

Universidad de las Ciencias Informáticas

Facultad 6



Título: Diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el
Centro de Inmunología Molecular

Trabajo de Diploma para optar por el título de
Ingeniero en Ciencias Informáticas

Autores: Themis Patricia Díaz Morales

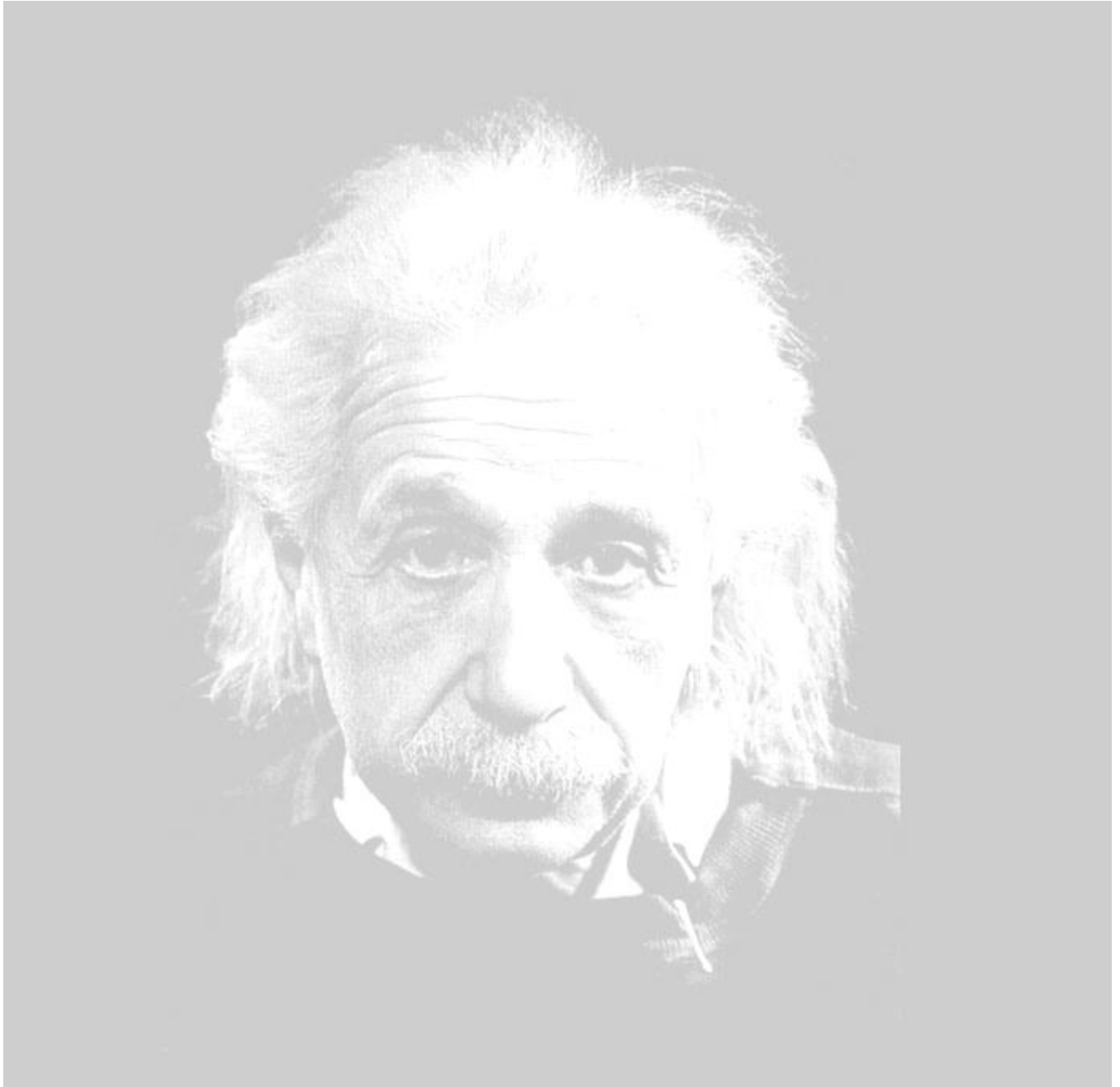
José Salvador Bermúdez Rodríguez

Tutores: Ing. Martha Denia Hernández Ramírez

Ing. Anthony Rafael Sotolongo León

Co-tutora: MSc. Sandy Henríquez Villafruela

Junio de 2010



Si supiese qué es lo que estoy haciendo, no lo llamaría investigación...
Albert Einstein

DECLARACIÓN DE AUTORÍA

Declaramos ser autores de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmo la presente a los ____ días del mes de _____ del año _____.

Themis Patricia Díaz Morales

José Salvador Bermúdez Rodríguez

Firma del Autor

Firma del Autor

Ing. Martha Denia Hernández Ramírez

Ing. Anthony Rafael Sotolongo León

Firma del Tutor

Firma del Tutor

MSc. Sandy Henríquez Villafruela

Firma de la Co-tutora

DATOS DE CONTACTO

Tutores:

Tutor: Ing. Martha Denia Hernández Ramírez
Especialidad de graduación: Ingeniería en Ciencias Informáticas
Categoría docente: Instructor en Adiestramiento
Categoría Científica: Ingeniero
Años de experiencia en el tema: 0
Años de graduado: 1
Correo Electrónico: mdhernandez@uci.cu

Tutor: Ing. Anthony Rafael Sotolongo León
Especialidad de graduación: Ingeniería Informática
Categoría docente: Instructor
Categoría Científica: Ingeniero
Años de experiencia en el tema: 0
Años de graduado: 3
Correo Electrónico: asotolongo@uci.cu

Co-tutora:

Co-tutora: MSc. Sandy Henríquez Villafruela
Especialidad de graduación: Licenciada en Lengua Inglesa
Categoría docente: Asistente
Categoría Científica: Máster
Años de experiencia en el tema: 0
Años de graduado: 15
Correo Electrónico: sandyh@uci.cu

-A Fidel Castro y a la Revolución por la increíble oportunidad de estudiar en esta universidad de excelencia y por tanta confianza depositada en nosotros.

-A nuestros padres, por nutrirnos día a día con todo el amor, el apoyo y la dedicación que un hijo puede desear. Por ser quienes en todos estos años se han mantenido a nuestro lado, estimulándonos al esfuerzo continuo por alcanzar siempre metas más altas.

-A nuestras familias, por ser la raíz de este árbol que hoy florece y mañana dará nuevos frutos. Por esperar siempre lo mejor de nosotros y confiar en nuestras decisiones aunque no siempre fueran las mejores.

-A Martica, Sandy y Anthony por su eterna paciencia y comprensión. Por tener siempre una respuesta para todas nuestras preguntas. Por el cariño que se han ganado y lo que significan hoy para nosotros.

-A todos los profesores que contribuyeron a nuestra formación como ingenieros, especialmente a Mariño, Yanet, Nara, Edgar, Olga, Yuraisy y Elennis. Por cada duda aclarada y cada consejo brindado.

-A los compañeros de los eternos 6106 y 6506, que a lo largo de todos estos años cada mañana al

despertar y cada noche al finalizar el día han estado ahí junto a nosotros. Por estar siempre al tanto de nuestras vidas, en lo personal y profesional. Por tantos recuerdos imborrables.

-A todos aquellos amigos que mañana serán parte insustituible de nuestros mejores años como estudiantes y siempre tuvieron presente nuestro desempeño en esta etapa que no vuelve atrás: Eli, Sury, Ludmi, Kaly, Yayi, Ailón, Anita, Mary, Susy, Katia, Alberto, Diana, Karen, Chuchi, Chela, Clau, Yili, Charly, Nelson, Fito, Ramsés, Javi (el jerry), Cuan y José Augusto.

-A todos los que de una forma u otra han aportado en nuestra preparación como jóvenes comprometidos con la Revolución para hacer este sueño realidad: Rosy, Yuniór, Zoili, Bertha, Mercedes, Mileydis y Yoannis.

-A todas las trabajadoras del departamento de estadísticas del Centro de Inmunología Molecular, en especial a Patricia Luaces, Patricia Piedra y Carmen, por las horas dedicadas en aras de lograr los objetivos propuestos.

-Y más que nada, agradecemos a Dios, por todo cuanto tenemos y somos.

Themis Patricia y José Salvador

-A mi mamá, por ser a cada instante más que una madre, por ser mi amiga, mi hermana, mi consejera. Por todo el cariño que día a día me regala, por todos sus esfuerzos por lograr mi desarrollo profesional, por guiarme en la vida y encontrar siempre en ella todo el apoyo y el cariño que necesito para seguir adelante.

-A mi abuelita, por dedicar parte de su vida a ayudarme a crecer, a complacerme en mis caprichos y entregarme todo su amor sin pedir nada a cambio. Por siempre estar ahí para mí, con ese cariño incomparable que no tiene límites. Por haberme brindado junto a mami un hogar lindo, lleno de amor y comprensión.

- A mi padre, que aunque se ha ido y no está hoy entre nosotros, me acompaña en cada momento de mi vida ayudándome a ser la persona que soy, donde quiera que se encuentre espero esté orgulloso de mí, a él gracias por entregarme el tesoro más grande, por ser un ejemplo a seguir y enseñarme que no existen límites para lograr un sueño.

-A mis tíos y primas: LunaK, Victorio, Mayra, Lázaro, Mayvic, Jessica y Araya, por hacerme sentir siempre parte de la familia.

-A Salvi, por ser más que un compañero, mi amigo, mi hermano durante estos 5 largos años, por el cariño y el amor que le tengo, por el lugar que ocupa hoy en mi corazón.

- A mis amigos y compañeras de estudio en estos inolvidables años que quedarán como parte de mi vida a Carly, Ailón, Anita, la piny, Geo, Roberto y Asencio por brindarme su amistad, sus consejos y por su apoyo en los momentos más difíciles.

- A los que ya no están, por cuidarme, amarme y desear un día este sueño hecho realidad.

Themis Patricia Díaz Morales

-A mis padres, por guiarme y aconsejarme siempre. Por la confianza depositada cuando creyeron que esta no era la carrera correcta. Por todo el amor brindado hasta hoy. Por estar en cada momento de felicidad y tristeza. Por ser simplemente, los mejores padres del mundo.

-A Tita, que hoy no está pero estoy seguro estuviera orgullosa de mí en estos momentos.

- A toda mi familia, especialmente a mi tía Ramona, mi tía Concha, Ame, Aya, Tata, Yanet, Lianet, Alci, Ferman, Saikil y Nena por demostrarme que son parte fundamental de mi existir. Por apoyarme y estar siempre al tanto de todos los pasos que doy en la vida.

- A Themura, por ser mi compañera, mi amiga, mi hermana, pero por sobre todas las cosas mi eterno dúo. Gracias por estar desde el primer día, hasta hoy y para siempre.

- A todos los que han llegado para quedarse a lo largo de estos años y ya forman parte de mi vida: Mary, Susy, Katia, Geo, Dieter, Eilys, Chela y Clau.

José Salvador Bermúdez Rodríguez

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). El principal objetivo de este último es gestionar, almacenar y analizar toda la información que se recoge en los Ensayos Clínicos (EC) una vez que se aplica un producto determinado a pacientes que padecen de cáncer. En el presente Trabajo de Diploma se realiza el diseño de un Datawarehouse (DWH) para los EC que se gestionan en el CIM con el propósito de almacenar gran cantidad de información y con ello viabilizar la integración de los datos de la institución que posteriormente serán analizados. Al estudiar las metodologías que existen para el diseño de un DWH, se observa que han sido creadas con el fin de ser aplicadas en sectores productivos, financieros y de servicios. Conociendo las peculiaridades de la gestión de la información en el CIM, se elabora un procedimiento para la etapa de diseño tomando lo mejor y más aplicable de cada metodología analizada. Para lograr la evaluación del diseño, se propone una lista de chequeo ya que no existe en la bibliografía consultada una técnica de evaluación para esta etapa.

PALABRAS CLAVES:

Centro de Inmunología Molecular, Datawarehouse, Diseño, Ensayos Clínicos, Procedimiento.

