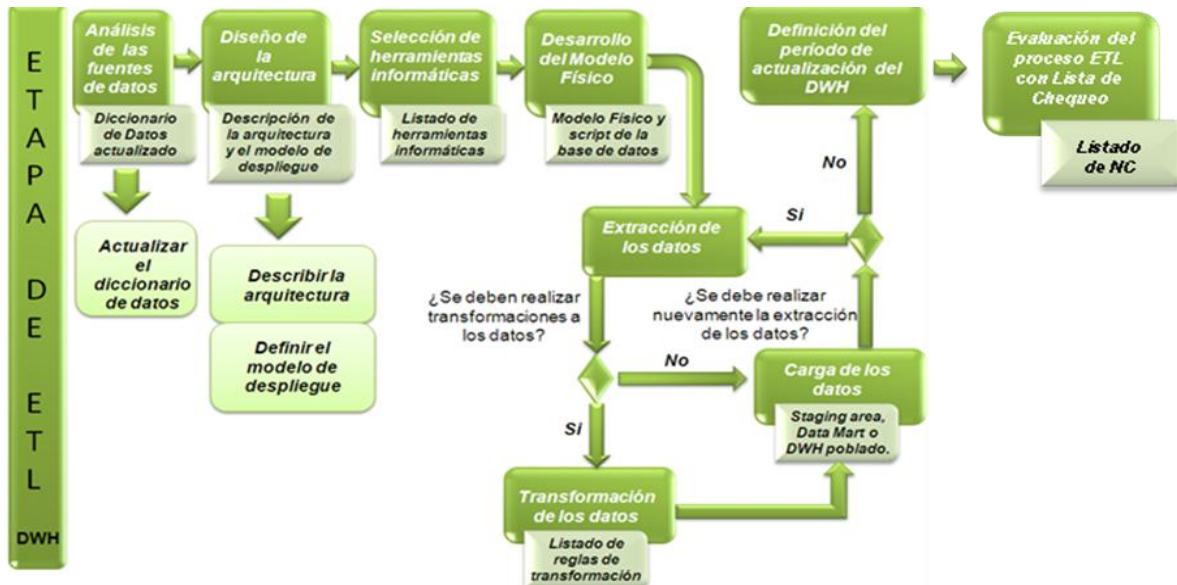


Figura 5 Procedimiento para la etapa de ETL

2.2.1. Análisis de las fuentes de datos

Para la realización de este paso se deben revisar los datos fuentes con los que cuenta la institución. Para esto se deben revisar los diccionarios de datos generados en la etapa de diseño para conocer la correspondencia de las perspectivas con sus datos fuentes y el significado para el negocio de cada una de las ellas. Además de conocer en qué modelos se encuentran las perspectivas y los posibles valores que estas pueden tomar. A partir de la revisión de los diccionarios se define el tipo de dato de cada variable que va a intervenir en la extracción de los datos y se especifica el formato en que se encuentran las fuentes de datos que van a alimentar el DWH.

Entregable: Diccionario de datos actualizado.

2.2.2. Diseño de la arquitectura

Este paso hace referencia al proceso de elegir y diseñar la estructura interna del DWH, basándose en que la misma estará formada por diversos componentes que interactúan entre sí y que cumplen una función

específica dentro del sistema. Debe definirse la estrategia arquitectónica a usar, ya sea Inmon (descendente) o Kimball (ascendente); además se define si se hará uso del *Staging area* (área temporal) y los metadatos (información del proceso). Se deben especificar los requerimientos de hardware necesarios para asumir la implementación del DWH así como la estructura del despliegue del DWH.

Entregable: Descripción de la arquitectura y modelo de despliegue.

Caso práctico: La arquitectura del DWH (ver figura 6) se diseñó teniendo en cuenta la estrategia ascendente propuesta por Kimball, donde cada Data Mart representa a un producto del CIM y el DWH integra los resultados de análisis para cualquier localización y producto a partir de los distintos Data Marts. En el caso del modelo de despliegue se define la estructura física de los procesos de ETL usados en la carga de los datos hacia el DWH (ver figura 7).

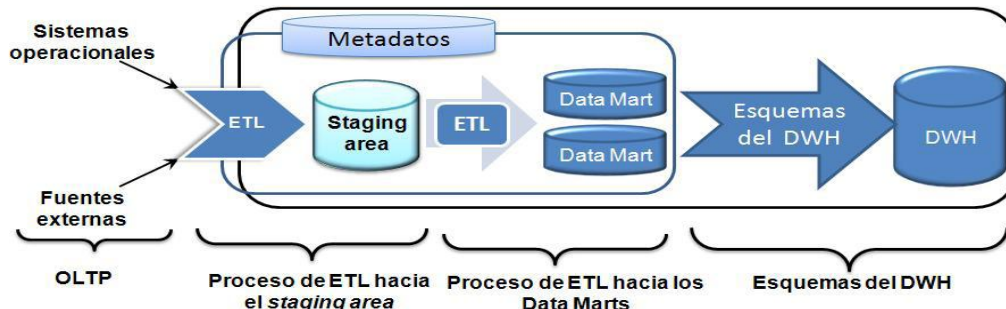


Figura 6 Arquitectura de DWH para el CIM

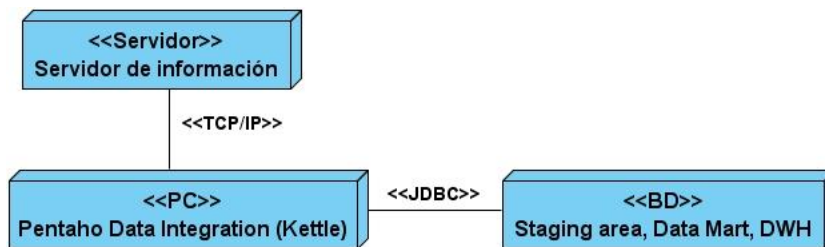


Figura 7 Modelo de despliegue

2.2.3. Selección de las herramientas informáticas para implementar el proceso de ETL

Existen varios sistemas que facilitan realizar la actividad de extraer datos desde diferentes fuentes, para luego integrarlos, filtrarlos y depurarlos como son: Kettle, Talend, Scriptella, Octopus. Para saber cuál de ellos es conveniente utilizar es necesario hacer un estudio de sus características, teniendo en cuenta una serie de aspectos importantes relacionados con el negocio. Algunos de estos aspectos podrían ser: la capacidad de la herramienta para reconocer todas las fuentes de datos definidas en el negocio y el

soporte para el modelo de despliegue definido en el paso anterior. Además se debe tener en cuenta el gestor de bases de datos a utilizar según los requerimientos de hardware definidos en el paso anterior y el modelado de los datos (14, p. 8.5).

Entregable: Listado de herramientas informáticas.

Caso práctico: Para la implementación y configuración del proceso de ETL se decidió optar por la herramienta informática Kettle en su versión 3.1 y como gestor de bases de datos PostgreSQL en su versión 8.3.

2.2.4. Desarrollo del modelo físico

En este paso es donde se transforma el modelo lógico, que se obtuvo a partir de las necesidades del negocio identificadas durante la etapa de diseño, a un modelo físico. En él se especifican los tipos de datos de las perspectivas y la cardinalidad entre las tablas. Además se genera el script de la base de datos teniendo en cuenta el gestor seleccionado en el paso anterior (14, p. 8.6) (44).

Entregable: Modelo físico y script de la BD.

Caso práctico: Se definieron dos modelos físicos: uno para el staging area y otro para el data mart. Este último se obtuvo a partir del modelo lógico obtenido en la etapa de diseño (ver figura 8).

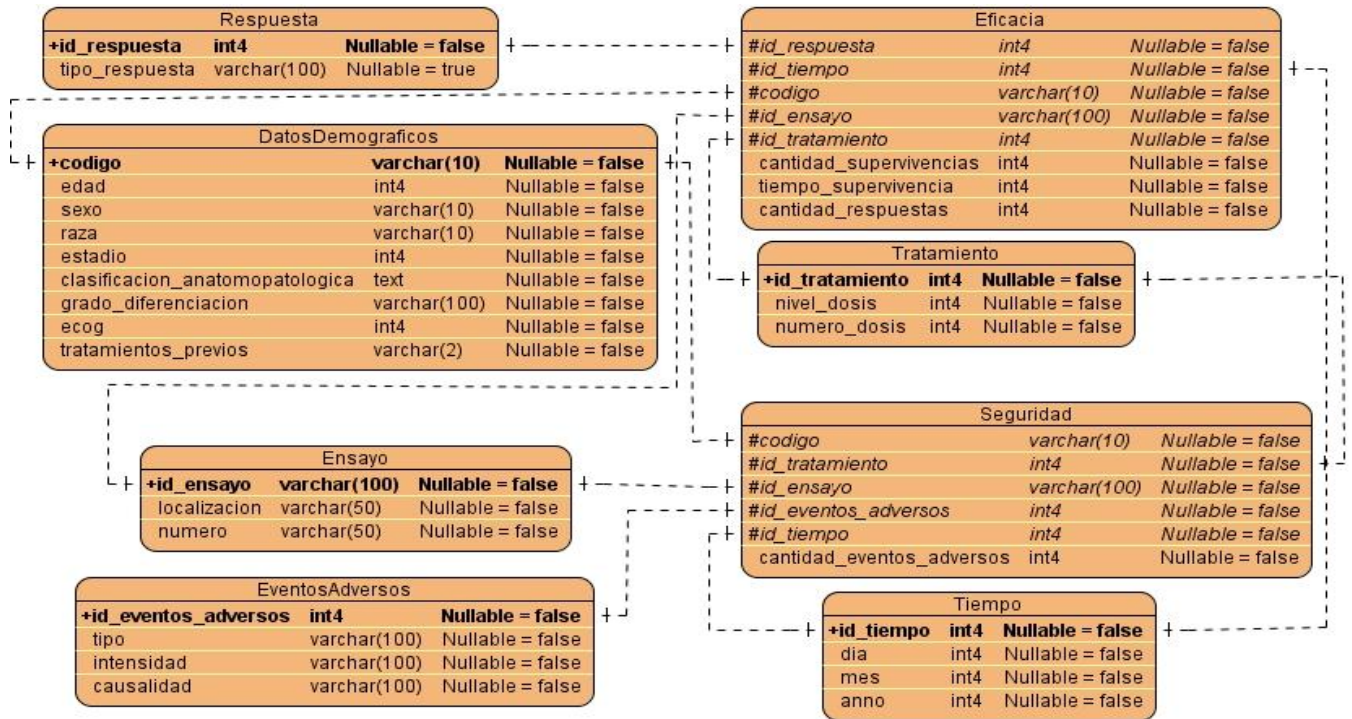


Figura 8 Modelo físico del Data Mart

2.2.5. Extracción, transformación y carga

En esta fase se realiza el proceso de extracción, transformación y carga de los datos hacia el DWH haciendo uso de la herramienta seleccionada. Es válido destacar que este proceso puede repetirse o ajustarse el orden de las acciones, teniendo en cuenta la arquitectura definida anteriormente (12, p. 97) (14, p. 1.6) (45).

Caso práctico: Para poblar el Data Mart del producto hR3 con los datos fuentes que se obtuvieron en la actualización del diccionario, se hizo necesario realizar el proceso de ETL dos veces, cumpliendo de esta forma con la arquitectura definida. Primero se pasó toda la información necesaria contenida en los ficheros excel hacia el Staging area, realizando el menor número de transformaciones posibles. Luego se insertaron los datos del Staging area dentro del Data Mart del producto hR3, donde se realizaron las transformaciones necesarias para que los datos quedaran integrados.

Actividad 1: Extracción de datos

A partir del diccionario de datos actualizado se procede a configurar e implementar el proceso de extracción con la herramienta previamente seleccionada. Una vez concluido este paso los datos quedarán

convertidos en un formato preparado para iniciar el proceso de transformación o para ser cargados dependiendo de la arquitectura definida y la naturaleza de los datos. Si se decide transformar los datos se deben definir las transformaciones, de lo contrario se pasa al paso de la carga.

Actividad 2: Definición de las transformaciones

En este paso se tiene como entrada la especificación de las reglas del negocio y la especificación de los indicadores que se obtienen del análisis con el cliente durante el diseño; los cuales se convierten en transformaciones necesarias sobre los datos extraídos. Es válido aclarar que pueden surgir transformaciones adicionales para facilitar el proceso de ETL, sin afectar el negocio. Luego de haber definido todas las transformaciones, se inicia la configuración e implementación de las mismas en la herramienta con la que se va a trabajar.

Entregable: Reglas de transformación.

Caso práctico: La definición de las transformaciones se realizó en dos ocasiones: antes de cargar los datos hacia el staging area y antes de cargar los datos al Data Mart. Hacia el staging area se definieron cuatro transformaciones que tienen correspondencia con algunas de las reglas del negocio identificadas en el diseño del DWH, de las cuales dos se presentan a continuación:

- *Definir si tuvo tratamientos previos: los únicos tratamientos previos que se tendrán en cuenta para el análisis serán quimioterapia, radioterapia y cirugía. Si el paciente participó en al menos uno de ellos, la variable tratamientos previos toma valor verdadero, en caso contrario se almacena como falso.*
- *Cambio de variable: se refiere a llevar todos los nombres de variables de las fuentes de datos a un lenguaje con más significado para el negocio, por ejemplo: la variable v6 es utilizada en los modelos excel para hacer referencia a la edad del paciente, aplicándole esta transformación quedaría con el nombre edad.*

En el caso del Data Mart se definieron 12 transformaciones, las cuales se corresponden con las reglas del negocio identificadas en el diseño del DWH, de las cuales se presentan a continuación tres de ellas:

- *Placebo: si un paciente está en un grupo de tratamiento donde se definió en el protocolo que recibe placebo (ningún producto) o es control, se define que el nivel de dosis es 0.*
- *Nivel de dosis igual cero: con esta transformación si el nivel de dosis aplicado a un paciente es cero y el mismo presenta eventos adversos, la causalidad del evento adverso se toma como no relacionada.*
- *Calcular edad: esta transformación es aplicada cuando la edad de un paciente no está registrada y la misma se puede obtener a partir de la resta de la fecha de fallecimiento y nacimiento de esta persona. Para la realización de este cálculo se ha escrito una función en código JavaScript.*

Actividad 3: Carga de los datos

Una vez que los datos han pasado por todo el proceso de extracción y/o transformación, solo queda almacenarlos en el DWH. Para la realización de este paso se debe configurar e implementar la carga mediante la herramienta seleccionada.

Entregable: DWH poblado.

2.2.6. Definición del período de actualización del datawarehouse

En este paso es donde se define la periodicidad de carga de los datos hacia el DWH teniendo en cuenta las necesidades del cliente. Luego se configura en la herramienta el período de actualización definido (14, p. 2.5) (46).

2.2.7. Evaluación del proceso ETL con lista de chequeo

La lista de chequeo que a continuación se presenta (ver tabla 6), se utiliza para la evaluación del proceso ETL desarrollado. Esta fue confeccionada a partir de elementos claves del proceso ETL (12, p. 97), (43, p. 1044). Los elementos que la conforman y la forma de evaluar se describen en el epígrafe 2.1.5.

Tabla 6 Lista de chequeo para evaluar el proceso ETL

Peso	Indicadores a evaluar	Eval	NP	Cantidad de elementos afectados	Comentarios
Estructura del documento					
crítico	1. ¿Los entregables contienen las secciones obligatorias de la plantilla estándar definidas para un expediente de proyecto? (Portada, Control de Versiones, Reglas de Confidencialidad, Tabla de Contenidos y Contenido). (ver Expediente de Proyecto)				
Indicadores definidos por la etapa					
	1. ¿La arquitectura satisface las necesidades del proyecto?				
	2. ¿La arquitectura soporta el incremento del proyecto?				
	3. ¿Se utilizó el menor número de transformaciones posibles al cargar los datos hacia el <i>Staging area</i> ?				
crítico	4. ¿Se creó el modelo físico a partir del modelo lógico?				
crítico	5. ¿Cumple la implementación del proceso de ETL con la arquitectura definida?				

Peso	Indicadores a evaluar	Eval	NP	Cantidad de elementos afectados	Comentarios
	6. ¿Se tuvo en cuenta los formatos fuentes y tipos de datos de las perspectivas de análisis?				
	7. ¿La extracción de los datos se realiza a partir de las fuentes de datos?				
	8. ¿Se realiza una limpieza de los datos antes de realizar la carga de los mismos?				
crítico	9. ¿No afectan las restricciones aplicadas sobre una dimensión para analizar una tabla hecho en específico, a otra(s) tabla(s) hecho(s) que comparten esta misma dimensión?				
crítico	10. ¿Se cargan primero los datos de las tablas de dimensiones y luego los de las tablas de hechos?				
crítico	11. ¿De utilizar un esquema copo de nieve, se comienza cargando las tablas de dimensiones del nivel más general al más detallado?				
	12. ¿Se establecen las políticas de actualización y refresco de los datos?				
	13. ¿Se crearon claves en el DWH diferentes a las claves de los OLTP?				
Semántica del documento					
crítico	1. ¿Se han identificado errores ortográficos en los entregables o en los modelos diseñados?				
crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?				
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?				

Caso práctico: Por último se procedió a evaluar el proceso ETL implementado utilizando la lista de chequeo, donde se identificó un indicador con dificultad, del cual se generó una NC como se muestra en la tabla 7. Esta fue tratada durante la etapa Implementación del proceso ETL por parte de los especialistas técnicos, a través de la identificación de las reglas de transformación. De manera general el proceso ETL es evaluado de “Bien” ya que no hubo ningún indicador crítico evaluado de mal y no existieron problemas con los formatos de las plantillas. Además no se encontraron errores ortográficos en los documentos revisados y la NC generada fue solucionada.

Tabla 7 No conformidades

No	No conformidad	Fase correspondiente	Significación	Recomendaciones	Estado NC	Respuesta del equipo de desarrollo

No	No conformidad	Fase correspondiente	Significación	Recomendaciones	Estado NC	Respuesta del equipo de desarrollo
1	No se definieron 13 de las correspondencias que deben existir entre las variables de las tablas, BD o modelos determinados con sus perspectivas.	Implementación del proceso de ETL.	No significativa		Resuelta	No se definieron porque no se encontraron los valores de las perspectivas en los modelos, debido a que cuando se condujo el ensayo no se recogieron estos datos. No obstante se decidió para cada caso darle solución a través de la creación de transformaciones a partir de las reglas del negocio.

2.3. Análisis de datos, con técnicas OLAP

La etapa de análisis de datos aplicando técnicas OLAP (ver figura 9) consta de seis pasos, de los cuales los cinco primeros fueron definidos según el ciclo de vida Kimball, incluyendo como sexto paso la evaluación del diseño a través de una lista de chequeo y la identificación de no conformidades.

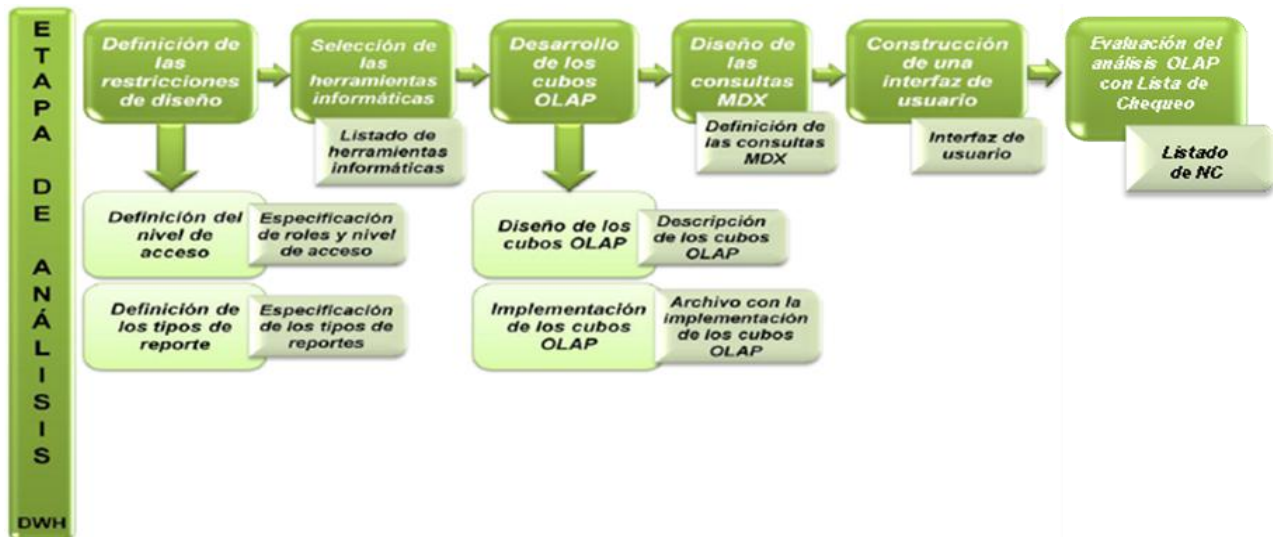


Figura 9 Procedimiento para la etapa de análisis con técnicas OLAP

2.3.1. Definición de las restricciones del diseño

Para una correcta definición de las restricciones de diseño que deberá tener la aplicación de análisis, se debe definir con claridad el nivel de acceso que tendrán los clientes a la hora de visualizar los reportes a partir de la definición de roles; además se deben definir los tipos de reportes que se desean obtener por

los clientes; es decir, el tipo de gráfica que por defecto desean utilizar en la visualización de los resultados del análisis.