TABLA DE CONTENIDO

INTRODUCCIÓN 1

CAPÍTULO 1: FUNDAMENTOS TEÓRICOS SOBRE EL DESARROLLO DE UN DATAWAREHOUSE4

1.1 Introducción4

1.2 Ensayos clínicos4

 1.2.1 ¿Qué es un Ensayo Clínico?4

 1.2.2 Gestión de los Ensayos Clínicos en el Centro de Inmunología Molecular.....4

1.3 Datawarehouse.....7

 1.3.1 Revisión conceptual.....7

 1.3.2 Características principales8

 1.3.3 Ventajas del uso de un Datawarehouse12

 1.3.4 Tendencias Actuales12

 1.3.5 Modelo Multidimensional.....13

1.4 Metodologías para el diseño de un DWH.....14

 1.4.1 Metodología Hefesto.....15

 1.4.2 Ciclo de vida Kimball16

 1.4.3 Desarrollo de almacenes de datos dirigidos por modelos (Trujillo)17

 1.4.4 Data Warehouse Engineering Process (DWEPE)18

 1.4.5 Rapid Warehousing Methodology (RWM)18

1.5 Etapa de diseño de un DWH20

 1.5.1 Modelo conceptual.....20

 1.5.2 Modelo lógico20

 1.5.3 Evaluación del diseño de un DWH.....22

1.6 Lenguaje de modelado unificado23

1.7 Herramientas de modelado.....23

1.8 Técnicas de captura de requisitos24

1.9 Conclusiones.....25

CAPÍTULO 2: DISEÑO DE UN DATAWAREHOUSE PARA LOS ENSAYOS CLÍNICOS QUE SE GESTIONAN EN EL CENTRO DE INMUNOLOGÍA MOLECULAR26

2.1 Introducción.....26

2.2 Propuesta de procedimiento26

 2.2.1 Análisis del negocio27

 2.2.2 Especificación de las necesidades28

 2.2.3 Desarrollo del modelo conceptual29

 2.2.4 Desarrollo del modelo lógico.....30

2.3 Aplicación del procedimiento33

 2.3.1 Análisis del negocio33

 2.3.2 Especificación de las necesidades37

2.3.3 Desarrollo del modelo conceptual	40
2.3.4 Desarrollo del modelo lógico.....	45
2.4 Conclusiones	50
CAPÍTULO 3: EVALUACIÓN DEL DISEÑO REALIZADO	52
3.1 Introducción.....	52
3.2 ¿Por qué lista de chequeo?.....	52
3.2.1 ¿Qué es una lista de chequeo?.....	52
3.3 Elaboración y evaluación de la lista de chequeo	52
3.4 Evaluación de la etapa de diseño a través de la lista de chequeo	55
3.5 Conclusiones	59
CONCLUSIONES	60
RECOMENDACIONES	61
REFERENCIAS BIBLIOGRÁFICAS	62
BIBLIOGRAFÍA	64
ANEXOS	66

INTRODUCCIÓN

Siempre ha sido de gran interés para las personas mejorar su calidad de vida y para ello se hace indispensable la búsqueda de nuevos tratamientos para diversas enfermedades como el cáncer, una de las principales causas de muerte, a la que se enfrenta toda la humanidad. A pesar de los esfuerzos realizados por miles de especialistas en todo el mundo, todavía se hace difícil encontrar la cura para padecimientos como estos que ponen en peligro la existencia de la raza humana.

Según la Organización Mundial de la Salud (OMS) el cáncer es: *"...un proceso de crecimiento y diseminación incontrolados de células. Puede aparecer prácticamente en cualquier lugar del cuerpo. Suele invadir el tejido circundante y puede provocar metástasis en puntos distantes del organismo..."* [1]. Otros términos utilizados son neoplasias y tumores malignos. Una de las características que define el cáncer es la generación rápida de células anormales que crecen más allá de sus límites normales y pueden invadir zonas adyacentes del organismo o diseminarse a otros órganos en un proceso que da lugar a la formación de las llamadas metástasis. Cada año aparecen en el planeta unos 10 millones de casos nuevos y mueren por esta causa 7 millones de personas. Se espera que para el 2020 aumenten a 15 millones los casos nuevos, el 60% de ellos en países en desarrollo, que cuentan sólo con el 5% de los recursos destinados a combatir este mal [2].

Existen en el mundo numerosos centros de investigaciones científicas, los cuales se dedican a la creación de fármacos para la prevención y/o tratamiento de este tipo de enfermedades que hasta hoy no tienen cura. Aunque la inmensa mayoría de estos centros se encuentra en países con un alto nivel de desarrollo, Cuba no está exenta de todas estas investigaciones para tratar de propiciarles una vida más estable y duradera a los pacientes que padecen de tan agobiante enfermedad. Para lograr esta premisa, en Cuba se han creado después del triunfo de la Revolución varios centros biotecnológicos y científicos - investigativos asociados a la producción; dentro de los cuales se encuentra el Centro de Inmunología Molecular (CIM), creado el 5 de diciembre de 1994 [3].

Este centro desde sus inicios se ha dedicado al desarrollo de biomoléculas y otros fármacos para el tratamiento de diferentes enfermedades relacionadas con el sistema inmunológico, principalmente el cáncer. Para ello, se realizan los análisis pertinentes con el objetivo de probar el funcionamiento de los fármacos y luego se aplican en pacientes que padecen diferentes enfermedades. Para la aprobación, prueba e introducción de dichos productos en el mercado se realizan 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..."* [4]. Estos presentan un protocolo,

documento que establece la razón de ser del estudio, sus objetivos, diseño, métodos y el análisis previsto de sus resultados, así como las condiciones bajo las que se realizará y desarrollará el estudio.

Actualmente, el CIM se encuentra trabajando con 28 productos diferentes que abarcan 62 EC, algunos multinacionales con pronóstico de inclusión de 2300 pacientes durante el 2009, además de seguir con el tratamiento a más de 1500 incluidos durante el año 2008. Los EC se encuentran involucrados en más de 500 hospitales y en casi todas las provincias del país y los ensayos post registros, fases IV de farmacovigilancia se encuentran en marcha en 13 de las 14 provincias del país [5].

Los EC que se gestionan en los diferentes centros involucrados, llevan asociados un gran volumen de documentación. Toda esta información se almacena en bases de datos las cuales se pueden encontrar en diferentes formatos. Estas bases de datos al no encontrarse integradas ni estandarizadas tornan engorroso el proceso para el manejo de la información por parte de los directivos de la institución y dificultan la realización de análisis estadísticos complejos dentro de un mismo EC o entre diferentes EC, corriendo el riesgo de que se pierda información útil y valiosa con el decursar del tiempo e impidiendo una adecuada integración de los datos que contribuya a elevar la efectividad del tratamiento de la información. Por toda la situación anteriormente descrita, se plantea como **problema científico**: ¿cómo modelar la integración de los datos de los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular?

La investigación tiene como **objeto de estudio** el proceso de desarrollo de un Datawarehouse, enmarcado en el **campo de acción** la etapa de diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular.

El **objetivo general** de este trabajo es diseñar un Datawarehouse que contribuya a la integración de los datos de los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular.

En correspondencia con el objetivo general, se plantean como **objetivos específicos**:

- ✓ Elaborar un procedimiento para la etapa de diseño de un Datawarehouse que integre los datos de los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular.
- ✓ Aplicar el procedimiento en la etapa de diseño de un Datawarehouse para integrar los datos de los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular.
- ✓ Elaborar una lista de chequeo para evaluar el diseño realizado.
- ✓ Evaluar el diseño realizado a partir de la aplicación de la lista de chequeo.

Para el cumplimiento de estos objetivos se realizaron esencialmente las siguientes **tareas investigativas**:

- ✓ Identificación de la tipología de los datos de los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular.
- ✓ 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 cada una de las metodologías de desarrollo existentes para el diseño de un DWH.
- ✓ Elaboración del procedimiento para la etapa de diseño de un Datawarehouse que integre los datos de los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular.
- ✓ Identificación de las necesidades del negocio.
- ✓ Realización del modelo conceptual para el diseño del Datawarehouse.
- ✓ Evaluación del modelo conceptual del Datawarehouse.
- ✓ Realización del modelo lógico para el diseño del Datawarehouse.
- ✓ Validación del procedimiento propuesto a través del método Delphi.
- ✓ Elaboración de una lista de chequeo para evaluar el diseño realizado.
- ✓ Evaluación del diseño realizado a partir de la aplicación de la lista de chequeo.

El Trabajo de Diploma está estructurado de la siguiente manera: introducción, tres capítulos, conclusiones, recomendaciones, referencias bibliográficas, bibliografía, anexos y glosario de términos.

Capítulo 1: Fundamentos teóricos sobre el desarrollo de un Datawarehouse

En este capítulo se abordan definiciones y conceptos importantes sobre DWH y la gestión de los EC en el CIM. Asimismo, se hace un estudio bibliográfico de las metodologías existentes a nivel mundial y en Cuba para el desarrollo de un DWH, con una explicación detallada de todos los aspectos a tener en cuenta en la etapa de diseño. Se fundamenta el uso de las distintas herramientas y tecnologías utilizadas y se hace una comparación con otras muy difundidas en el contexto internacional.

Capítulo 2: Diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular

En este capítulo se realiza el diseño de un DWH para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular, se propone y aplica un procedimiento para esta etapa explicando paso a paso la propuesta.

Capítulo 3: Evaluación del diseño realizado

En este capítulo se elabora y aplica una lista de chequeo con el objetivo de evaluar el diseño realizado.

CAPÍTULO 1: FUNDAMENTOS TEÓRICOS SOBRE EL DESARROLLO DE UN DATAWAREHOUSE

1.1 Introducción

En este capítulo se abordan definiciones y conceptos importantes sobre DWH y la gestión de los EC en el CIM. Asimismo, se hace un estudio bibliográfico de las metodologías existentes a nivel mundial y en Cuba para el desarrollo de un DWH, con una explicación detallada de todos los aspectos a tener en cuenta en la etapa de diseño. Se fundamenta el uso de las distintas herramientas y tecnologías utilizadas y se hace una comparación con otras muy difundidas en el contexto internacional.

1.2 Ensayos clínicos

1.2.1 ¿Qué es un Ensayo Clínico?

Un EC es un tipo de estudio clínico en el que se evalúan la eficacia o seguridad de nuevos fármacos o tratamientos médicos a través de la aplicación a seres humanos con un protocolo de investigación estrictamente controlado. Los EC permiten a los médicos determinar si un nuevo tratamiento, medicamento o dispositivo; contribuirá a prevenir, detectar o tratar una enfermedad.

Un EC medicamentoso es toda evaluación experimental de una sustancia o medicamento, a través de su administración o aplicación a seres humanos, orientada a alguno de los siguientes fines:

- ✓ Poner de manifiesto sus efectos farmacodinámicos o recoger datos referentes a su absorción, distribución, metabolismo y excreción en el organismo humano.
- ✓ Establecer la eficacia para una indicación terapéutica, profiláctica o diagnóstica determinada.
- ✓ Conocer el perfil de sus reacciones adversas y establecer su seguridad [6].

Cuando se habla de un ensayo clínico, es importante conocer el término localización, lugar del cuerpo donde está concentrado el tumor y sobre el cual se aplican los fármacos que se crean en el centro.

1.2.2 Gestión de los Ensayos Clínicos en el Centro de Inmunología Molecular

Uno de los fármacos que se desarrollan en el CIM y sobre el cual se realizan los EC es el Nimotuzumab, anticuerpo monoclonal recombinante contra el Receptor del Factor de Crecimiento Epidérmico (EGFR, siglas en inglés). Nimotuzumab 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. Nimotuzumab recibe alternativamente las denominaciones comerciales de CIMAher®, TheraCIM®-hR3, Theraloc®, BIOMAb®-EGFR, YMB 1000 y VECTHIX®. Internacionalmente, se encuentra en Canadá, la

India, China y Alemania; de ahí la importancia que representa para el CIM garantizar la seguridad y eficacia de este producto.

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. Hasta la fecha, Nimotuzumab ha sido administrado a más de 1000 pacientes en ensayos clínicos terminados y en más de 5000 pacientes en estudios en curso.

Los eventos adversos más frecuentemente reportados con la administración de Nimotuzumab en ensayos clínicos incluyen escalofríos, fiebre, náuseas, vómitos, dolores musculares, cefalea y fatiga. En los ensayos en población pediátrica, los eventos adversos no graves, clasificados como posiblemente relacionados con Nimotuzumab son mareos, parestesias, somnolencia, fiebre, fatiga, eritema, erupción, leucopenia, trombocitopenia, anemia, hipocaliemia, incremento de la transaminasa glutámico pirúvica (TGP) y disminución de la hemoglobina.