Actividad 1: Definición del nivel de acceso

En este paso se definen los diferentes roles y el nivel de acceso que tendrán los clientes para visualizar los reportes, en aras de garantizar la seguridad de la información.

Entregable: Especificación de roles y nivel de acceso.

Caso práctico: Se definieron los tipos de roles que tendrán acceso a la información del almacén (ver tabla 8), garantizando así la seguridad de la información y la característica de no volátil de un DW; ya que el rol definido solo tendrá acceso a los datos cargados y no podrá modificar ni eliminar ninguna información.

Tabla 8 Especificación de roles y nivel de acceso

Rol	Nivel de acceso	Descripción
Experto	Administrador del sistema	Encargado de filtrar, reestructurar, actualizar el reporte que se quiere en tiempo real, configurar los datos y sus conjuntos significativos. Visualizador de la información.

Actividad 2: Definición de los tipos de reportes

En este paso se analizan los indicadores que se identificaron durante el análisis del negocio de la etapa de diseño del DWH, para definir los tipos de reportes (tipos de gráficas: pastel, barra, etc.) que se mostrarán a los clientes. Seguidamente se determinan los tipos de reportes para cada indicador.

Entregable: Especificación de los tipos de reportes.

Caso práctico: Se definieron los tipos de reportes que se desean obtener para garantizar el análisis de las necesidades de la organización identificadas en el diseño (ver tabla 9).

Tabla 9 Especificación de los tipos de reportes

Nombre del reporte	Descripción de la consulta	Tipo de reporte
Cantidad de respuestas	Cantidad de respuestas por tipos de respuesta para un tratamiento según datos demográficos.	Gráfica de barras en tres dimensiones (3D)
	Cantidad de respuestas por tipos de respuesta para una cantidad determinada de número de dosis según datos demográficos.	
Tiempo de supervivencia	Tiempo de supervivencia para un tratamiento según datos demográficos.	Gráfica de líneas
	Tiempo de supervivencia para una cantidad determinada de número de dosis según datos demográficos.	
Cantidad de	Cantidad de supervivencias para un tratamiento según datos	Gráfica de líneas

<i>Nombre del reporte</i>	<i>Descripción de la consulta</i>	<i>Tipo de reporte</i>
<i>supervivencias</i>	<i>demográficos.</i>	
	<i>Cantidad de supervivencias para una cantidad determinada de número de dosis según datos demográficos.</i>	
<i>Cantidad de eventos adversos (EA)</i>	<i>Cantidad de eventos adversos por tipos, por intensidad y por causalidad de los EA según el tratamiento.</i>	<i>Gráfica de barras en tres dimensiones (3D)</i>
	<i>Cantidad de eventos adversos por tipos, por intensidad y por causalidad de los EA según el número de dosis.</i>	
<i>Cantidad de pacientes</i>	<i>Cantidad de pacientes con EA por tratamiento.</i>	<i>Gráfica de barras en tres dimensiones (3D)</i>
	<i>Cantidad de pacientes con EA para una cantidad determinada de número de dosis.</i>	

2.3.2. Selección de las herramientas informáticas para implementar análisis OLAP

Existe un grupo de sistemas que facilitan el análisis a los datos para mostrar mediante gráficas los resultados del mismo; y así apoyar la toma de decisiones por parte de los expertos. Alguna de estas herramientas informáticas son: Jedox PALO, JRubik, Mondrian y JPivot. Para saber cuál de estas es conveniente utilizar es necesario hacer un estudio de sus características, teniendo en cuenta las características del negocio. Algunos de estos aspectos son: el soporte de la herramienta para con la fuente de datos original, la capacidad de la herramienta para realizar los distintos tipos de gráficas necesarias para el correcto análisis de los datos definidos en el negocio. Otro criterio de selección es que la herramienta sea de fácil manejo en el entorno en que se dará uso a la misma, así como que garantice la total funcionalidad requerida en el negocio. Después de tener en cuenta todos estos criterios se procede a seleccionar para el análisis de datos del DWH un servidor OLAP, un cliente que permita recoger datos del servidor OLAP seleccionado y realizar operaciones sobre estos; un desarrollador de cubos OLAP que permita el desarrollo y prueba de los mismos y un servidor web donde se montará el servidor OLAP para soportar el análisis en línea de los datos almacenados en el DWH (47).

Entregable: Listado de herramientas informáticas.

Caso práctico: Para la implementación del análisis OLAP a los datos del producto hR3 se decidió como servidor OLAP: Mondrian 3.0, como cliente OLAP: JPivot 3.0, para el desarrollo de los cubos OLAP: Shema Workbench 3.0.3.11016 y como servidor Web: Apache JakartaTomcat 5.5.

2.3.3. Desarrollo de los cubos OLAP

Para desarrollar los cubos OLAP estos deben ser diseñados e implementados, haciendo uso de las herramientas seleccionadas con este fin. Los cubos OLAP se obtienen de un análisis del negocio, y se corresponden con objetivos decisionales que se identifiquen.

Caso práctico: Teniendo en cuenta los dos objetivos estratégicos se diseñaron e implementaron dos cubos OLAP, uno para la eficacia y otro para la seguridad.

Actividad 1: Diseño de los cubos OLAP

En este paso debe hacerse un estudio sobre las necesidades del negocio, los indicadores y perspectivas de análisis identificados durante la etapa de diseño del DWH; así como los tipos de reportes. Además, debe analizarse la granularidad presente en el DWH, debido a que todo esto es necesario para el diseño de los cubos OLAP, ya que según el reporte y el grado de detalle que se quiera obtener en este, se definen las medidas y dimensiones que serían los ejes dimensionales del cubo OLAP.

Entregable: Descripción de los cubos OLAP.

Actividad 2: Implementación de los cubos OLAP

En este paso a partir de las dimensiones y medidas del cubo OLAP descrita en el paso anterior, se procede a crear cada uno de los cubos necesarios para obtener los reportes que se desean, usando para ello las herramientas seleccionadas.

Entregable: Archivo con la implementación de los cubos OLAP.

2.3.4. Diseño de las consultas MDX

En este paso se deben definir las consultas MDX, las cuales deben tener correspondencia con la implementación de los cubos OLAP obtenidos en el paso anterior (48).

Entregable 6: Definición de las consultas MDX.

Caso práctico: Se definieron las diez consultas, descritas en lenguaje MDX, para obtener los resultados de análisis de los datos del DWH que están en correspondencia con las diez necesidades del negocio.

2.3.5. Construcción de una interfaz de usuario

En este paso se debe realizar una interfaz de usuario para que el cliente visualice los resultados de la ejecución de las consultas MDX diseñadas en el paso anterior. La visualización se hará en correspondencia con los tipos de reportes identificados. Finalmente como una tarea clave se encuentra la capacitación de los clientes para el buen funcionamiento de la herramienta de análisis OLAP, la cual apoyará el proceso de toma de decisiones de la institución (14, p. 1.6).

Entregable: Interfaz de usuario.

Caso práctico: Se desarrolló una interfaz de usuario, con la herramienta Kompozer, para permitir la navegabilidad de los usuarios por los reportes de análisis del DWH poblado para el producto hR3. Esta interfaz, además de permitir la autenticación de los usuarios, cumpliendo con el requisito de seguridad, ofrece a los mismos la opción de navegar por un entorno agradable y ameno (ver figura 10 y 11).



Figura 10 Interfaz principal

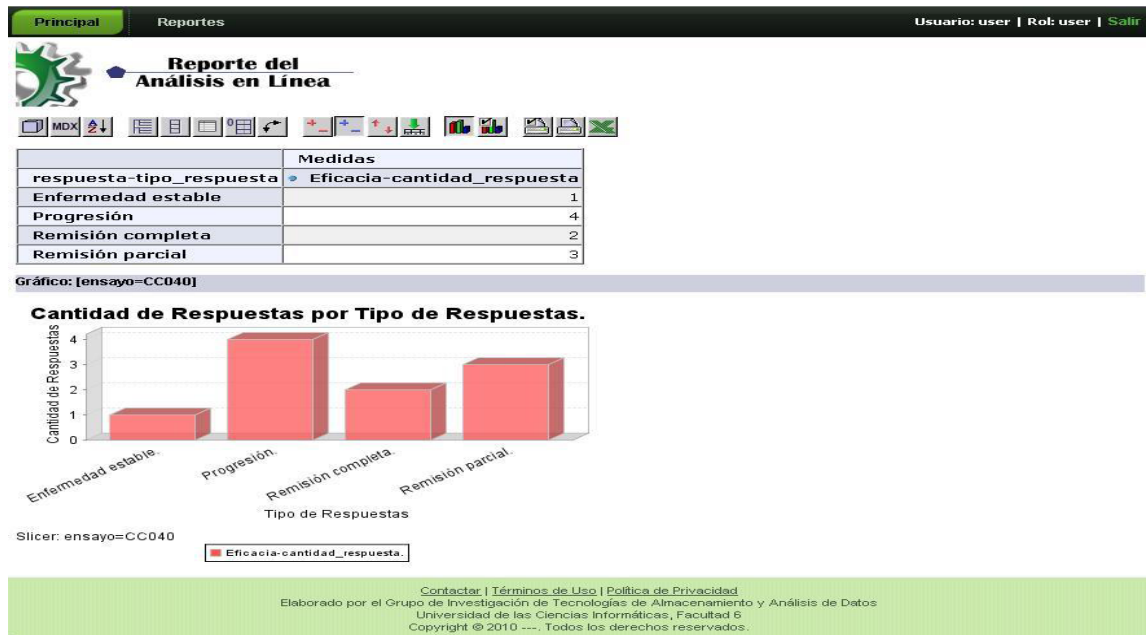


Figura 11 Interfaz para visualizar reportes

2.3.6. Evaluación del análisis con lista de chequeo

La lista de chequeo que a continuación se presenta (ver tabla 10), se utiliza para la evaluación del análisis de datos desarrollado. Esta fue confeccionada a partir de las reglas de Cood (43, p. 1095) que especifican las leyes del procesamiento analítico en línea. Los elementos que la conforman y la forma de evaluar se describen en el epígrafe 2.1.5.