A continuación se presenta una tabla con algunos de los EC relacionados al Nimotuzumab para tener una idea de la cantidad de pacientes que se enfrentan a este tratamiento (ver tabla 1) [7].

Número del estudio (país)	Indicación	Fase	Dosis de Nimotuzumab	Fármaco(s)	Número de sujetos tratados/dosis
IIC RD EC040 (Cuba)	Cáncer de Cabeza y Cuello	I/II	50, 100, 200 ó 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 ó 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	Cáncer de Cabeza y Cuello	I	200 mg	Nimotuzumab ± RT	10

(Cuba)					
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
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 ó 400 mg	Nimotuzumab	3/50 mg 3/100 mg 3/200 mg 3/400 mg

Tabla 1: Ensayos clínicos relacionados al Nimotuzumab

Toda la información que genera cada uno de estos ensayos (datos primarios, imágenes, Cuadernos de Recogida de Datos (CRD), aprobaciones y modificaciones), debe mantenerse almacenada durante 15 años con el objetivo de cumplir con las buenas prácticas clínicas exigidas por las agencias regulatorias a nivel mundial. Se puede afirmar que cada paciente genera aproximadamente un volumen de información de mil datos diferentes; desde información referida a mediciones cuantitativas del estado de un paciente (presión sanguínea, hemoglobina), hasta imágenes que caracterizan el tamaño del tumor, así como información referida al proceso de intercambio de mensajes entre los investigadores que conducen los estudios.

El proceso de monitorización de toda la información clínica (cuyo objetivo es controlar la calidad y uniformidad de los datos recogidos) se lleva a cabo mediante visitas periódicas a cada uno de los sitios que participan en esta actividad (que se refiere a todas las provincias involucradas). Luego de llenar los CRD en los hospitales, estos se guardan en bases de datos electrónicas para facilitar los resultados estadísticos que de ellos se derivan. En el CIM, la gestión de los EC se realiza a través del sistema EPIDATA, utilizado para la recopilación de la información. Este sistema genera reportes en diferentes formatos: SPSS, Excel, SASS y Text. Igualmente almacena información en otros formatos: .rec, .ges, .eix, .chk, .bak y .not. Otra característica importante es que a este sistema acceden diferentes especialistas, que la información está disgregada en diferentes modelos y que estos no presentan igual estructura de diseño. Al no encontrarse integradas, se torna engorroso el proceso para el manejo de la información por parte de los directivos de la institución, dificultando la realización de análisis estadísticos complejos dentro de un mismo EC o entre diferentes EC; corriendo el riesgo de que se pierda información útil y valiosa con el decursar del tiempo e impidiendo una adecuada integración de los datos que contribuya a elevar la efectividad del tratamiento de la información.

1.3 Datawarehouse

1.3.1 Revisión conceptual

En cualquier revisión que se realice sobre lo que se entiende por DWH, es difícil encontrar una concepción acabada y compartida por los autores, por el contrario, existen diversas aproximaciones teóricas; lo que demuestra que se trata de una herramienta en evolución y de compleja concepción. En la tabla 2, se muestran las concepciones de catorce autores [8].

ALMACÉN DE DATOS. CONCEPTOS.
(1) Depósito (o archivo) de la información reunida a partir de varias fuentes, guardada según un esquema unificado en un único lugar. Una vez reunidos, los datos se guardan durante un tiempo, lo que permite el acceso a datos históricos.
(2) Un almacén de datos es un repositorio de datos que almacena e integra información procedente de toda la organización, asegurando una gestión más eficiente de la misma y proyectando una visión única de la realidad de la compañía.
(3) 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.
(4) Un Datawarehouse es la colección de datos, organizados, integrados, historizados y disponibles para facilitar la toma de decisiones de usuarios finales.
(5) Un Datawarehouse es una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis.
(6) Colección de datos orientados a temas, integrada, variante en el tiempo, no volátil, que añade la geografía del dato.

(7) Depósito donde se almacenan los datos que la organización utiliza para saber cómo está funcionando. El almacenamiento de datos concentra mucha información proveniente de los procesos, de los sistemas operativos y financieros de los ERP y CRM entre otros.
(8) Almacenamiento efectivo, filtrado y ordenado de los datos estratégicos, tácticos y operativos, que permita su extracción y análisis de manera ordenada, clara, funcional y efectiva.
(9) Bodega donde están almacenados todos los datos necesarios para realizar las funciones de gestión de la empresa, de manera que puedan utilizarse fácilmente según se necesiten.
(10) Base de datos que almacena una gran cantidad de datos transaccionales integrados para ser usados para análisis gestionables por usuarios especializados (tomadores de decisión de la empresa).
(11) Colección de datos en la cual se encuentra integrada la información de la Institución y que se usa como soporte para el proceso de toma de decisiones gerenciales.
(12) Almacén de información temática orientado a cubrir las necesidades de aplicaciones de los sistemas de Soporte de Decisiones (DSS) y de la Información de Ejecutivos (EIS), que permite acceder a la información corporativa para la gestión, control y apoyo a la toma de decisiones.
(13) Almacenamiento de información homogénea y fiable, en una estructura basada en la consulta y el tratamiento jerarquizado de la misma, y en un entorno diferenciado de los sistemas operacionales.
(14) Herramienta que se nutre de las bases de datos de gestión y de otras externas que permite, con total flexibilidad y en tiempo real, obtener y combinar todo tipo de datos, indicadores, comparativas y simulaciones para la simple información, el conocimiento, el análisis y la toma de decisiones.

Tabla 2: Almacén de datos. Conceptos

1.3.2 Características principales

Las principales características que definen un DWH son [9]:

- ✓ Orientado por temas
- ✓ Integrados
- ✓ No volátiles
- ✓ Variables en el tiempo

Orientado por temas:

El diseño del DWH está orientado en cuanto a temas globales y por las principales áreas temáticas de la empresa que han sido definidas en el modelo de datos de alto nivel, como se muestra en la figura 1. Esta

La calidad contrasta con el clásico método que era orientado al proceso y funcionamiento de las aplicaciones utilizadas en sistemas operacionales más antiguos.

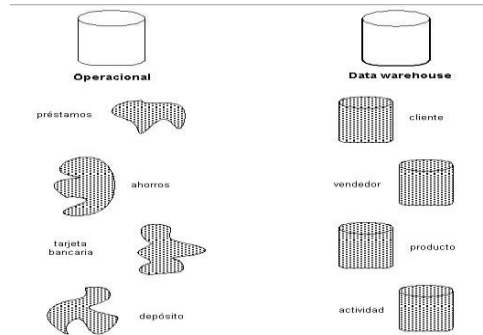


Figura 1: Orientado por temas

Integrados:

Los datos necesitan ser almacenados en el DWH de una forma globalmente aceptable y singular, aunque el programa operacional los almacene de una forma distinta (ver figura 2). Los datos deben de ser consistentes siempre dentro del DWH e integrados de distintas fuentes de datos operacionales. Por ejemplo:

Código: al introducir una información, ésta debe llegar al DWH de una forma consistente independientemente de cómo fue introducido en el programa de aplicación.

Medidas de atributos: independientemente de cuál sea la medida utilizada (por ejemplo: centímetros, metro, yardas, pulgadas), al ser almacenados los datos en el DWH deben tener una misma medida.

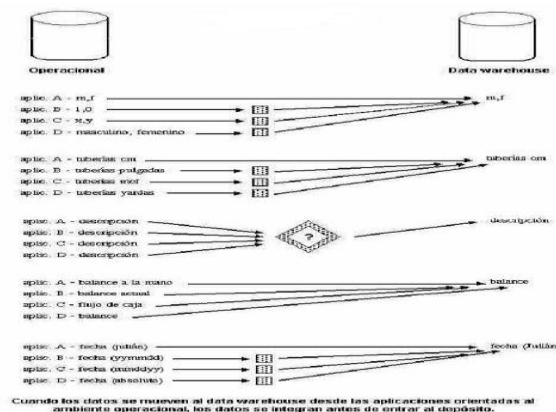


Figura 2: Integrados

No volátil:

La manipulación de datos en el DWH es mucho más simple. Sólo existen dos tipos de operaciones que se llevan a cabo en el DWH: cuando se cargan inicialmente los datos y cuando se acceden a él (ver figura 3). Los datos almacenados no se modifican ni actualizan nunca, solo se añaden nuevos datos.

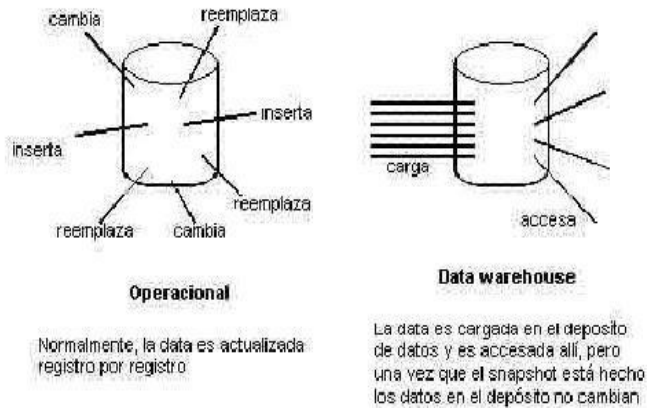


Figura 3: No volátil

Variables en el tiempo:

Las características de los datos contenidos en el DWH son distintas a aquellos en el ambiente operacional (ver figura 4). En el ambiente operacional, al momento de acceder a los datos estos son exactos a aquellos que se espera recibir. Sin embargo, en el DWH, los datos son exactos en algún momento del tiempo. Las variantes del tiempo se pueden notar de tres formas:

Límite de tiempo: el margen de tiempo del DWH es mucho mayor en cuanto a los datos (puede contener datos entre 5 y 10 años de almacenamiento). Por otro lado, en el ambiente operacional, el margen de tiempo de almacenamiento de los datos es mucho menor por lo que la información a la cual se accede es mucho más “fresca” (60- 90 días); ya que un programa de aplicación para trabajar eficientemente debe llevar la mínima cantidad de datos necesarios para realizar las transacciones.

Clave de estructura: los datos en el DWH contienen un elemento de tiempo (Ej.: día, semana, mes y año).

Actualizaciones: los datos una vez almacenados correctamente en el DWH no pueden ser alterados, por lo tanto no se pueden actualizar.

Los usuarios de un DWH necesitan acceder a los datos complejos, frecuentemente desde fuentes múltiples y de formas no predecibles.

Los usuarios que accedan a los datos operacionales, comúnmente efectúan tareas predefinidas que, generalmente requieren acceso a una sola base de datos de una aplicación. Por el contrario, los usuarios que accedan al DWH, efectúan tareas que requieren acceso a un conjunto de datos desde fuentes múltiples y frecuentemente no son predecibles. Lo único que se conoce (si es modelada correctamente) es el conjunto inicial de datos que se han establecido en el depósito.

1.3.3 Ventajas del uso de un Datawarehouse

En este epígrafe de la investigación se presentan algunas de las ventajas más sobresalientes que trae consigo la utilización de un DWH y que ejemplifican de mejor modo sus características y cualidades:

- ✓ Transforma datos orientados a las aplicaciones en información orientada a la toma de decisiones.
- ✓ Integra y consolida diferentes fuentes de datos en una única plataforma sólida y centralizada.
- ✓ Provee la capacidad de analizar y explotar toda la información que posee.
- ✓ Permite reaccionar rápidamente a los cambios del mercado.
- ✓ Aumenta la competitividad en el mercado.
- ✓ Mejora la entrega de información, es decir, información completa, correcta, consistente, oportuna y accesible.
- ✓ Aprovecha el enorme valor potencial de los recursos de información y los transforma en valor verdadero.
- ✓ Permite al usuario adquirir mayor confianza acerca de sus propias decisiones y de las del resto, y lograr así, un mayor entendimiento de los impactos ocasionados.
- ✓ Los usuarios pueden acceder directamente a la información en línea, lo que contribuye a su capacidad para operar con mayor efectividad en las tareas rutinarias o no.
- ✓ Los usuarios pueden tener a su disposición una gran cantidad de información multidimensional, presentada coherentemente como fuente única, confiable y disponible en sus estaciones de trabajo.

1.3.4 Tendencias Actuales

Los sistemas de DWH son el centro de la arquitectura de los sistemas de información de los 90's. Han surgido como respuesta a la problemática de extraer información sintética a partir de datos atómicos almacenados en bases de datos. Constituyen un recurso corporativo primario y parte importante del

patrimonio de la institución, manejando gran cantidad de datos en forma centralizada y manteniendo sus sistemas en línea. En este acápite, se pretende dar una panorámica de cómo se comporta, tanto en Cuba como en el mundo, la utilización de estos sistemas.

Existen muchas empresas que dependen del uso de los almacenes de datos para llevar a cabo un negocio de mercado. La compañía Wal-Mart, considerada la empresa más grande a nivel mundial cuenta con el DWH más voluminoso y poblado del mundo; el cual usa para tomar decisiones acerca de todos los procesos que realizan en el mercado internacional, elevar su economía y mantenerse en competencia respecto a otras compañías. Igualmente, Twentieth Century Fox utiliza la información relacionada con las películas que se proyectan en distintos lugares de los Estados Unidos para predecir qué actores, argumentos y filmes serán populares, con el objetivo de ganar audiencia en sus producciones.

El uso del DWH no sólo se aplica en áreas dentro del negocio, es aplicable al 100% de las áreas fuera de este. Se evidencia su utilización en Venezuela, aplicado al tema de la seguridad ciudadana; así como en hospitales de Perú para la sectorización de pacientes en el consumo de medicamentos.

Nuestro país no se encuentra exento de la aplicación de esta herramienta para la toma de decisiones. La empresa CIMEX, destacada por el crecimiento constante y la estabilidad financiera, tanto dentro como fuera del país utiliza un DWH para la gestión de inventarios. Además, en la Universidad de las Ciencias Informáticas se ha desarrollado un almacén de datos para la toma de decisiones en cuanto al consumo energético.

1.3.5 Modelo Multidimensional

El modelo multidimensional es la base de los almacenes de datos. Los DWHs con frecuencia almacenan información histórica y agregada, extraída de múltiples, heterogéneas, autónomas y distribuidas fuentes de información, por ello, la supervivencia de las organizaciones depende de la correcta gestión, seguridad y confidencialidad de la información.

Para el diseño de un DWH, en general, se emplea la representación de un modelo multidimensional. El modelado multidimensional es una técnica de diseño lógico que busca presentar los datos en un estándar y facilitar una recuperación adecuada de estos. Los datos son almacenados como hechos y dimensiones en un modelo de datos relacional. Para poder entender la definición presentada así como el modelo multidimensional, se deben comprender 3 conceptos fundamentales: hechos, dimensiones y atributos.

Se llama hecho a una operación que se realiza en el negocio la cual está estrechamente relacionada con el tiempo y es objeto de análisis para la toma de decisiones. También puede verse como un valor numérico que representa una actividad específica casi siempre con cifras que se suman entre sí.

Se conoce como dimensión a la 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. Son agrupaciones lógicas de atributos con un significado común y atómico. Por lo general son estables.

Por su parte, los atributos se utilizan en las tablas de dimensiones, para búsquedas, filtrado o clasificación de los hechos [10].

1.4 Metodologías para el diseño de un DWH

Para diseñar un Datawarehouse se deben analizar las opciones que se tienen en cuanto a las metodologías de arquitectura (Kimball e Inmon) y las metodologías de desarrollo existentes. En este epígrafe, se pretende hacer un análisis de ambos tipos de metodologías con el objetivo de seleccionar las más adecuadas para la arquitectura y el desarrollo del DWH.

Ralph Kimball (principal promotor del enfoque dimensional para el diseño de almacenes de datos), considera que un Datawarehouse es una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis.

Bill Inmon (conocido por muchos como el padre del Datawarehouse), plantea que un Datawarehouse es un conjunto de datos orientados por temas, integrados, variantes en el tiempo y no volátiles, que tienen por objetivo dar soporte a la toma de decisiones.

La metodología Kimball es la más aceptada en todo el mundo como la metodología más efectiva para desarrollar una solución de construcción de DWH. Apunta a una solución completa que se puede implementar en poco tiempo. Esto cumple 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.

La metodología Inmon por el contrario puede tener una implementación mucho más tardada, y es recomendada cuando se hace demasiado 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; si es necesario utilizar esta metodología, se recomienda hacerlo en iteraciones ya avanzadas y siempre empezar con Kimball [11].

De las dos metodologías de arquitectura analizadas, se decide utilizar Kimball por ser la más aceptada en todo el mundo como la metodología más efectiva para desarrollar una solución de construcción de DWH.

1.4.1 Metodología Hefesto

Hefesto es una metodología propia, cuya propuesta está fundamentada en una muy amplia investigación, comparación de metodologías existentes y experiencias propias en procesos de confección de almacenes de datos.

La metodología Hefesto puede resumirse a través de la siguiente figura (ver figura 5):



Figura 5: Metodología Hefesto

Esta metodología cuenta con las siguientes características:

- ✓ Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- ✓ Se basa en los requerimientos del usuario, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.
- ✓ Reduce la resistencia al cambio, ya que involucra al usuario final en cada etapa para que tome decisiones respecto al comportamiento y funciones del DWH.
- ✓ Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.

- ✓ Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- ✓ Es independiente de las herramientas que se utilicen para su implementación.
- ✓ Es independiente de las estructuras físicas que contengan el DWH y de su respectiva distribución.
- ✓ Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.
- ✓ Se aplica tanto para data mart como para DWH [12].

1.4.2 Ciclo de vida Kimball

A continuación se presenta una figura donde está el ciclo de vida concebido para el desarrollo de esta metodología (ver figura 6):

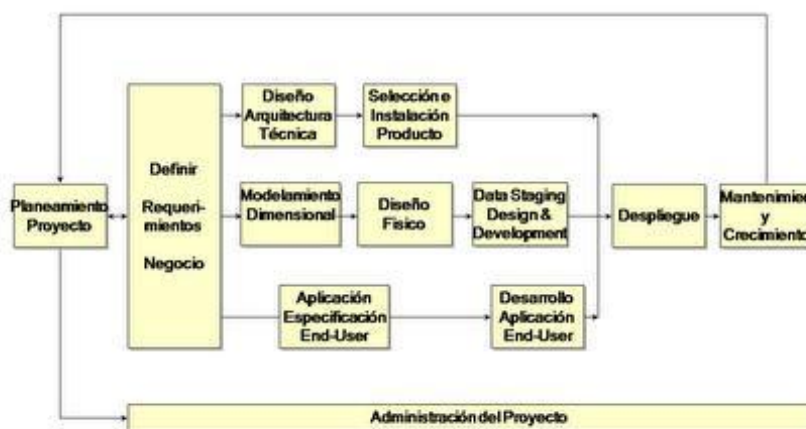


Figura 6: Ciclo de vida Kimball

El ciclo 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 con todo esto el plan de proyecto. En la segunda etapa de este ciclo se definen los requerimientos del negocio. Luego de definir los requerimientos del negocio, el proyecto se enfoca en tres líneas concurrentes: tecnología, datos y aplicaciones de la inteligencia de negocios (ver figura 6).

La etapa de diseño del DWH, está enmarcada en la línea de datos, donde se realiza el modelo dimensional y se analizan los datos del negocio para identificar la granularidad de las tablas de hechos, dimensiones y atributos asociados. Las construcciones primarias que se hacen en esta etapa son las tablas de dimensiones y de hechos. Las primeras contienen las métricas derivadas de los procesos de negocio o eventos. Hay que tener en cuenta que la granularidad debe ser lo más atómica posible, lo que

permite una mayor flexibilidad y extensibilidad. Por su parte, las tablas de dimensiones contienen la descripción de atributos y características asociadas con medidas de eventos tangibles y específicos. Para terminar el ciclo de vida se realiza el despliegue con el objetivo de dejar sentadas las bases de crecimiento y mantenimiento del DWH [13].

1.4.3 Desarrollo de almacenes de datos dirigidos por modelos (Trujillo)

Esta metodología propone el uso de una arquitectura dirigida por modelos (conocida por sus siglas en inglés como MDA: Model Driven Architecture) para el desarrollo multidimensional del DWH de una manera estructurada que facilite la tarea del diseñador mediante la automatización del proceso de 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.

En esta propuesta (ver figura 7) a partir de los requisitos de los usuarios del DWH se construye un modelo computacional independiente, cumpliendo con los establecido en el primer nivel (NIVEL CIM). De este nivel, se deriva un modelo conceptual multidimensional o PIM (Platform Independent Model) inicial que contiene los elementos multidimensionales necesarios para cumplir con los requisitos sin tener en cuenta ninguna tecnología de bases de datos. En este mismo nivel se definen un conjunto de relaciones QVT (Query/View/Transformation) para reconciliar este modelo con las fuentes operacionales, obteniéndose un PIM enriquecido que será consistente con las fuentes de datos.

El siguiente paso consiste en el desarrollo de diferentes PSMs (Platform Specific Models) a partir del PIM (mediante la definición de transformaciones QVT), dependiendo de la tecnología en la que se base la implementación del DWH [14].

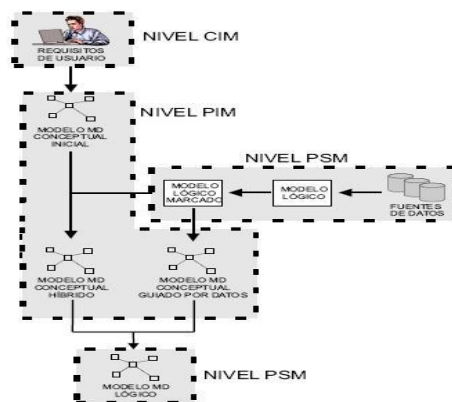


Figura 7: Desarrollo de almacenes de datos dirigidos por modelos (Trujillo)

1.4.4 Data Warehouse Engineering Process (DWEPE)

Metodología que permite modelar todos los aspectos fundamentales en los modelos de datos (lógico, conceptual y físico) de los DWH, está basada en el proceso unificado (en inglés Unified Process, UP). El DWEPE mantiene las cuatro fases del UP: inicio, elaboración, construcción y transición.

Se presentan gráficamente en la figura 8 la relación existente entre los flujos de trabajo y las fases tanto de UP como de DWEPE [15].

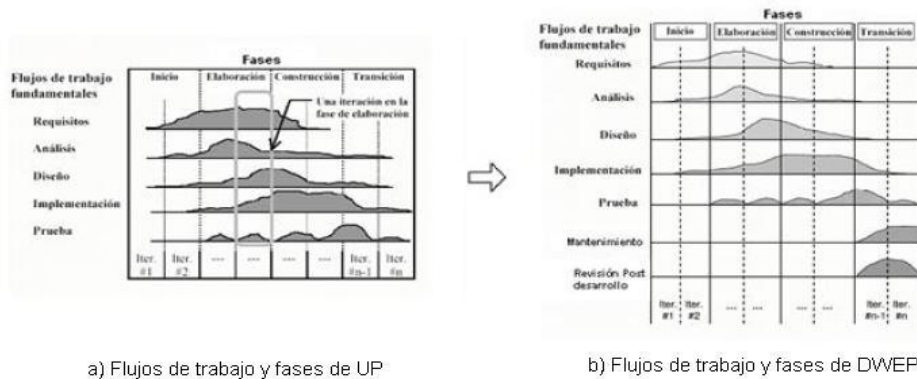


Figura 8: Relación que existe entre los flujos de trabajo y las fases tanto de UP como de DWEPE

1.4.5 Rapid Warehousing Methodology (RWM)

Metodología iterativa, está basada en el desarrollo incremental del proyecto de DWH dividido en cinco fases: definición de los objetivos, definición de los requerimientos de información, diseño y modelización, implementación y revisión (ver figura 9).