Tabla 10 Lista de chequeo para evaluar el análisis OLAP

Peso	Indicadores a evaluar	Eval	NP	Cantidad de elementos afectados	Comentarios
Estructura del documento					
crítico	1. ¿Los entregables contienen las secciones obligatorias de la plantilla estándar definidas para un expediente de proyecto? (Portada, Control de Versiones, Reglas de Confidencialidad, Tabla de Contenidos y Contenido) (ver Expediente de Proyecto)				
Indicadores definidos por la etapa					
	1. ¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?				

Peso	Indicadores a evaluar	Eval	NP	Cantidad de elementos afectados	Comentarios
	2. ¿Se realizó una interfaz amigable para hacer consultas?				
crítico	3. ¿Los reportes son configurables a través de la interfaz del sistema?				
	4. ¿El rendimiento de los reportes no se afecta cuando el número de dimensiones del modelo se incrementa?				
	5. ¿Presenta la capacidad de crear todo tipo de dimensiones con funcionalidades aplicables de una dimensión a otra?				
	6. ¿La interfaz está orientada a facilitar el uso de las funciones del sistema por parte de los usuarios?				
crítico	7. ¿No existen restricciones para construir cubos OLAP con dimensiones y niveles de agregación ilimitados?				
crítico	8. ¿Los usuarios son capaces de manipular los resultados de manera que se ajusten a sus necesidades, conformando nuevos reportes?				
	9. ¿El sistema responde de una forma rápida y veraz a la información que le sea solicitada por el usuario?				
crítico	10. ¿El sistema refleja cualquier lógica del negocio para poder responder a preguntas específicas?				
crítico	11. ¿El sistema garantiza la confidencialidad y seguridad de acceso a los datos por rol de los usuarios?				
	12. ¿Los datos e información derivados del proceso de análisis realizado mediante la aplicación, apoyan la toma de decisiones en la Institución?				
crítico	13. ¿Los cambios en los datos se reflejan automáticamente en los reportes de forma instantánea?				
Semántica del documento					
crítico	1. ¿Se han identificado errores ortográficos en los entregables o en los modelos diseñados?				
crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?				
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?				

Caso práctico: Se evaluó el análisis OLAP utilizando la lista de chequeo, y no se identificó ningún indicador con dificultad. El proceso es evaluado de "Bien" ya que no hubo ningún indicador crítico evaluado de mal y no existieron problemas con los formatos de las plantillas. Además no se encontraron errores ortográficos en los documentos revisados.

2.4. Conclusiones parciales

- ✓ Se definió un procedimiento por etapas de desarrollo que facilita la implementación de soluciones BI en los ensayos clínicos que se gestionan en el CIM; tomando como bases las metodologías que sustentaron la propuesta a partir del análisis realizado en el Capítulo 1.
- ✓ Se obtuvieron tres listas de chequeo para evaluar el diseño, la implementación de ETL y el análisis de los datos de un datawarehouse realizado en el Centro de Inmunología Molecular. Las cuales están avaladas por el Centro Nacional de Software CALISOFT.
- ✓ Se logró la implementación de un datawarehouse para los ensayos clínicos del producto hR3, el cual contará con información histórica, íntegra y estándar de todos los ensayos conducidos.
- ✓ Se obtuvieron análisis estadísticos sobre eficacia y seguridad del producto hR3 desarrollado en el CIM, para el tratamiento del cáncer en diferentes localizaciones.
- ✓ Se preparó al personal del centro cliente en la aplicación del procedimiento.

Capítulo 3: Evaluación del procedimiento BI para los EC del CIM

En este capítulo se aplica el método seleccionado para la evaluación del procedimiento BI para los EC del CIM, a través de la técnica Delphi mediante criterio de expertos; detallándose los resultados arrojados y los cuestionarios utilizados para obtener los criterios de los encuestados.

3.1. Aplicación del método Delphi

El método consta de cuatro pasos fundamentales, las cuales se exponen a continuación y se pueden ver representados en dos fases como se muestra en la figura 12:

- ✓ Elección de expertos.
- ✓ Elaboración del cuestionario para validación de la propuesta.
- ✓ Establecimiento de la concordancia de los expertos mediante el uso del coeficiente de Kendall.
- ✓ Desarrollo práctico y explotación de resultados.

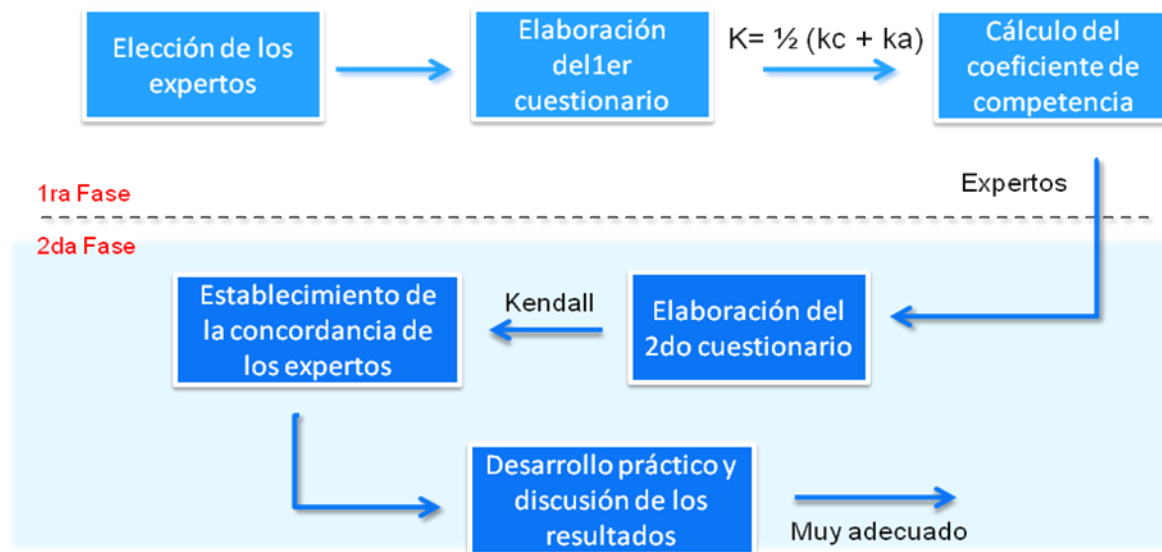


Figura 12 Fases del método Delphi

3.1.1. Elección de los expertos

Los expertos fueron seleccionados teniendo en cuenta el cumplimiento de los siguientes criterios: graduado de nivel superior, un año de experiencia como mínimo, vinculación al desarrollo de productos informáticos, conocimientos sobre el desarrollo de un datawarehouse y sobre alguna de

las tres etapas de desarrollo de una solución BI; así como presentar una noción básica sobre el proceso de gestión de ensayos clínicos.

Para aplicar este método no existe una norma generalizada para determinar el número de experto que podrán intervenir en el proceso de validación. Se recomienda que dicho método tenga entre siete y 30 expertos debido a que con siete el error se disminuye exponencialmente y con más de 30, aunque siga disminuyendo, es de forma poco significativa y no compensa el incremento de costo y esfuerzo. Por lo que se seleccionaron 15 expertos del Centro de Tecnologías de Datos (DATEC) de la Universidad de las Ciencias Informáticas, para que formaran parte del panel de expertos. De estos se seleccionaron siete expertos por etapa (alguno de ellos experto en más de una etapa) para un total de 21 encuestas a analizar. A las personas seleccionadas se les aplicó una encuesta (ver anexo 1) con el objetivo de determinar sus coeficientes de competencia y recopilar información detallada sobre la labor que desempeñan, los años de experiencia en el tema, la calificación profesional y la categoría docente. Al obtener los resultados de la encuesta aplicada al panel de experto, se procedió con el cálculo del coeficiente de competencia. Este coeficiente se determina mediante la fórmula: $K = \frac{1}{2}(k_c + k_a)$, donde k_c es el coeficiente de conocimientos y k_a es el coeficiente de argumentación.

El coeficiente de conocimiento (k_c) se obtiene de la primera tabla de la encuesta, la cual recoge una autoevaluación del posible experto. El posible experto marcará en la casilla enumerada según su criterio acerca de la capacidad que tiene sobre el tema que se ha sometido a su consideración, en una escala del uno al diez; considerando uno como no tener ningún conocimiento y diez el de pleno conocimiento de la problemática tratada. Después para ajustarla a la teoría de las probabilidades se multiplica por 0.1 el valor de la casilla seleccionada. El coeficiente de argumentación (k_a) se calcula de la siguiente forma: se utiliza la tabla correspondiente a la segunda pregunta del cuestionario. El experto debe marcar, según su criterio, su grado de competencia sobre los aspectos sometidos a su consideración. Esta selección se traduce a puntos, según la siguiente escala (tabla 11):

Tabla 11 Grados de influencia en la determinación del coeficiente de argumentación

Fuentes de argumentación	Grado de influencia de cada una de las fuentes en sus criterios		
	Alto	Medio	Bajo
Análisis teóricos realizados por usted	0.3	0.2	0.1

Fuentes de argumentación	Grado de influencia de cada una de las fuentes en sus criterios		
	Alto	Medio	Bajo
Su experiencia obtenida	0.5	0.4	0.2
Trabajos de autores nacionales	0.05	0.05	0.05
Trabajos de autores extranjeros	0.05	0.05	0.05
Su propio conocimiento del estado del problema en el extranjero	0.05	0.05	0.05
Su intuición	0.05	0.05	0.05

El coeficiente de argumentación (k_a) será igual a la suma de los valores donde el posible experto haya marcado. Al tener calculado k_c y k_a ya se cuenta con los datos suficientes para obtener el coeficiente de competencia K , el cual es analizado según los siguientes intervalos para determinar el nivel del coeficiente de competencia (K):

Si $0,8 < K < 1,0$ el coeficiente de competencia es alto.