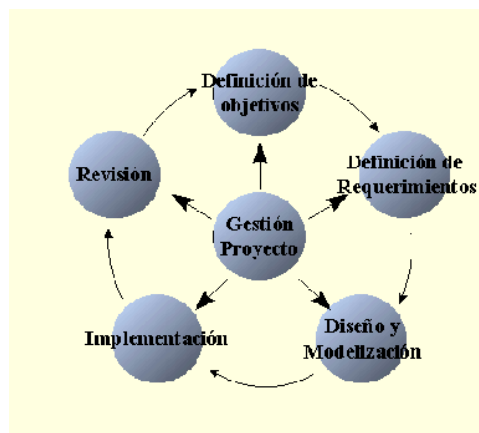


Figura 9: Rapid Warehousing Methodology

La etapa de diseño de esta metodología, está enmarcada en las tres primeras fases: definición de los objetivos, definición de los requerimientos de información, así como diseño y modelización.

En la primera fase de esta metodología se definen los objetivos del negocio para luego, en la segunda etapa, definir los requerimientos de información tal como sucede en todo tipo de proyectos.

Los requerimientos de información identificados durante la anterior fase proporcionarán las bases para realizar el diseño y la modelización del DWH, aquí es donde se enmarca la etapa de diseño. En esta fase se identificarán las fuentes de los datos y las transformaciones necesarias para, a partir de dichas fuentes, obtener el modelo lógico de datos del DWH. Este modelo estará formado por entidades y relaciones que permitirán resolver las necesidades de negocio de la organización. El modelo lógico se traducirá posteriormente en el modelo físico de datos que se almacenará en el DWH y que definirá la arquitectura de almacenamiento del DWH adaptándose al tipo de explotación que se realice del mismo. La mayor parte estas definiciones de los datos del DWH estarán almacenadas en los metadatos y formarán parte del mismo [16].

En el caso de las metodologías de desarrollo existentes y teniendo en cuenta que cada una presenta características específicas que establecen sus diferencias, optar por alguna de ellas no facilitará el proceso de desarrollo del DWH. Cuando se realiza la selección, se descartan las metodologías DWEP y RWM por la poca documentación encontrada en la bibliografía consultada. Luego de analizada la metodología basada en modelos, se evidencia que a pesar de tener una base consistente en la especificación de los requisitos a través de la identificación de objetivos, 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. En el caso de la metodología Hefesto, a pesar de poseer una explicación detallada de cómo analizar las necesidades de la organización identificadas, así como realizar el diseño de los modelos de datos, no es factible utilizarla debido a que esta no facilita la comunicación entre especialistas del negocio y especialistas técnicos, como consecuencia del tipo de información que se gestiona en el CIM. Por otra parte, el ciclo de vida Kimball, a pesar de poseer una estructura bien organizada para el desarrollo de un DWH, no explica detalladamente los elementos necesarios para la etapa de diseño, ni cómo se deben desarrollar. Sin olvidar que estas metodologías han sido creadas con el fin de ser aplicadas en sectores productivos, financieros y de servicios; conociendo además las peculiaridades de la gestión de los Ensayos Clínicos, se propone tomar lo mejor y más aplicable de cada metodología

analizada. A partir de esto, se elaborará un procedimiento para la etapa de diseño de un DWH que pueda ser aplicado a la gestión de los Ensayos Clínicos en el Centro de Inmunología Molecular.

1.5 Etapa de diseño de un DWH

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.

1.5.1 Modelo conceptual

Modelo conceptual: descripción de alto nivel de la estructura de la base de datos, en la cual la información se representa a través de indicadores y perspectivas.

Los **indicadores** son valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: promedios, cantidades, sumatorias y fórmulas.

Por su parte, las **perspectivas** se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de responder a las preguntas planteadas, por ejemplo: pacientes, proveedores, sucursales, países y productos. Teniendo en cuenta las características de los DWHs de ser variables en el tiempo, se sugiere incluir siempre el **tiempo** como una perspectiva más.

A la hora de modelar conceptualmente el DWH, se debe tener en cuenta la **granularidad**, la cual 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 o sumarizados. Es decir, los datos que posean granularidad fina (nivel de detalle) se podrán resumir 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 [12].

1.5.2 Modelo lógico

Modelo lógico: representación de una estructura de datos, que puede procesarse y almacenarse en algún Sistema Gestor de Bases de Datos (SGBD). A diferencia de los modelos lógicos comunes, el modelo lógico de un DWH tiene en cuenta tres elementos fundamentales para su modelación: tipología de esquema que será utilizada, tabla de dimensiones y tabla de hechos. Otros elementos que distinguen este modelo son las uniones y las jerarquías, que se tienen en cuenta en el proceso de diseño.

Primeramente, se debe seleccionar cuál será la **tipología 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 o copo de nieve, ya que esta decisión afectará considerablemente la elaboración del modelo lógico.

Existen tres tipos de esquemas:

- ✓ Esquemas en estrella
- ✓ Esquema en bola de nieve o copo de nieve
- ✓ 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**.

Para diseñar **tablas de dimensiones** se debe tener en cuenta que en 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 se realizará 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.

Las **tablas de hechos** son las que contendrán los hechos a través de los cuales se construirán los indicadores de estudio.

Para los tres tipos de esquemas, se realizarán las **uniones** correspondientes entre sus tablas de dimensiones y sus tablas de hechos.

A nivel de dimensiones es posible definir **jerarquías**, las cuales son grupos de atributos que siguen un orden preestablecido. Una jerarquía implica una organización de niveles dentro de una dimensión, donde cada nivel representa el total agregado de los datos del nivel inferior. Una dimensión típica soporta una o más jerarquías naturales. Una jerarquía puede, pero no exige contener todos los valores existentes en la dimensión [12].

1.5.3 Evaluación del diseño de un DWH

Del diseño realizado dependen la velocidad de búsqueda, la calidad y oportunidad obtenida en las respuestas dadas por el almacén de datos. Una decisión mal tomada en este aspecto puede significar demoras de días para la obtención de las respuestas requeridas, o incluso, que el almacén de datos no esté en capacidad de responder las preguntas claves para la organización. Por esta razón es necesario comprobar que el diseño que se haga para la construcción del DWH sea el correcto [17].

En la bibliografía consultada no se encontró ninguna forma de evaluar el diseño de un DWH; sin embargo, se puede hacer referencia a elementos importantes que no deben faltar una vez que se realice el diseño y que serían de mucha importancia que se tenga en cuenta a la hora de evaluar el diseño propuesto [12-14-17-18]:

- ✓ El diseño requiere soportarse en un análisis profundo de la institución o de áreas funcionales responsables de la utilización del almacén, las fuentes de datos que poblarán el DWH y las personas capacitadas en su correcta estructuración.
- ✓ Debe existir una total correspondencia entre el análisis del negocio realizado y los principales objetivos estratégicos de la institución.
- ✓ Es muy importante especificar correctamente las necesidades de la organización a través de los requisitos funcionales, los cuales deben ser: completos, consistentes, no ambiguos, precisos, verificables, modificables y que puedan ser leídos; de esta forma se garantizaría que cumplieran con los objetivos planteados por el cliente.
- ✓ Se deben descomponer y analizar profundamente las necesidades de la organización para lograr identificar correctamente los indicadores que se utilizarán y las perspectivas de análisis que intervendrán; así como la relación con los datos fuentes.
- ✓ Un buen diseño de un modelo conceptual poseerá un alto nivel de definición de los datos y permitirá ser explicado fácilmente a los involucrados.
- ✓ Se debe definir la granularidad teniendo en cuenta que representará el nivel de detalle con el que se desea almacenar la información sobre el negocio analizado.
- ✓ Seleccionar una tipología de esquema que se adapte mejor a los requerimientos y necesidades del usuario para así lograr una correcta elaboración del modelo lógico.
- ✓ Diseñar correctamente las tablas de dimensiones y tablas de hechos teniendo en cuenta que cada perspectiva en el modelo conceptual constituirá una tabla dimensión y que las tablas de hechos contendrán los hechos a través de los cuales se construirán los indicadores de estudio.

- ✓ Realizar uniones correspondientes con el objetivo de evitar ambigüedad en las tablas de dimensiones así como determinar jerarquías entre los atributos que posean una relación padre-hijo.
- ✓ Un buen diseño de un modelo lógico, el cual se realiza a partir del modelo conceptual, permite tener una primera representación de la estructura de los datos.

Al no encontrar una forma de evaluación específica para el diseño de un Datawarehouse, se propone elaborar y aplicar una lista de chequeo con el propósito de dar cumplimiento a este objetivo.

1.6 Lenguaje de modelado unificado

A lo largo de los años, el desarrollo de los proyectos de software causan grandes confusiones y malas interpretaciones en los requerimientos de los clientes y usuarios; en parte, debido a la abundancia de notaciones, metodologías y conceptos que hacen que los desarrolladores de sistemas no se pongan de acuerdo en lo que realmente están elaborando. En un esfuerzo para estandarizar las notaciones y procesos a utilizar, se conformó el Lenguaje Unificado de Modelado (UML, por sus siglas en inglés, Unified Modeling Language) siendo hoy el lenguaje de modelado de sistemas de software más conocido y utilizado en la actualidad. Es un lenguaje gráfico para visualizar, especificar, construir y documentar un sistema de software. Ofrece un estándar para describir un "plano" del sistema (modelo); incluyendo aspectos conceptuales tales como procesos de negocios y funciones del sistema, y aspectos concretos como expresiones de lenguajes de programación, esquemas de bases de datos y componentes de software reutilizables. Es importante resaltar que UML es un "lenguaje" para especificar y no para describir métodos o procesos. Se utiliza para definir un sistema de software, para detallar los artefactos en el sistema y para documentar y construir [19]. Se decide utilizar UML como lenguaje de modelado en su versión 2.0.

1.7 Herramientas de modelado

Existen varias herramientas creadas para el desarrollo de la Ingeniería de Software, con el fin de desarrollar programas, utilizando técnicas de diseño y metodologías bien definidas, soportadas por herramientas automatizadas. Ejemplo de ello, son las herramientas CASE (Computer Aided Software Engineering), las cuales constituyen un conjunto de ayudas para el desarrollo de programas informáticos.

Entre las herramientas CASE orientadas a UML se encuentran:

- ✓ Rational Rose
- ✓ ArgoUML

- ✓ Poseidón
- ✓ Visual Paradigm
- ✓ MagicDraw UML
- ✓ Borland Together

En la presente investigación se decidió utilizar Visual Paradigm para UML en su versión 6.1 (versión 3.1 de la suite del producto) porque es multiplataforma y permite su uso en cualquier sistema operativo. Utiliza UML como lenguaje de modelado y posibilita una rápida construcción de las aplicaciones con alta calidad. Es factible a la hora de dibujar diagramas de clases, y generar script para diferentes SGBD. Además, permite una integración con sistemas de control de versiones que almacenan centralmente los artefactos y realizan un seguimiento de los cambios realizados sobre un proyecto. Los desarrolladores lo utilizan para facilitar el modelado simultáneo, almacenar los archivos de proyectos y hacer un seguimiento de los cambios [19].

1.8 Técnicas de captura de requisitos

Desde el inicio del desarrollo de sistemas, los ingenieros han presentado dificultad con la identificación de los requisitos del sistema. Esto es debido a que no es un proceso que se pueda determinar matemáticamente, sino un proceso en el cual los datos son extraídos de las personas y estos datos pueden variar; dependiendo de la persona a la que se esté consultando. Se le ha dado solución a esto a través del desarrollo de técnicas que permitan hacer este proceso de una forma más eficiente y segura [20].

Algunas de las técnicas más utilizadas para la captura de requisitos son las siguientes:

- ✓ Entrevistas
- ✓ Tormenta de ideas
- ✓ Cuestionarios
- ✓ Observaciones
- ✓ Discusiones
- ✓ Análisis de protocolo
- ✓ Casos de uso
- ✓ Introspección

Ninguna técnica es efectiva por sí misma, requiere ser combinada con otras para lograr “una correcta” especificación de requisitos. Para esta investigación se utilizaron principalmente las entrevistas, tormenta de ideas, observaciones y discusiones.

1.9 Conclusiones

Este capítulo abarca una panorámica general del proceso de desarrollo de un Datawarehouse, así como el estudio de las tendencias actuales de las metodologías de arquitectura y desarrollo, tecnologías y herramientas que se utilizarán en este Trabajo de Diploma. Luego de esta investigación se arribaron a las siguientes conclusiones:

- ✓ Se decidió utilizar como metodología de arquitectura la propuesta por Kimball, siendo la más aceptada en todo el mundo como la metodología más efectiva para desarrollar una solución de construcción de Datawarehouse.
- ✓ Se seleccionó lo mejor y más aplicable de cada metodología de desarrollo analizada con el objetivo de elaborar un procedimiento para la etapa de diseño de un Datawarehouse; que permita integrar los datos de los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular.
- ✓ Se decidió utilizar el Lenguaje de Modelado Unificado como lenguaje de modelado en su versión 2.0.
- ✓ Se decidió utilizar Visual Paradigm para el Lenguaje de Modelado Unificado en su versión 6.1 (versión 3.1 de la suite del producto) como herramienta de modelado, porque es multiplataforma lo que permite su uso en cualquier sistema operativo.
- ✓ Se utilizaron principalmente las entrevistas, tormenta de ideas, observaciones y discusiones como técnicas de captura de requisitos.
- ✓ Se decidió elaborar y aplicar una lista de chequeo con el propósito de evaluar el diseño propuesto.

CAPÍTULO 2: DISEÑO DE UN DATAWAREHOUSE PARA LOS ENSAYOS CLÍNICOS QUE SE GESTIONAN EN EL CENTRO DE INMUNOLOGÍA MOLECULAR

2.1 Introducción

En este capítulo, se realiza el diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular y se propone un procedimiento para esta etapa explicando paso a paso la propuesta.

2.2 Propuesta de procedimiento

En este acápite se realiza la propuesta de procedimiento para la etapa de diseño del desarrollo de un Datawarehouse para los Ensayos Clínicos del Centro de Inmunología Molecular (ver figura 10)

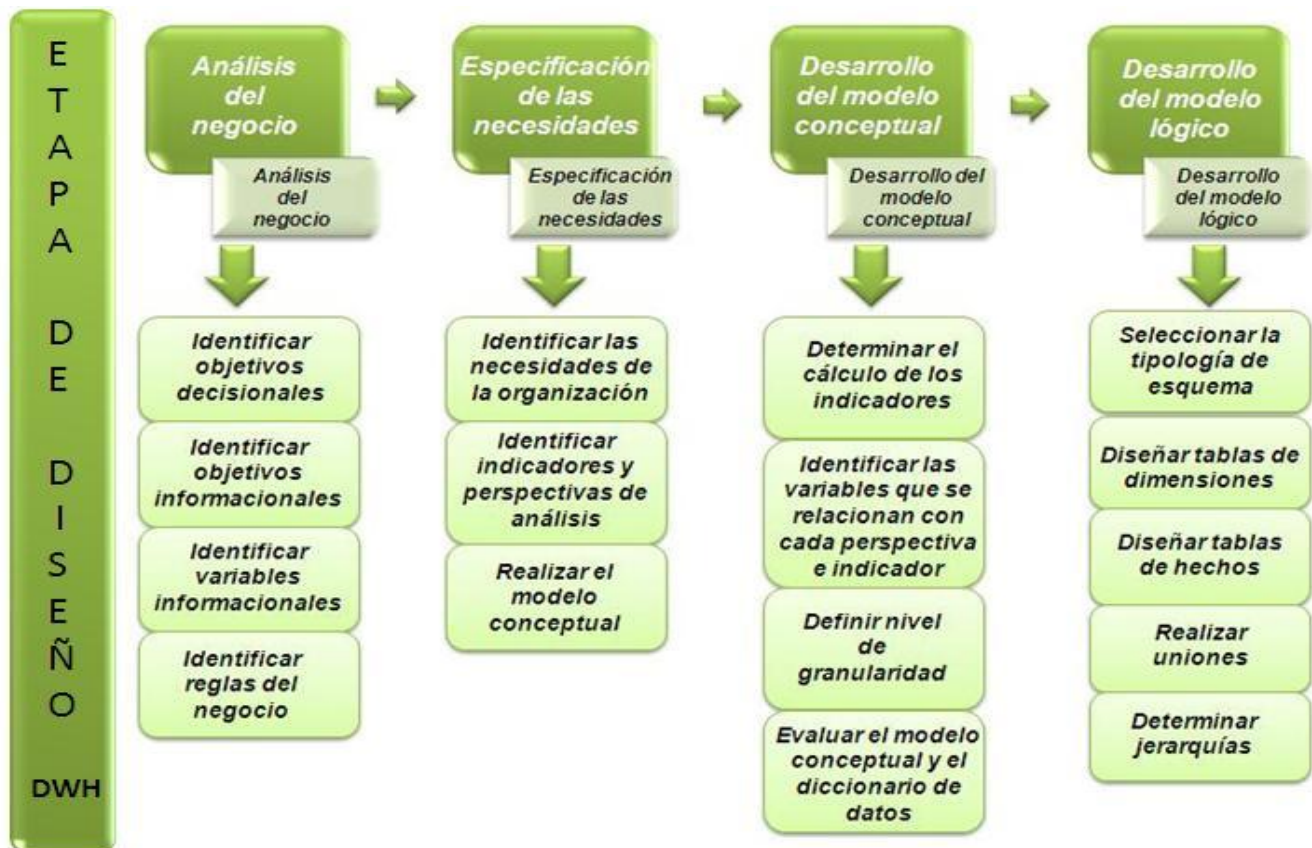


Figura 10: Procedimiento para la etapa de diseño de un DWH para los EC que se gestionan en el CIM

2.2.1 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 variadas y diferentes técnicas, cada una de las cuales poseen características inherentes y específicas, como: 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.

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 a tomar para cumplir con un objetivo estratégico. El cumplimiento de un objetivo decisional sólo es beneficioso para la organización si ayuda a alcanzar un objetivo estratégico.

Identificar objetivos informacionales

Los objetivos informacionales están relacionados con la información necesaria para alcanzar un objetivo decisional. De esta manera, los requisitos de información se obtienen fácilmente de estos objetivos informacionales, y se evita la complejidad de intentar especificarlos de manera directa.

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.

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.

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.

2.2.2 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, se analizarán a fin de 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.

Identificar las necesidades de la organización

Se analizan cada uno de los objetivos informacionales identificando claramente 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, verificables, modificables y que puedan ser leídas.

Una forma de asegurarse que se ha realizado un buen análisis e identificado correctamente las necesidades de la organización, es corroborar que el resultado del mismo haga explícitos los objetivos estratégicos planteados por la institución.

Identificar indicadores y perspectivas de análisis

En este paso, se debe proceder a la descomposición de las necesidades para descubrir 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 el **tiempo** se debe tener en cuenta como una perspectiva más de análisis.

Realizar el modelo conceptual

En este paso, se construirá un modelo conceptual a partir de los indicadores y perspectivas obtenidas en el paso anterior. A través de este modelo, se podrá observar con claridad cuáles son los alcances del proyecto, para luego poder trabajar sobre ellos; además al poseer un alto nivel de definición de los datos, permite que pueda ser presentado ante los usuarios y explicado con facilidad.

A la izquierda se colocan las perspectivas seleccionadas, que serán unidas a un óvalo central que representa y lleva el nombre de la relación que existe entre ellas. La relación la constituyen los principales objetivos estratégicos de la institución. De dicha relación y entrelazadas con flechas, se desprenden los indicadores, estos se ubican a la derecha del esquema.

2.2.3 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 el diccionario de datos, identificando para eso las variables que se relacionan. 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.

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: SUM, AVG, COUNT.

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 esto se realiza un análisis detallado de las mismas definiendo la correspondencia que existe entre las variables de las tablas, bases de datos o modelos determinados con sus perspectivas.

Se propone identificar estas correspondencias a través del diccionario de datos (DD) (ver tabla 5).

Capítulo 2: Diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular

Perspectiva	Significado en el negocio	Modelo (s) donde se encuentra	Correspondencia con variable (s) de los datos fuentes	Posibles valores

Tabla 5: Estructura del diccionario de datos

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 (nivel de detalle) 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.

En este paso se determina la variable que integrará cada perspectiva, actualizándose el diccionario de datos. Debe prestarse mucha atención, ya que esta acción determinará la granularidad de la información encontrada en el DWH.

Evaluar el modelo conceptual y el diccionario de datos

En este paso finalmente, después de actualizado el diccionario de datos se realiza una evaluación del mismo, 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. Para tener constancia de esta evaluación, el cliente debe emitir un aval de aceptación.

2.2.4 Desarrollo del modelo lógico

A continuación se confeccionará el modelo lógico de la estructura del DWH, teniendo como base el modelo conceptual que ya ha sido creado. Para ello, primero se definirá el tipo de modelo que se utilizará 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.

Entregable: Desarrollo del modelo lógico.

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 o copo de nieve, ya que esta decisión afectará considerablemente la elaboración del modelo lógico.

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.

Diseñar tablas de hechos

En este paso, se definirán las tablas de hechos, que son las que contendrán los hechos a través de los cuales se construirán los indicadores de estudio.

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 estratégicos.
- ✓ 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.

Capítulo 2: Diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular

- ✓ 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.

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, el cual 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.

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.

2.3 Aplicación del procedimiento

En este acápite se aplica la propuesta de procedimiento para la etapa de diseño del desarrollo de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular.

2.3.1 Análisis del negocio

Identificar objetivos decisionales

Se identificaron dos objetivos decisionales de la organización partiendo de los objetivos estratégicos relacionados a la gestión de los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular:

- ✓ Verificar la **eficacia** del producto hr3 en los pacientes a los cuáles se aplica.
- ✓ Verificar la **seguridad** del producto hr3 en los pacientes a los cuáles se aplica.

Identificar objetivos informacionales

Se identificaron 14 objetivos informacionales que son los que están relacionados con la información necesaria para alcanzar los objetivos decisionales identificados en el paso anterior:

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.

Capítulo 2: Diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular

Para cada objetivo identificado se debe tener en cuenta el tiempo, la localización, el número del ensayo y los datos demográficos.

Identificar variables informacionales

Se identificaron las siguientes variables informacionales a partir de los objetivos informacionales identificados en el paso anterior:

Generales

Localización: es el lugar donde se manifiesta la enfermedad. Estas pueden ser: Cabeza y Cuello (C y C), Glioma, Esófago, Metástasis cerebral (Metacerebral), Tumores Sólidos (T.Sólidos) y Mama.

Ensayo: Está en correspondencia con la localización. Estos son: C y C 040, C y C 046, C y C 055, C y C 076, Glioma 053, Glioma 069, Esófago 075, Metacerebral 079, T.Sólidos 035 y Mama 070.

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.

Datos demográficos: sexo, edad, raza, estadío, clasificación anatomopatológica, grado de diferenciación, ECOG (Karnofsky) y tratamientos previos.

- ✓ **Sexo:** nos dice el sexo del paciente.
- ✓ **Edad:** nos dice la edad del paciente, en el caso que no aparezca esta variable es calculable (restando la fecha de inclusión menos la fecha de nacimiento).
- ✓ **Raza:** nos dice la raza del paciente, puede ser Blanca, Negra, Mestiza y Amarilla.
- ✓ **Estadío:** gravedad de la enfermedad al diagnosticar el paciente (I, II, III, IV).
- ✓ **Clasificación anatomopatológica:** clasificación del tumor al diagnosticar al paciente.
- ✓ **Grado de diferenciación:** forma parte de la clasificación anatomopatológica, puede ser: bien diferenciada, moderado y poco diferenciado o indiferenciado.
- ✓ **ECOG (Karnofsky):** estado general del paciente cuando se diagnostica.
- ✓ **Tratamientos previos:** tratamiento que ha recibido el paciente antes del ensayo, pueden ser: quimioterapia, radioterapia y cirugía.

Para eficacia

Respuestas: son las respuestas que dan los pacientes del ensayo según el tratamiento o el número de dosis que reciben. Los tipos de respuestas que existen son: Respuesta Completa (RC), Respuesta Parcial (RP), Progresión (P), Enfermedad Estable (EE) y Muerte Precoz (MP).

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).

Supervivencia: indica la cantidad de pacientes vivos o muertos dado un tratamiento determinado.

Para seguridad

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.

Intensidad: refleja la intensidad del evento adverso, pueden ser: ligero, moderado, severo, muy severo y muerte.

Causalidad: especifica si el evento adverso está relacionado o no con el tratamiento. Las clasificaciones son: relacionado (posible, probable, muy probable y definitiva) y no relacionado (improbable, no relacionado y desconocido).