Si $0,5 < K < 0,8$ el coeficiente de competencia es medio.

Si $K < 0,5$ el coeficiente de competencia es bajo.

Los expertos seleccionados para formar parte del panel para la validación de la propuesta tienen que ser aquellos cuyos resultados arrojaron un coeficiente de competencia alto o medio. De los 15 expertos a los que se les aplicó la encuesta de autoevaluación, todos arrojaron un coeficiente de competencia alto o medio por lo que fueron seleccionados para continuar con la ejecución del método (ver anexo 2), los resultados se muestran en la tabla 12 que se presenta a continuación:

Tabla 12 Coeficiente de competencia de los expertos

Experto	Competencia	Experto	Competencia
E1	Medio (Diseño) Alto (Análisis de datos)	E9	Alto (ETL)
E2	Medio (Diseño)	E10	Medio (ETL)
E3	Medio (Diseño)	E11	Alto (ETL), Medio (Análisis de datos)
E4	Alto (Diseño y Análisis de datos)	E12	Alto (ETL)
E5	Medio (Diseño y Análisis de datos)	E13	Medio (ETL)
E6	Alto (En las tres etapas)	E14	Alto (Análisis de datos)
E7	Alto (Diseño)	E15	Alto (Análisis de datos)
E8	Medio (ETL)		

3.1.2. Elaboración del cuestionario para la validación de la propuesta

Cuando se culmina con la selección del panel de expertos, se prosigue con la elaboración de la encuesta para la evaluación del procedimiento. Para ello se conforma un cuestionario el cual no debe tener demasiadas preguntas, pero el mismo debe abordar aspectos medulares sobre la investigación realizada. Para elaborar la encuesta se identificó la variable a evaluar la cual se operacionalizó para encontrar los indicadores que permitieran evaluarla. Los cuestionarios fueron creados de forma tal que las respuestas fueran categorizadas en Muy adecuado (C1), Bastante adecuado (C2), Adecuado (C3), Poco adecuado (C4) y No adecuado (C5) (ver anexo 3). El objetivo que se persiguió fue conocer si el procedimiento propuesto facilitaba la implementación de soluciones BI en el CIM, para ello se verificó que el procedimiento propuesto se caracterizara por:

1. Tener en cuenta las características del negocio analizado.
2. Estar comprendido como un ciclo completo BI que presenta las tres etapas: diseño, ETL y análisis del DWH. Cada etapa del procedimiento debe contener una secuencia lógica de pasos que se correspondan con las actividades comprendidas en cada etapa según la teoría. Estos pasos deben estar descrito de forma precisa, consistente y sin ambigüedades.
3. Presentar un artefacto que permita lograr la correspondencia entre las perspectivas de análisis y los datos fuentes.
4. Definir dentro de los elementos de la arquitectura de un DWH las áreas temporales y los metadatos.
5. Tener en cuenta la evaluación de la solución BI (o de algunos pasos que lo requiera) y presentar un artefacto que lo permita.

A continuación, en la tabla 13, se muestran los valores porcentuales para cada indicador del total de respuestas analizadas, por encuestas, de las preguntas relacionadas con ese indicador. Se puede apreciar que todos los indicadores son evaluados de Muy adecuado en más de un 50 por ciento.

Tabla 13 Resultados porcentuales de las encuestas

Indicador	Muy adecuado	Bastante adecuado	Adecuado	Poco adecuado	No adecuado	Cantidad de preguntas relacionadas con los indicadores			Total respuestas
						Diseño	ETL	OLAP	

Indicador	Muy adecuado	Bastante adecuado	Adecuado	Poco adecuado	No adecuado	Cantidad de preguntas relacionadas con los indicadores			Total respuestas
						Diseño	ETL	OLAP	
1	78.6%	21.4%	-	-	-	2	1	3	42
2	70%	30%	-	-	-	3	3	4	70
3	71.4%	28.6%	-	-	-	1	1	0	14
4	57.1%	42.9%	-	-	-	0	1	0	7
5	64.3%	35.7%	-	-	-	2	1	1	28

3.1.3. Establecimiento de la concordancia de los expertos

Para que la propuesta tenga una mayor validez es necesario que exista un adecuado acuerdo entre los integrantes del panel de expertos, esto se comprueba mediante el cálculo del coeficiente de concordancia de Kendall, el cual ayuda a precisar el grado de coincidencia de las valoraciones arrojadas por los expertos. Para determinar este coeficiente se construye una tabla de aspectos a evaluar contra expertos, estos datos fueron obtenidos al aplicar las encuestas por etapas para la validación del procedimiento. Luego de haber confeccionado dicha tabla se siguen algunos pasos para lograr calcular Chi cuadrado real, que se compara con el de las tablas estadísticas de Siegel, y si: $X^2_{real} < X^2(\alpha, N - 1)$ existe una concordancia entre el panel de expertos. A continuación se presenta la comparación de los resultados (tabla 14):

Tabla 14 Comparación de Chi cuadrado real

α	N-1	$X^2_{real} < X^2(\alpha, N - 1)$	Etapas BI
0.01	11	0.7546 < 33.4087	Diseño
0.01	8	0.2688 < 31.3078	ETL
0.01	11	0.7546 < 33.4087	Análisis de datos

Por lo que se demuestra que existe concordancia entre el panel de experto confeccionado, para validar cada etapa del procedimiento propuesto.

3.1.4. Desarrollo práctico y explotación de los resultados

Los expertos que integraron el panel recibieron un resumen de la propuesta de solución para responder la encuesta de evaluación del procedimiento. Se realizó una sola ronda de preguntas y al recibir los resultados se prosiguió a analizarlos. Para la recopilación y visualización de los resultados aportados por el panel de expertos se confeccionaron tablas utilizando el Microsoft Excel 2007.

Luego de tener todos los datos almacenados en las tablas de frecuencias acumuladas, se siguen una serie de pasos para obtener los demás resultados:

Paso 1: Se construye una tabla de frecuencias absolutas acumuladas donde cada número en la fila, excepto el primero, se obtiene sumándole el anterior. En estas tablas desaparece la columna del total que se observa en las de frecuencias acumuladas.

Paso 2: Se construye la tabla de frecuencias relativas acumulativas, la cual se obtiene dividiendo por la cantidad de expertos, cada uno de los números de las tablas de frecuencias absolutas acumuladas obtenidas en el paso anterior. En esta tabla se elimina una columna pues hay cinco categorías de evaluación y solo se necesitan cuatro puntos de corte para obtener los cinco intervalos.

Paso 3: Se crea una nueva tabla de puntos de corte a partir de la anterior, a la cual se le adicionan columnas y filas para colocar los resultados que se mencionan a continuación: suma de las columnas, suma de filas, y promedio de las columnas.

Los promedios de las filas se obtienen de forma similar, en este caso también se divide por cuatro porque quedan cuatro categorías ya que la última se eliminó. Para hallar N, se divide la suma de las sumas entre el resultado de multiplicar el número de indicadores por el número de preguntas. El valor N-P da el valor promedio que otorgan los expertos para cada indicador propuesto. A continuación se muestra las tablas para cada etapa, que resume lo expuesto en los puntos anteriores (ver tablas 15, 16 y 17):

Tabla 15 Tabla de puntos de cortes (diseño)

Puntos de corte							N =	2.29	Grado de adecuación
No	Elementos	C1	C2	C3	C4	Suma	P	N-P	
1	A1	3.72	3.72	3.72	3.72	14.88	3.72	-1.43	Muy adecuado
2	A2.1	1.07	3.72	3.72	3.72	12.22	3.06	-0.77	Muy adecuado
3	A2.2	-0.57	3.72	3.72	3.72	10.59	2.65	-0.36	Muy adecuado
4	A2.3	-0.57	0.57	3.72	3.72	7.44	1.86	0.43	Muy adecuado
5	A3.1	3.72	3.72	3.72	3.72	14.88	3.72	-1.43	Muy adecuado
6	A3.2	-0.18	0.18	3.72	3.72	7.44	1.86	0.43	Muy adecuado
7	A3.3	1.07	1.07	3.72	3.72	9.57	2.39	-0.11	Muy adecuado
8	A4	0.57	3.72	3.72	3.72	11.72	2.93	-0.64	Muy adecuado
9	A5	3.72	3.72	3.72	3.72	14.88	3.72	-1.43	Muy adecuado

Puntos de corte							N =	2.29	Grado de adecuación
No	Elementos	C1	C2	C3	C4	Suma	P	N-P	
10	A6	0.57	3.72	3.72	3.72	11.72	2.93	-0.64	Muy adecuado
11	A7	-0.57	3.72	3.72	3.72	10.59	2.65	-0.36	Muy adecuado
12	A8	0.18	3.72	3.72	3.72	11.34	2.83	-0.55	Muy adecuado
Suma		12.73	35.28	44.63	44.63	137.27	Total Muy adecuado		12
Puntos de corte		1.06	2.94	3.72	3.72		Total Bastante adecuado		0

Tabla 16 Tabla de puntos de cortes (ETL)