Identificar reglas del negocio

Se identificaron las 22 reglas del negocio que aparecen a continuación:

- ✓ 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.
- ✓ Si el nivel de dosis es 0 y presenta eventos adversos, la causalidad del evento adverso es no relacionada.
- ✓ 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.
- ✓ La variable que identifica a un paciente en las bases de datos de gliomas (glioma 053 y glioma 069) no puede ser solamente una concatenación entre sus iniciales y el número de inclusión en el ensayo pues hay pacientes que forman parte del estudio en ambos ensayos. Se determinó que la variable que identifica a los pacientes en estas bases de datos fuera una concatenación entre sus iniciales, el número de inclusión en el ensayo y la localización (Ej.: JSB-9-glioma053, JSB-9-glioma069).
- ✓ Cuando la localización del tumor es mama (base de datos mama 070) se asume que el sexo del paciente es femenino.
- ✓ Cuando el ensayo se aplica a pacientes que se encuentran en fase I aparece el nivel de dosis recibido; sin embargo, si se está en fase II, lo que aparece es el grupo de tratamiento y se debe buscar en el protocolo el nivel de dosis de cada grupo.

Capítulo 2: Diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular

- ✓ Si la intensidad del evento adverso es normal entonces no se incluye como evento adverso.
- ✓ En el caso de las perspectivas que no se les encuentren relación con los datos fuentes en algunas de las bases de datos porque no se recogieron en ese ensayo (Ej.: raza, causalidad, grado de diferenciación) se registra en el DWH como missing.
- ✓ Cuando no aparece la edad se toma la diferencia que existe entre la fecha actual y la fecha de nacimiento.
- ✓ En el caso de la base de datos c y c 046, se generan los estados generales aleatorios de ECOG 0, 1 ó 2 para que los pacientes cumplan con los criterios de inclusión en el ensayo.
- ✓ En el caso de la base de datos metacerebral 079 el cliente determinó que el estadio para todos los pacientes era fase IV.
- ✓ La clasificación anatomopatológica está recogida en las bases de datos en diferentes variables. Se determinó concatenar estas variables para obtener esta clasificación.
- ✓ El grado de diferenciación forma parte de la clasificación anatomopatológica y a los especialistas del negocio les hace falta tenerlo separado pues es otra perspectiva de análisis. Se determinó extraerlo de la clasificación anatomopatológica según los valores que puede tomar esta perspectiva.
- ✓ El sexo se recoge en la mayoría de las bases de datos como 1 y 2. Se determinó cambiar estos valores por masculino y femenino respectivamente.
- ✓ La raza se recoge en la mayoría de las bases de datos como 1, 2, 3 y 4. Se determinó cambiar estos valores por blanca, negra, mestiza y amarilla respectivamente.
- ✓ Los únicos tratamientos previos que se tendrán en cuenta para el análisis serán quimioterapia, radioterapia y cirugía. Se determinó que en los ensayos donde no se hayan recogido, ese campo aparecerá como missing.
- ✓ Para los tipos de respuestas se determinó que independientemente de las respuestas que pueda tener un mismo paciente a los diferentes eventos adversos, se selecciona la mejor respuesta (la mínima).
- ✓ Para el número de dosis se determinó que independientemente de la cantidad de dosis que reciba, se selecciona la última recibida (número máximo de dosis).
- ✓ 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.

- ✓ El estadio para el caso de la base de datos mama 070 se recogió como IIIa y IIIb. Se determinó que todos los pacientes incluidos en este ensayo tenían estadio III.
- ✓ En el caso de la causalidad e intensidad de los eventos adversos se determinó recogerlas como mismo está establecido en la identificación de las variables informacionales.
- ✓ En las bases de datos tumores sólidos 035, metacerebral 079, glioma 053 y glioma 069 no se recogió ECOG sino Karnofsky. Se determinó establecer la correspondencia que existe entre ECOG y Karnofsky a través de la escala evaluación de la capacidad funcional (ver anexo 1).

2.3.2 Especificación de las necesidades

Identificar las necesidades de la organización

Las necesidades de la organización identificadas fueron las siguientes:

Para el proceso de eficacia:

- ✓ Se desea conocer la cantidad de respuestas por tipos de respuestas para un tratamiento determinado, lo que se traduce en: "**Cantidad de respuestas por tipos de respuestas para un tratamiento**".
- ✓ Se desea conocer la cantidad de respuestas por tipos de respuestas para una cantidad determinada de número de dosis, lo que se traduce en: "**Cantidad de respuestas por tipos de respuestas para una cantidad determinada de número de dosis**".
- ✓ Se desea conocer cuál fue el tiempo de supervivencia de los pacientes que se enfrentaron a un tratamiento determinado, lo que se traduce en: "**Tiempo de supervivencia para un tratamiento**".
- ✓ Se desea conocer cuál fue el tiempo de supervivencia de los pacientes que se enfrentaron a una cantidad determinada de número de dosis, lo que se traduce en: "**Tiempo de supervivencia para una cantidad determinada de número de dosis**".
- ✓ Se desea conocer la cantidad de pacientes vivos o muertos en la actualidad que se enfrentaron a un tratamiento determinado, lo que se traduce en: "**Cantidad de supervivencias para un tratamiento**".
- ✓ Se desea conocer la cantidad de pacientes vivos o muertos en la actualidad que se enfrentaron a una cantidad determinada de número de dosis, lo que se traduce en: "**Cantidad de supervivencias para una cantidad determinada de número de dosis**".

Para el proceso de seguridad:

- ✓ Se desea conocer cuántos eventos adversos hubo por tipos, por intensidad y por causalidad de los eventos adversos según el tratamiento, lo que se traduce en: "**Cantidad de eventos adversos por tipos, por intensidad y por causalidad de los eventos adversos según el tratamiento**".
- ✓ Se desea conocer cuántos eventos adversos hubo por tipos, por intensidad y por causalidad de los eventos adversos según el número de dosis, lo que se traduce en: "**Cantidad de eventos adversos por tipos, por intensidad y por causalidad de los eventos adversos según el número de dosis**".
- ✓ Se desea conocer cuántos pacientes tuvieron eventos adversos ante un tratamiento determinado, lo que se traduce en: "**Cantidad de pacientes con eventos adversos según tratamiento**".
- ✓ Se desea conocer cuántos pacientes tuvieron eventos adversos ante una cantidad determinada de número de dosis, lo que se traduce en: "**Cantidad de pacientes con eventos adversos para una cantidad determinada de número de dosis**".

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.

Identificar indicadores y perspectivas de análisis

Los indicadores son:

Para el proceso de eficacia:

- ✓ Cantidad de respuestas
- ✓ Tiempo de supervivencia
- ✓ Cantidad de supervivencias

Para el proceso de seguridad:

- ✓ Cantidad de eventos adversos
- ✓ Cantidad de pacientes

Las perspectivas de análisis son:

Para el proceso de eficacia:

- ✓ Tipos de respuestas
- ✓ Tratamiento
- ✓ Número de dosis

Para el proceso de seguridad:

- ✓ Tipos de eventos adversos
- ✓ Intensidad de los eventos adversos
- ✓ Causalidad de los eventos adversos
- ✓ Tratamiento
- ✓ Número de dosis

Teniendo en cuenta las características de los DWHs de ser variables en el tiempo, se decide incluir el **tiempo** como una perspectiva más.

Teniendo en cuenta que para cada objetivo identificado se debe conocer la **localización**, el **número del ensayo** y los **datos demográficos**, se deciden incluirlos como tres perspectivas más.

Realizar el modelo conceptual

Para el proceso de eficacia (ver figura 11):

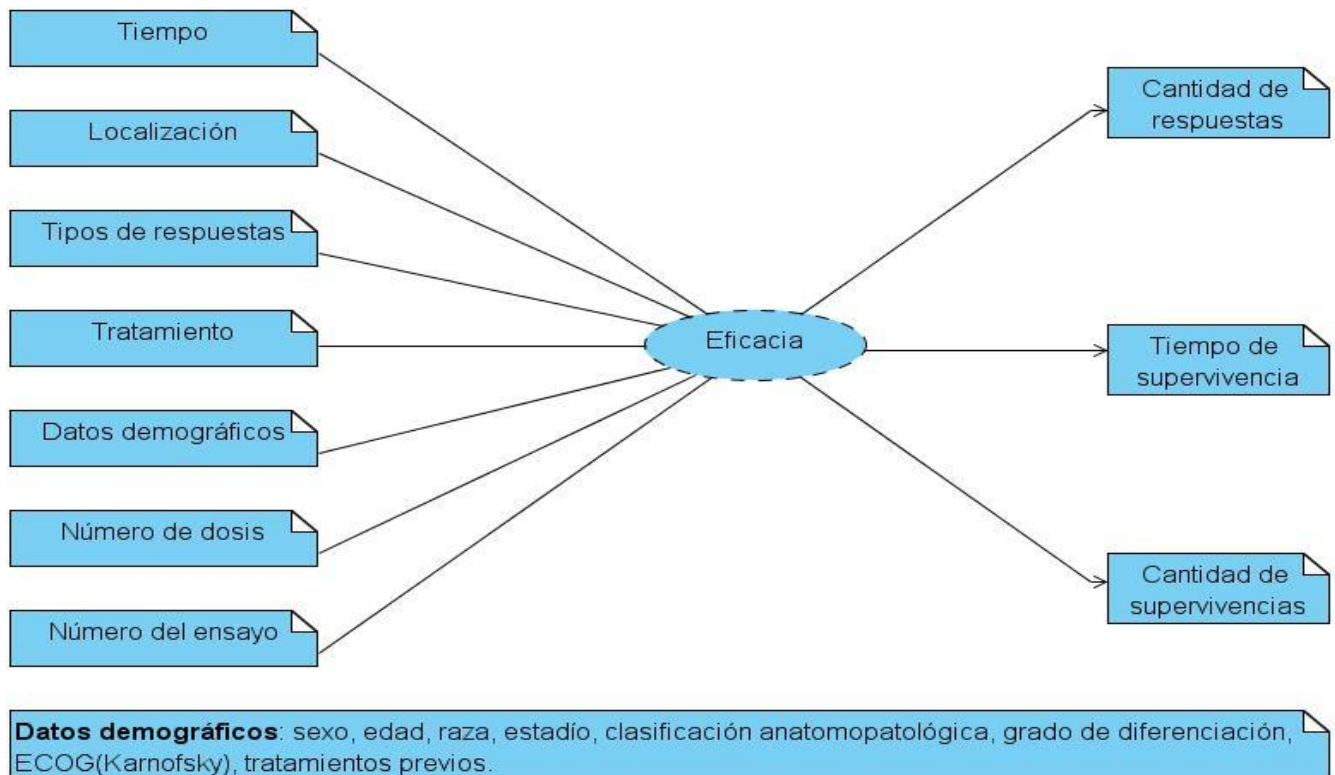


Figura 11: Modelo conceptual para eficacia

Para el proceso de seguridad (ver figura 12):