Puntos de corte							N =	2.56	Grado de adecuación
No	Elementos	C1	C2	C3	C4	Suma	P	N-P	
1	A1	3.72	3.72	3.72	3.72	14.88	3.72	-1.16	Muy adecuado
2	A2.1	1.07	3.72	3.72	3.72	12.22	3.06	-0.49	Muy adecuado
3	A2.2	1.07	3.72	3.72	3.72	12.22	3.06	-0.49	Muy adecuado
4	A2.3	0.18	3.72	3.72	3.72	11.34	2.83	-0.27	Muy adecuado
5	A3	3.72	3.72	3.72	3.72	14.88	3.72	-1.16	Muy adecuado
6	A4	0.18	3.72	3.72	3.72	11.34	2.83	-0.27	Muy adecuado
7	A5	0.18	3.72	3.72	3.72	11.34	2.83	-0.27	Muy adecuado
8	A6	3.72	3.72	3.72	3.72	14.88	3.72	-1.16	Muy adecuado
9	A7	1.07	3.72	3.72	3.72	12.22	3.06	-0.49	Muy adecuado
Suma		14.90	33.47	33.47	33.47	115.31	Total Muy adecuado		9
Puntos de corte		1.66	3.72	3.72	3.72		Total Bastante adecuado		0

Tabla 17 Tabla de puntos de cortes (análisis de datos)

Puntos de corte							N =	2.29	Grado de adecuación
No	Elementos	C1	C2	C3	C4	Suma	P	N-P	
1	A1	3.72	3.72	3.72	3.72	14.88	3.72	-1.39	Muy adecuado
2	A2.1	1.07	3.72	3.72	3.72	12.22	3.06	-0.73	Muy adecuado
3	A2.2	-0.18	1.07	3.72	3.72	8.33	2.08	0.25	Muy adecuado
4	A2.3	0.57	3.72	3.72	3.72	11.72	2.93	-0.60	Muy adecuado
5	A3.1	0.57	3.72	3.72	3.72	11.72	2.93	-0.60	Muy adecuado
6	A3.2	-0.18	3.72	3.72	3.72	10.98	2.74	-0.42	Muy adecuado
7	A3.3	1.07	3.72	3.72	3.72	12.22	3.06	-0.73	Muy adecuado
8	A4	0.57	3.72	3.72	3.72	11.72	2.93	-0.60	Muy adecuado
9	A5	1.07	3.72	3.72	3.72	12.22	3.06	-0.73	Muy adecuado
10	A6	1.07	3.72	3.72	3.72	12.22	3.06	-0.73	Muy adecuado

Puntos de corte							N =	2.29	Grado de adecuación
No	Elementos	C1	C2	C3	C4	Suma	P	N-P	
11	A7	0.18	3.72	3.72	3.72	11.34	2.83	-0.51	Muy adecuado
12	A8	-1.07	3.72	3.72	3.72	10.09	2.52	-0.19	Muy adecuado
Suma		8.44	41.98	44.63	44.63	137.27	Total Muy adecuado		12
Puntos de corte		0.70	3.50	3.72	3.72		Total Bastante adecuado		0

Las sumas obtenidas en las cuatro primeras columnas dan los puntos de corte, los cuales se utilizan para determinar la categoría o grado de adecuación de cada criterio según la opinión de los expertos consultados.

Por los resultados arrojados por las tablas anteriores se puede llegar a la conclusión que el procedimiento sometido a evaluación por el panel de expertos resultó "Muy adecuado" en sus tres etapas.

3.2. Conclusiones parciales

- ✓ A partir de la evaluación realizada en el trabajo con los expertos el procedimiento propuesto fue catalogado de Muy adecuado.
- ✓ Se cuenta con un aval emitido por el Centro de Inmunología Molecular, donde se describe como satisfactorio y de vital importancia el trabajo realizado, en aras de facilitar el análisis de los datos y con ello la toma de decisiones en el departamento de la dirección de investigaciones clínicas. Demostrando su utilidad práctica para continuar con su implementación en otros ensayos clínicos que se gestionen en el CIM.

Conclusiones

El estudio del desarrollo de soluciones BI y las propuestas metodológicas que existen, proporcionaron la elaboración del presente trabajo, cuyo resultado fue una propuesta de procedimiento para el desarrollo de un sistema de inteligencia de negocios para los ensayos clínicos que se gestionan en el Centro de Inmunología Molecular. La investigación contribuyó a la integración y análisis de los datos de la institución y arrojó los siguientes resultados:

- ✓ Se definió un procedimiento por etapas de desarrollo, para facilitar el proceso de inteligencia de negocios de los ensayos clínicos que se gestionan en el Centro de Inmunología Molecular, a partir de lo mejor y más aplicable de las metodologías de desarrollo BI.
- ✓ Se aplicó el procedimiento de inteligencia de negocios desarrollado, a partir de los datos de los ensayos clínicos del producto Nimotuzumab (hR3), que se gestiona en el Centro de Inmunología Molecular.
- ✓ Se evaluó de Muy adecuado el procedimiento propuesto a través del método de expertos.
- ✓ Se obtuvo un aval del centro cliente donde se valora de satisfactorio el trabajo realizado al facilitar el proceso BI a los especialistas de la institución, sirviendo de guía en el desarrollo de la solución.

Recomendaciones

- ✓ Utilizar el procedimiento propuesto para guiar las etapas de desarrollo de inteligencia de negocios si se desea agregar nuevos datos de ensayos clínicos de los productos del Centro de Inmunología Molecular al datawarehouse; así como para otros proyectos que desarrollen soluciones de este tipo.
- ✓ Implementar el proceso BI a partir de los resultados que se obtendrán del sistema alasClínicas una vez desplegado, haciendo uso del procedimiento propuesto.
- ✓ Incluirle a la propuesta de procedimiento los pasos para realizar análisis de datos en los ensayos clínicos que se gestionan en el Centro de Inmunología Molecular, aplicando técnicas de minería de datos.

Referencias Bibliográficas

1. Transforming Knowledge Into Action. *Artículos BI en la Práctica*. [En línea] 29 de marzo de 2010. [Citado el: 19 de marzo de 2010.] <http://www.siskle.com/spanish/articulo01.html>
2. *Business Intelligence y Data Warehousing en Windows*. 2005, Danysoft.
3. **Brito Sarasa, Ing. Raycos**. *Minería de Datos aplicada a la Gestión Docente del Instituto Superior Politécnico José Antonio Echeverría*. CUJAE. 2008. Tesis de maestría.
4. **Hernández López, Ing. Asnioby**. *Almacenes de Datos aplicada a la Seguridad Ciudadana*. UCI. 2009. Tesis de maestría.
5. *Una metodología para sectorizar pacientes en el consumo de medicamentos aplicando datamart y datamining en un hospital*. **Tapia Rivas, Ivan, Ruiz Rivera, María y Ruiz L, Edgar**. 2007. <http://revistas.concytec.gob.pe>.
6. III Taller Internacional de Inteligencia Empresarial y Gestión del Conocimiento en la Empresa. *Sistema DataWarehouse Comercial de la Corporación CIMEX*. [En línea] 2002. [Citado el: 19 de marzo de 2010.] <http://www.redciencia.cu/empres/Intempres2002/Seminarios/CIMEXSAponencia.pdf>.
7. **Ramón Cueto, Ariagna y Díaz García, Joannis**. *Implementación de un Data Warehouse para el control del Recurso Humano de la Salud*. UCI. 2009. Tesis.
8. **González García, Alejandro**. *Evento II Taller Nacional de Gestión del Conocimiento en Energía*. [Documento pdf] 2005. Desarrollo metodológico para la implementación del proceso de datawarehousing en Cubaenergía.
9. *Propuesta de Sistema de gestión y control de energía en la Universidad de Ciencias Informáticas*. **Falcón Márquez, Odalys Rosa**. 2008, RevistaCiencia.com.
10. Clinical Trials. *Understanding Clinical Trials*. [En línea] National Institutes of Health, 2007. [Citado el: 19 de marzo de 2010.] <http://clinicaltrials.gov/ct2/info/understand>
11. Centro Nacional Coordinador de Ensayos Clínicos (CENCEC). *Experiencias*. [En línea] 2009. [Citado: 19 de marzo 2009.] <http://www.cencec.sld.cu/pgs/resultados.htm>
12. **Ricardo Darío, Bernabeu**. *DATA WAREHOUSING: Investigación y Sistematización de Conceptos-HEFESTO: Metodología propia para la construcción de un Datawarehouse*. Córdoba, Argentina: s.n., 2009. pág. 146.

13. **Trujillo, Juan; Soler, Emilio; Zubcoff, José; Mazon, José Norberto; Glorio, Octavio y Pardillo, Jesús.** *Desarrollo de almacenes de datos dirigido por modelos.* España, Universidad de Alicante, 2007.
14. **Kimball, Ralph y Roos, Margy.** *The Data Warehouse Lifecycle Toolkit.* 2da edición. Canadá : Wiley Publishing, 2002. ISBN: 0471200247.
15. **Río, Luis Méndez del.** *Más allá del Business Intelligence: 16 experiencias de éxito.* Barcelona : s.n., 2006. pág. 219. 978-84-96612-10-5.
16. **Abelló, Alberto, Rollón, Emma y Rodríguez, M. Elena.** *Diseño y administración de bases de datos.* Barcelona : s.n., 2006. 84-8301-880-8.
17. **Heredía, Mariano y de Almeida, Rodrigo.** *BI-FLOSS: Business Intelligence - Free/Libre Open Source Software.* 2008.
18. **Trujillo, Juan C. y Palomar, Manuel.** *Uso y Diseño de Bases de Datos Multidimensionales y Almacenes de Datos.* [Documento pdf] España : s.n., 2002.
19. **English, Larry.** Business Intelligence Defined. [En línea] 2005. [Citado el: 31 de marzo de 2010.] <http://www.b-eye-network.com/view/1119>.
20. **Davis, Jim, Miller, Gloria J. y Russell, Allan.** *La revolución de la información: cómo utilizar el modelo de evolución de la información para que su empresa crezca.* Barcelona : s.n., 2006. 978-84-936084-1-5.
21. *Assessing benefits of business intelligence systems – a case study.* **Borut Hocevar, Jurij Jaklic, Hocevar, Borut y Jaklic, Jurij.** 2010.
22. *Conceptual model of business value of business intelligence systems.* **Popovic, Ales, Turk, Tomas y Kaklic, Jurij.** 2010.
23. **Serrano, Manuel.** *Data Warehouses: Almacenamiento y Recuperación de la información.* 2010.
24. **Talaya, Águeda Esteban.** *Principio de marketing.* 2006. pág. 791.
25. **Peralta, Verónica.** *Un caso de estudio sobre diseño lógico de Data Warehouses.* [Documento pdf] Uruguay : s.n., 2001.
26. **Wang, Jon.** *Encyclopedia of data warehousing and mining.* New York : s.n., 2009. 9781605660103.

27. **Pelegrín Tamayo, Neyaris y Casaña Vinagera, Virgen.** *Proceso de la migración de datos hacia un Data Warehouse para el módulo Análisis Químico del proyecto LIMS Control de Calidad.* Universidad de las Ciencias Informáticas. 2009. Tesis.
28. **Zorrilla, Marta.** *Data warehouse y OLAP.* [Documento pdf] 2008.
29. Centro de Inmunología Molecular. [En línea] [Citado el: 31 de marzo de 2010.] <http://www.cim.co.cu/>.
30. **Suárez Martínez, G. y Morejón Zayas, O. M.** *Informe periódico de seguridad Nimotuzumab.* CIMAB. Ciudad de la Habana, Cuba: s.n., 2010. pág. 34.
31. Transforming Knowledge into action! *Artículos de BI en la práctica: Bill Inmon - Ralph Kimball.* [En línea] 2010. [Citado el: 2 de abril de 2010.] <http://kle.sisinfomanagement.com/spanish/articulo04.html>
32. **Mirabella, Pablo y Sánchez, Jesús.** *DataWarehouse.* [Documento word] Caracas, Venezuela : s.n., 2009.
33. **Lujan Mora, Sergio.** *Data Warehouse Design with UML* Tesis (Doctorado). España, Universidad de Alicante, 2005.
34. **Espinosa Milla, Roberto.** *El Rincón del BI.* Descubriendo el Business Intelligence [En línea] 2010. [Citado: 31 de marzo 2010.] <http://churriwifi.wordpress.com/2009/12/05/5-fases-en-la-implantacion-de-un-sistema-dw-metodologia-para-la-construccion-de-un-dw>
35. *El método Delphi: cuando dos cabezas piensan más que una en el desarrollo de guías de práctica clínica.* **Carreño Jaimes, Marisol Carreño Jaimes.** Colombia : s.n., 2009, Revista Colombiana de Psiquiatría. 0034-7450.
36. **Aedo, Raúl Rubén Fernández.** 2005. *Modelo Informático para la autogestión del aprendizaje para la universalización de la enseñanza.* [Documento PDF] Granada, España: s.n., 2005.
37. **Aragón, Salvador.** 2003. Business school. *El Método Delphi.* [En línea] 2003. [Citado el: 6 de abril de 2010.] http://profesores.ie.edu/salvador_aragon/Documentacion/2003-M%E9todo%20Delphi.pdf
38. **Colunga, Silvia, G,A.** 2003. *La Psicología Educativa, su objeto, métodos y problemas principales.* Universidad de Camagüey: s.n., 2003.
39. **Santelices, L.C. Á. M. D.** 2004. *Experimentos virtuales para la enseñanza del Electromagnetismo* Universidad de Camagüey. 2004.

40. **Díaz Morales, Themis Patricia y Bermúdez Rodríguez, José Salvador.** *Diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular.* UCI, 2010. Tesis.
41. **Rodríguez Sotolongo, Javier y Peralta Góngora, Yohan Orlando.** *Implementación del proceso de extracción, transformación y carga de un Datawarehouse para los Ensayos Clínicos del Centro de Inmunología Molecular.* UCI, 2010. Tesis.
42. **Simón Mir, Yailín y Iñiguez Bermúdez Góngora, Yoander.** *Análisis de datos a un Datawarehouse para los Ensayos Clínicos del Centro de Inmunología Molecular, aplicando técnicas OLAP.* UCI, 2010.
43. **Connolly, Thomas M., Begg, Carolyn E. y Wesl, Addison.** *Sistemas de Bases de Datos, un enfoque práctico para diseño, implementación y gestión.* 2005.
44. **Piattini, M., y otros.** *Tecnología y Diseño de Bases de Datos.* 2006. pág. 980. 8478977333.
45. **Walter Reynoso, Luis, Marcelo Algorry, Aldo y Assandri, Marina.** *SQLMax Connections. Data Warehousing.* [En línea] 2001. [Citado el: 7 de abril de 2010.] <http://www.sqlmax.com/dataw1.asp>
46. **Nader, Ing. Javier.** *Sistema de apoyo gerencial universitario.* Argentina : s.n., 2004. Tesis de maestría.
47. Grupo eGlu Bi. *Documentos Transcritos.* [En línea] 2009. [Citado el: 7 de abril de 2010.] http://www.slideshare.net/bernabeu_dario/grupo-eglu-bi
48. **Peña Rivera, Juan David y Suárez Daza, Jesús A.** *Minería de Datos aplicada a la Gestión Docente del Instituto Superior Politécnico José Antonio Echeverría.* 2005. Tesis.

Bibliografía

- **GTIC.** Grupo de Tecnologías de la Información y las Comunicaciones. *SSR - ETSI Telecomunicación - UPM.* [En línea] 2007. <http://www.gtic.ssr.upm.es/encuestas/delphi.htm>
- Corporate Information Factory. [En línea] Inmon Consulting Services, 2007. <http://www.inmoncif.com/home>
- **Huamantumba, Rayner.** *Datamart paso a paso* [En línea] 2010. <https://www.u-cursos.cl/ingenieria/2008/1/IN830/1/.../11313>
- **Berzal Galiano, Fernando.** Curso de doctorado “Extracción del Conocimiento”: *Introducción al Datawarehousing.* Universidad de Granada, España.
- **Mendez, A., Martínez, A., Britos, P. y García-Martínez, R.** *Fundamentos de Data Warehouse.* Instituto Tecnológico de Buenos Aires, Argentina.
- **De la caridad Suárez Giró, Milagros y González Sardina, José Raúl.** Realizar el diseño e implementación de un datawarehouse para el proyecto LIMS Control de Calidad. Trabajo de diploma. Ciudad de La Habana. Universidad de las Ciencias Informáticas, 2009.
- **López Díaz, Aidacelys y Rodríguez García, Lucía.** *Sistema de Manejo de Datos de Ensayos Clínicos.* Trabajo de diploma. Ciudad de La Habana. Universidad de las Ciencias Informáticas, 2007.
- Soler, E., Trujillo, J., & Fernández-Med, E. (2006). Conferencia: Actas del Taller sobre Desarrollo de Software Dirigido por Modelos. MDA y Aplicaciones. *Un conjunto de transformaciones QVT para el modelado de almacenes de datos seguros* . España.
- **Services, U.S. Department of Health and Human.** *Guidance for Industry. Computerized Systems Used in Clinical Investigations.* 2007. <http://www.fda.gov/cder/guidance/index.htm>.
- Leon Monzón, K. (2006) “*Propuesta de proyecto de colaboracion con la UCI*”, Departamento de Ensayos Clínicos, Centro de Inmunología Molecular, Ciudad de la Habana, Cuba
- *Diseño de protocolos de un estudio clínico : las denominadas "case report forms" o cuadernos de recogida de datos.* **Aznar Salatti, J.** España : s.n., 1996, Jano.
- **CECMED.** Regulación No. 45-2007. *Requerimientos para la notificación y el reporte de eventos adversos graves e inesperados en los Ensayos Clínicos.* Ciudad de la Habana : s.n., 2007.

- **Pentaho Corporation.** Pentaho. *Pentaho Data Integration (Kettle)*. [En línea] 2005-2010. <http://kettle.pentaho.com/>.
- **Talend.** Talend: open integration solutions. *The Open Source Data Management Company*. [En línea] 2006-2010. <http://es.talend.com/index.php>.
- **Marrero, Ivette.** Business Intelligence from cuban perspective. [En línea] 2009. <http://bi-topics.blogspot.com/>.
- **BARCAS Consultores.** *Metodología ASAP*. [Documento pdf] Febrero de 2006.
- **Durán, Martín.** *Business Intelligence y SaaS: herramientas para una gestión más eficiente en las Pymes*. Madrid : CEPREDE: Centro de Predicción Económica, 2010.
- **Florea, Alejandro.** *Inteligencia de negocio: Motor de las decisiones económicas de un país. Análisis del impacto de las herramientas de inteligencia de negocios en la SHCP de México*. México : s.n., 2006.
- **Grupo SAP.** *Historia de éxito SAP. Asegurando el crecimiento. Grupo Farma* [Documento pdf] España : s.n., 2006.
- **Grupo SAP.** *Soluciones SAP® para pequeñas y medianas empresas. A la medida de sus necesidades. Diseñadas para hacerle crecer*. [Documento pdf] España : s.n., 2007.
- **Planeaux, David y Daniel, Alvin.** *Informe Ejecutivo de Oracle: Oracle Business Intelligence Standard Edition One*. [Documento pdf] Estados Unidos : s.n., 2007.
- **Grupo Sinnexus.** *Sinnexus: Soluciones de Business Intelligence*. [Documento pdf] España : s.n., 2006.
- **Joyanes Aguilar, Luis.** *La Bioinformática como convergencia de la Biotecnología y la Informática*. [Documento pdf] Madrid : s.n., 2003.
- **Hernández León, Rolando Alfredo y Coello González, Sayda.** *El paradigma cuantitativo de la investigación científica*. Ciudad de la Habana : Universitaria, 2008. pág. 115. 978-959-16-0343-2.
- **Ramos Chagoya, Ena.** *Métodos y técnicas de investigación*. [En línea] 2008. <http://www.gestiopolis.com>.
- **Sistemas de Información orientados a la toma de decisiones: el enfoque multidimensional.** **Dapena Bosquet, Isabel, Muñoz San Roque, Antonio y Sánchez Miralles, Álvaro.** III, España : s.n., 2005, Anales de Mecánica y Electricidad, Vol. LXXXII, págs. 18-23. 0003-2506.

Anexos

Anexo # 1: Encuesta de autovaloración para determinar el coeficiente de competencia de los expertos

Compañero (a):

En la ejecución de la presente investigación, se desea someter a la valoración de un grupo de especialistas, la propuesta de procedimiento para el desarrollo de soluciones BI para los ensayos clínicos que se gestionan en el Centro de Inmunología Molecular con el objetivo de evaluar las principales actividades definidas. Para ello se necesita conocer el grado de dominio que usted posee acerca del tema de investigación y con ese fin se desea que responda lo que se le pide a continuación:

Nombre (s) y Apellidos:

Centro de trabajo:

Labor que realiza:

Años de experiencia:

Calificación profesional: Ingeniero ___ Licenciado en educación ___ Máster ___ Doctor ___

Categoría docente: Prof. Instructor___ Prof. Asistente___ Prof. Auxiliar___ Prof. Titular___ Prof. Adjunto___

1. Seleccione en una escala del 1 al 10 el valor que corresponda con el grado de conocimientos que usted posee acerca del tema de investigación que se desarrolla (desarrollo de soluciones BI), considerando 1 como no tener ningún conocimiento y 10 el de pleno conocimiento de la problemática tratada.

1	2	3	4	5	6	7	8	9	10

2. Valore el grado de influencia que cada una de las fuentes que se le presenta a continuación ha tenido en su conocimiento y criterios sobre el tema que se investiga.

Fuentes de argumentación	Grado de influencia de cada una de las fuentes en sus		
	Alto	Medio	Bajo
Análisis teóricos realizados por usted			

Fuentes de argumentación	Grado de influencia de cada una de las fuentes en sus		
	Alto	Medio	Bajo
Su experiencia obtenida			
Trabajos de autores nacionales			
Trabajos de autores extranjeros			
Su propio conocimiento del estado del problema en el extranjero			
Su intuición			

Anexo # 2: Listado de expertos

Experto	Nombre y apellidos	Competencia
Experto 1 (E1)	Ing. Yunier Santana Aldana	Diseño y Análisis de datos
Experto 2 (E2)	Ing. Yanisbel González Hernández	Diseño
Experto 3 (E3)	Ing. Mabel Medina Rodríguez	Diseño
Experto 4 (E4)	Ing. Eidy Marien Carbo Peña	Diseño y Análisis de datos
Experto 5 (E5)	Ing. Madelys Cuesta Villa	Diseño y Análisis de datos
Experto 6 (E6)	MSc. Asnioby Hernández López	Diseño, ETL y Análisis de datos
Experto 7 (E7)	Ing. Alfonso Claro Arceo	Diseño
Experto 8 (E8)	Ing. Daulemys Rigo Portillo	ETL
Experto 9 (E9)	Ing. Doris Medina Mustelier	ETL
Experto 10 (E10)	Ing. Esley Leon Valdes	ETL
Experto 11 (E11)	Ing. Osniel Hernandez Calvo	ETL y Análisis de datos
Experto 12 (E12)	Ing. Persy Morell Guerra	ETL
Experto 13 (E13)	Ing. Raul Martinez Aguirre	ETL
Experto 14 (E14)	Ing. Roberto Téllez Ibarra	Análisis de datos
Experto 15 (E15)	Ing. Anisley Delfino Rodríguez	Análisis de datos

Anexo # 3: Encuestas a expertos en diseño de DWH, proceso ETL y análisis de datos para la validación del procedimiento

Compañero (a):

La presente encuesta forma parte de la aplicación del método de valoración de expertos. Con este fin se solicita su valiosa colaboración para evaluar si las etapas y actividades; así como los artefactos de entrada y salida que se propusieron son correctos y facilitan la implementación de soluciones de inteligencia de negocios en el Centro de Inmunología Molecular. Para lograr este objetivo se han elaborado un conjunto de preguntas que permiten medir la efectividad del modelo. De antemano se le asegura que nadie podrá saber quién es el encuestado y además se garantiza que sus opiniones se tendrán en cuenta para la posterior aplicación del procedimiento. Valore el grado de factibilidad de cada pregunta de acuerdo a la siguiente escala: Muy adecuado (C1), Bastante adecuado (C2), Adecuado (C3), Poco adecuado (C4), No adecuado (C5)

Preguntas para expertos en diseño de DWH	Criterio del Experto				
	C1	C2	C3	C4	C5
1. La utilidad de un procedimiento para aplicarlo a la etapa de diseño de un datawarehouse con el objetivo de lograr modelar la integración de los datos de los ensayos clínicos que se gestionan en el Centro de Inmunología Molecular es:					
2. Se proponen tres actividades fundamentales para lograr la identificación de las necesidades del negocio, en una aproximación basada en objetivos. Categorice cada uno de ellos teniendo en cuenta el tipo de información que se maneja en el CIM:					
2.1 Identificar objetivos decisionales.					
2.2 Identificar objetivos informacionales.					
2.3 Identificar variables informacionales.					
3. Los siguientes aspectos forman parte del procedimiento. Categorice cada una de ellos:					
3.1 Las etapas propuestas					
3.2 Los pasos dentro de cada una de las etapas					
3.3 Los artefactos propuestos para cada una de las etapas					
4. La identificación de las reglas del negocio para modelar la integración de los datos de los ensayos clínicos que se gestionan en el Centro de Inmunología Molecular es:					
5. El diccionario de datos definido en el procedimiento para establecer la correspondencia entre las perspectivas y las variables de los datos fuentes proporcionadas es:					

Preguntas para expertos en diseño de DWH	Criterio del Experto				
	C1	C2	C3	C4	C5
6. La evaluación del modelo conceptual y el diccionario de datos en la etapa de desarrollo del modelo conceptual del procedimiento es:					
7. La lista de chequeo propuesta para la evaluación del diseño del DWH es:					
8. En sentido general emita su criterio acerca de la redacción de cada uno de los pasos dentro del procedimiento. Tenga en cuenta si la especificación de los pasos es clara, específica y sin ambigüedades:					

Preguntas para expertos en el proceso ETL	Criterio del experto				
	C1	C2	C3	C4	C5
1. La utilidad de un procedimiento para aplicarlo a la etapa de ETL de un DWH con el objetivo de lograr la integración de los datos de los EC que se gestionan en el CIM es:					
2. Los siguientes aspectos forman parte del procedimiento. Categorice cada uno de ellos:					
2.1 Las etapas propuestas.					
2.2 Los pasos dentro de cada una de las etapas.					
2.3 Los artefactos propuestos para cada una de las etapas.					
3. La creación de transformaciones a partir de las reglas del negocio, para lograr la integración de los datos de los EC que se gestionan en el CIM es:					
4. La actualización del diccionario de datos en el procedimiento, para establecer la correspondencia entre las perspectivas, tipos de variables de los datos fuentes y el formato en que se encuentra la información es:					
5. La definición de un área temporal y el uso de los metadatos en el diseño de la arquitectura es:					
6. La lista de chequeo propuesta para la evaluación del proceso de ETL es:					
7. En sentido general emita su criterio acerca de la redacción de cada uno de los pasos dentro del procedimiento. Tenga en cuenta si la especificación de los pasos es clara, específica y sin ambigüedades:					

Preguntas para expertos en Análisis de datos	Criterio del Experto				
	C1	C2	C3	C4	C5

Preguntas para expertos en Análisis de datos	Criterio del Experto				
	C1	C2	C3	C4	C5
1. La utilidad de un procedimiento para aplicarlo a la etapa de análisis de un datawarehouse con el objetivo de lograr el análisis multidimensional de los datos de los ensayos clínicos que se gestionan en el Centro de Inmunología Molecular es:					
2. Se proponen tres actividades que permiten asegurar la confidencialidad de los datos. Categorice cada uno de ellos:					
2.1 Definición del nivel de acceso.					
2.2 Definición de los tipos de reporte.					
2.3 Construcción de una interfaz de usuario.					
3. Los siguientes aspectos forman parte del procedimiento. Categorice cada una de ellos:					
3.1 Las etapas propuestas					
3.2 Los pasos dentro de cada una de las etapas					
3.3 Los artefactos propuestos para cada una de las etapas					
4. La creación de una interfaz de usuario es:					
5. Tener en cuenta para el desarrollo de los cubos OLAP las necesidades de la organización identificadas, así como los indicadores y perspectivas de análisis es:					
6. Los roles definidos para el acceso y realización del análisis de los datos en la fase definición de las restricciones del análisis dentro del procedimiento propuesto es:					
7. La lista de chequeo propuesta para la evaluación del análisis OLAP es:					
8. En sentido general emita su criterio acerca de la redacción de cada uno de los pasos dentro del procedimiento. Tenga en cuenta si la especificación de los pasos es clara, específica y sin ambigüedades:					

Glosario de Términos

A continuación se presentan los términos que podrían resultar de difícil comprensión, nuevos al lector o de diversos significados dependiendo del contexto que se analice. Esta sección tiene como objetivo facilitar la comprensión del contenido expuesto en el documento.

Atributo: se utiliza en las tablas de dimensiones, para búsquedas, filtrado o clasificación de los hechos.

CIM: Centro de Inmunología Molecular. Centro biotecnológico y científico - investigativo asociado a la producción. Creado en Cuba el 5 de diciembre de 1994.

Diccionario de datos: artefacto que se utiliza para definir la correspondencia que existe entre las variables de las tablas, bases de datos o modelos determinados con sus perspectivas.

Dimensión: característica de un hecho que permite su análisis posterior en el proceso de toma de decisiones y brinda una perspectiva adicional a un hecho dado.

DWH: Datawarehouse. Tecnología orientada a la toma de decisiones al depurar, integrar y almacenar los datos en un solo destino, después de haberlos gestionado de diversos formatos y orígenes. Colección de datos orientada al negocio, integrada, variante en el tiempo y no volátil para el soporte del proceso de toma de decisiones de la gerencia.

Data marts: son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones.

EA: Eventos Adversos. Reacciones negativas que manifiestan los pacientes durante el ensayo. Algunos tipos de EA son: cefalea, dolor de cabeza o migraña, fiebre, diarrea, vómitos.

EC: Ensayos Clínicos. Cualquier investigación en seres humanos dirigida a descubrir o verificar los efectos clínicos, farmacológicos u otros efectos farmacodinámicos de un producto en investigación.

QT: tratamiento al cáncer con quimioterapia.

RT: tratamiento al cáncer con radioterapia.