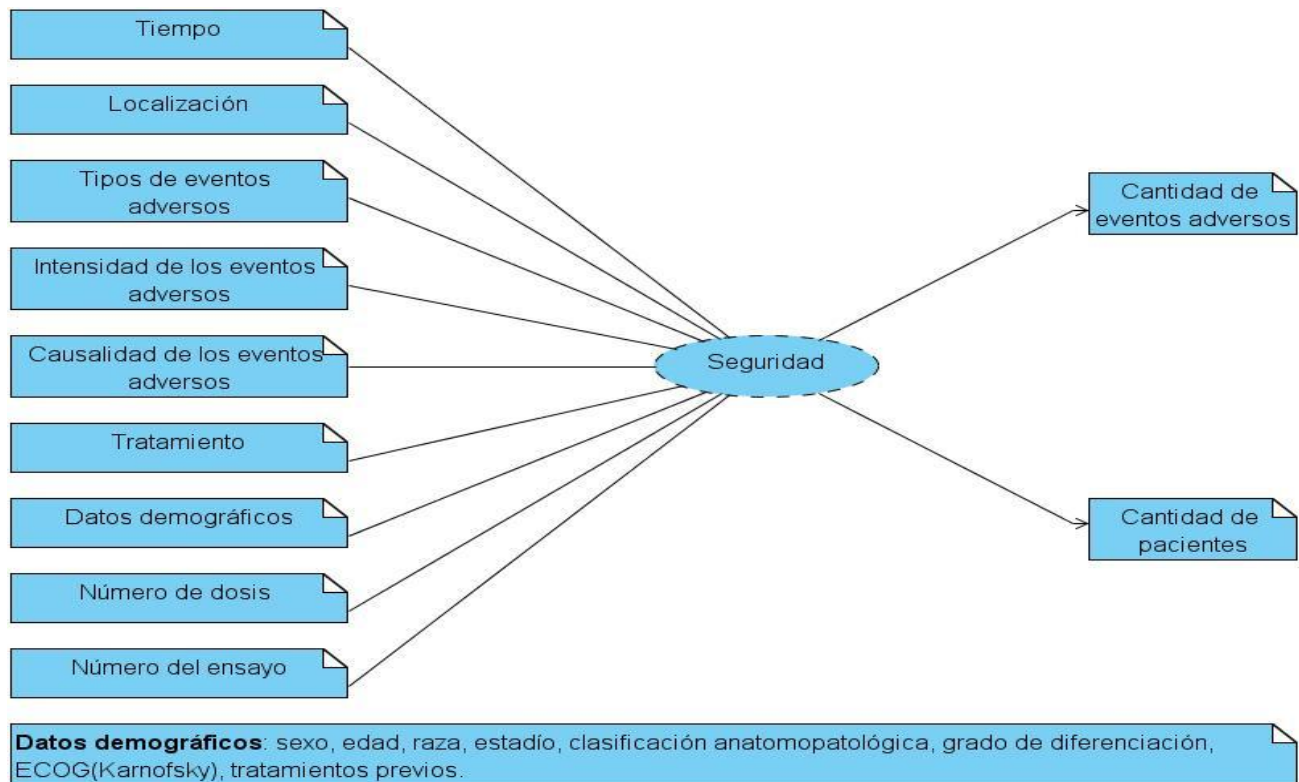


Figura 12: Modelo conceptual para seguridad

Como puede apreciarse en las figuras anteriores, el modelo conceptual permite de un sólo vistazo y sin poseer demasiados conocimientos previos, comprender cuáles serán los resultados que se obtendrán, cuáles serán las variables que se utilizarán para analizarlos y cuál es la relación que existe entre ellos.

2.3.3 Desarrollo del modelo conceptual

Determinar el cálculo de los indicadores

Los indicadores se calcularán de la siguiente forma:

Para el proceso de eficacia:

- ✓ La "**Cantidad de respuestas**" representa la cantidad de respuestas dadas ante un tratamiento o un número de dosis determinado. Este indicador puede ser analizado por cada una de las perspectivas de análisis definidas en la **Especificación de las necesidades**. Se obtiene al sumar cada una de las respuestas que cumplan con los criterios de análisis.

- ✓ El "**Tiempo de supervivencia**" representa el tiempo que duró un paciente desde que se le diagnostica la enfermedad hasta que muere, teniendo en cuenta el tratamiento que recibió, así como el número de dosis recibidas. Generalmente, para saber el tiempo de supervivencia de un paciente se resta la fecha de inclusión del paciente en el ensayo a la última fecha de contacto. Este indicador puede ser analizado por cada una de las perspectivas de análisis definidas en la **Especificación de las necesidades**. Se obtiene al calcular la mediana del tiempo de supervivencia de cada uno de los pacientes que cumplen con los criterios de análisis.
- ✓ La "**Cantidad de supervivencias**" representa la cantidad de pacientes vivos o muertos ante un tratamiento o un número de dosis determinado. Este indicador puede ser analizado por cada una de las perspectivas de análisis definidas en la **Especificación de las necesidades**. Se obtiene al sumar cada uno de los pacientes que cumplan con los criterios de análisis.

Para el proceso de seguridad:

- ✓ La "**Cantidad de eventos adversos**" representa la cantidad de eventos adversos por un tipo, una intensidad y una causalidad de evento adverso determinadas. Este indicador puede ser analizado por cada una de las perspectivas de análisis definidas en la **Especificación de las necesidades**. Se obtiene al sumar cada uno de los eventos adversos que cumplan con los criterios de análisis.
- ✓ La "**Cantidad de pacientes**" representa la cantidad de pacientes con eventos adversos. Este indicador puede ser analizado por cada una de las perspectivas de análisis definidas en la **Especificación de las necesidades**. Se obtiene al sumar cada uno de los pacientes que cumplan con los criterios de análisis.

Identificar las variables que se relacionan con cada perspectiva

Se revisaron todas las fuentes de información proporcionadas por el CIM, realizándose así un análisis exhaustivo y detallado de estas. Se encontraron las correspondencias existentes entre las variables de los modelos con las perspectivas identificadas, realizando el diccionario de datos. A continuación se presenta un ejemplo de uno de los diccionarios de datos confeccionados (Diccionario de datos - BD: hr3 C y C 040, ver tabla 6). El resto de los diccionarios de datos se encuentran en el artefacto: Desarrollo del modelo conceptual dentro del expediente de proyecto.

Capítulo 2: Diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular

Diccionario de datos - BD: hr3 C y C 040

Perspectiva	Significado en el negocio	Modelo (s) donde se encuentra	Correspondencia con variable (s) de los datos fuentes	Posibles valores
Datos demográficos				
Sexo	Nos dice el sexo del paciente	Modelo 1. Datos Demográficos	Sexo (v11)	1:Femenino 2:Masculino
Edad	Nos dice la edad del paciente	Modelo 1. Datos Demográficos	Edad (v6)	Entre 18 y 80
Raza	Nos dice la raza del paciente	Modelo 1. Datos Demográficos	Raza (v10)	1: Blanca 2: Negra 3: Mestiza
Estadío	Nos dice el estadío en que se encuentra el paciente	Modelo1. Datos Demográficos	Estadío (v18)	1: I 2: II 3: III 4: IV
Clasificación anatomopatológica	Nos dice la clasificación del tumor al diagnosticar al paciente	Modelo1. Datos Demográficos	V21a V21b	Texto hasta 80 caracteres. Estas variables deben concatenarse
Grado de diferenciación	Forma parte de la clasificación anatomopatológica	Modelo1. Datos Demográficos	V21a V21b	1.Bien Diferenciado 2.Moderado 3.Poco Diferenciado o Indiferenciado
ECOG (Karnofsky)	Nos dice el estado general del paciente cuando se diagnostica	Modelo1. Datos Demográficos	Estado (OMS) (v20)	De 0-2 según la escala de ECOG

Capítulo 2: Diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular

Tratamientos previos	Nos dice el tratamiento que ha recibido el paciente antes del ensayo	Modelo1. Datos Demográficos	Cirugía (v22) Radioterapia (v24) Quimioterapia (v26)	1: True 2: False
Tiempo de Supervivencia				
Fecha de inclusión en el ensayo	Nos dice la fecha de inclusión del paciente en el ensayo	Modelo 1. Datos Demográficos	Fecha (v3)	El valor de esta variable en los datos fuentes es una fecha
Fecha de último contacto con el paciente	Nos dice la última vez que se tuvo contacto con el paciente	Modelo 12. Salida del ensayo	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)	El valor de esta variable en los datos fuentes es una fecha
Tratamiento				
Nivel de dosis	Nivel de dosis aplicada al paciente	Modelo5. Monitoreo Clínico del Paciente.	(V5)	1: 50 mg 2: 100 mg 3: 200 mg 4: 400 mg
Número de dosis	Cantidad de dosis aplicadas al paciente	Modelo5. Monitoreo Clínico del Paciente.	(V6)	1: Primera Administración 2: Segunda Administración 3: Tercera Administración 4: Cuarta Administración 5: Quinta Administración 6: Sexta Administración

Capítulo 2: Diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular

Respuesta				
Tipos de respuestas	Tipos de respuestas del paciente	Modelo 10. Evaluación de la Respuesta Clínica	(V12)	1: Remisión Completa 2: Remisión Parcial 3: Estabilización 4: Progresión 5: Muerte Precoz
Eventos adversos				
Tipos de eventos adversos	Son las reacciones negativas que manifiestan los pacientes durante el ensayo	Modelo 6. Reacciones Adversas	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	ver anexo 2: Listado de eventos adversos
Causalidad de los eventos adversos	Especifica si el EA está relacionado o no con el tratamiento	No se recogió en ningún modelo. El cliente determinó llenar los campos de esta variable como missing.		
Intensidad de los eventos adversos	Grado de intensidad del EA	Modelo 6. Reacciones Adversas	variables 11, 18, 25, 32, 39, 46, 53, 60, 67, 74.	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.

Tabla 6: Diccionario de datos - BD: hr3 C y C 040

Definir nivel de granularidad

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 DWH y se actualizó el diccionario de datos diciendo finalmente que variable integrará cada perspectiva.

Evaluar el modelo conceptual y el diccionario de datos

Después de actualizado el diccionario de datos se realiza una evaluación de este, así como del modelo conceptual para pasar a la próxima etapa del procedimiento. Esta evaluación se realizó en conjunto con los clientes y los especialistas técnicos para verificar si tanto los modelos conceptuales como los diccionarios de datos estaban en correspondencia con las necesidades de la institución. La evaluación resultó satisfactoria y para tener constancia de ello, el cliente emitió un aval de aceptación.

2.3.4 Desarrollo del modelo lógico

Seleccionar la tipología de esquema

Teniendo en cuenta los requerimientos y necesidades del usuario, se decidió emplear como tipología de esquema constelación de hechos pues se tienen dos tablas de hechos que tienen relación con algunas tablas de dimensiones.

Diseñar tablas de dimensiones

Se realizó el diseño de las tablas de dimensiones, teniendo en cuenta que cada perspectiva definida en modelo conceptual constituye una tabla de dimensión.

Las tablas de dimensiones son:

Respuesta

RESPUESTA
-id_respuesta
-tipos_respuestas

Figura 13: Tabla de dimensión RESPUESTA

Tratamiento

TRATAMIENTO
-id_tratamiento
-nivel_dosis

Figura 14: Tabla de dimensión TRATAMIENTO

Número de dosis

NUMERO_DOSIS
-id_numero_dosis
-numero_dosis

Figura 15: Tabla de dimensión NUMERO_DOSIS

Tiempo

TIEMPO
-id_tiempo
-anno
-mes
-dia

Figura 16: Tabla de dimensión TIEMPO

Número del ensayo

NUMERO_ENSAYO
-id_numero_ensayo
-numero_ensayo

Figura 17: Tabla de dimensión NUMERO_ENSAYO

Localización

LOCALIZACION
-id_localizacion
-localizacion

Figura 18: Tabla de dimensión LOCALIZACION

Tipos de eventos adversos

EVENTOS_ADVERSOS
-id_eventos_adversos
-tipos_eventos_adversos

Figura 19: Tabla de dimensión EVENTOS_ADVERSOS

Datos demográficos

DATOS_DEMOGRAFICOS
-id_datos_demograficos
-edad
-sexo
-raza
-estadio
-clasificacion_anatomopatologica
-grado_diferenciacion
-ecog
-tratamientos_previos

Figura 20: Tabla de dimensión DATOS_DEMOGRAFICOS

Causalidad de los eventos adversos

CAUSALIDAD_EVENTOS_ADVERSOS
-id_causalidad_eventos_adversos
-causalidad_eventos_adversos

Figura 21: Tabla de dimensión CAUSALIDAD_EVENTOS_ADVERSOS

Intensidad de los eventos adversos

INTENSIDAD_EVENTOS_ADVERSOS
-id_intensidad_eventos_adversos
-intensidad_eventos_adversos

Figura 22: Tabla de dimensión INTENSIDAD_EVENTOS_ADVERSOS

Diseñar tablas de hechos

Las tablas de hechos son:

Eficacia

EFICACIA
+cantidad_respuestas()
+tiempo_supervivencia()
+cantidad_supervivencias()

Figura 23: Tabla de hechos EFICACIA

Seguridad

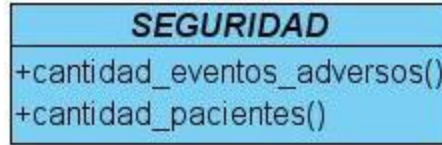


Figura 24: Tabla de hechos SEGURIDAD

Realizar uniones

Luego de diseñar las tablas de dimensiones y las tablas de hechos a partir de los indicadores y las perspectivas identificados en el modelo conceptual, se realizaron las uniones y se obtuvo el modelo lógico previo siguiente (ver figura 25):

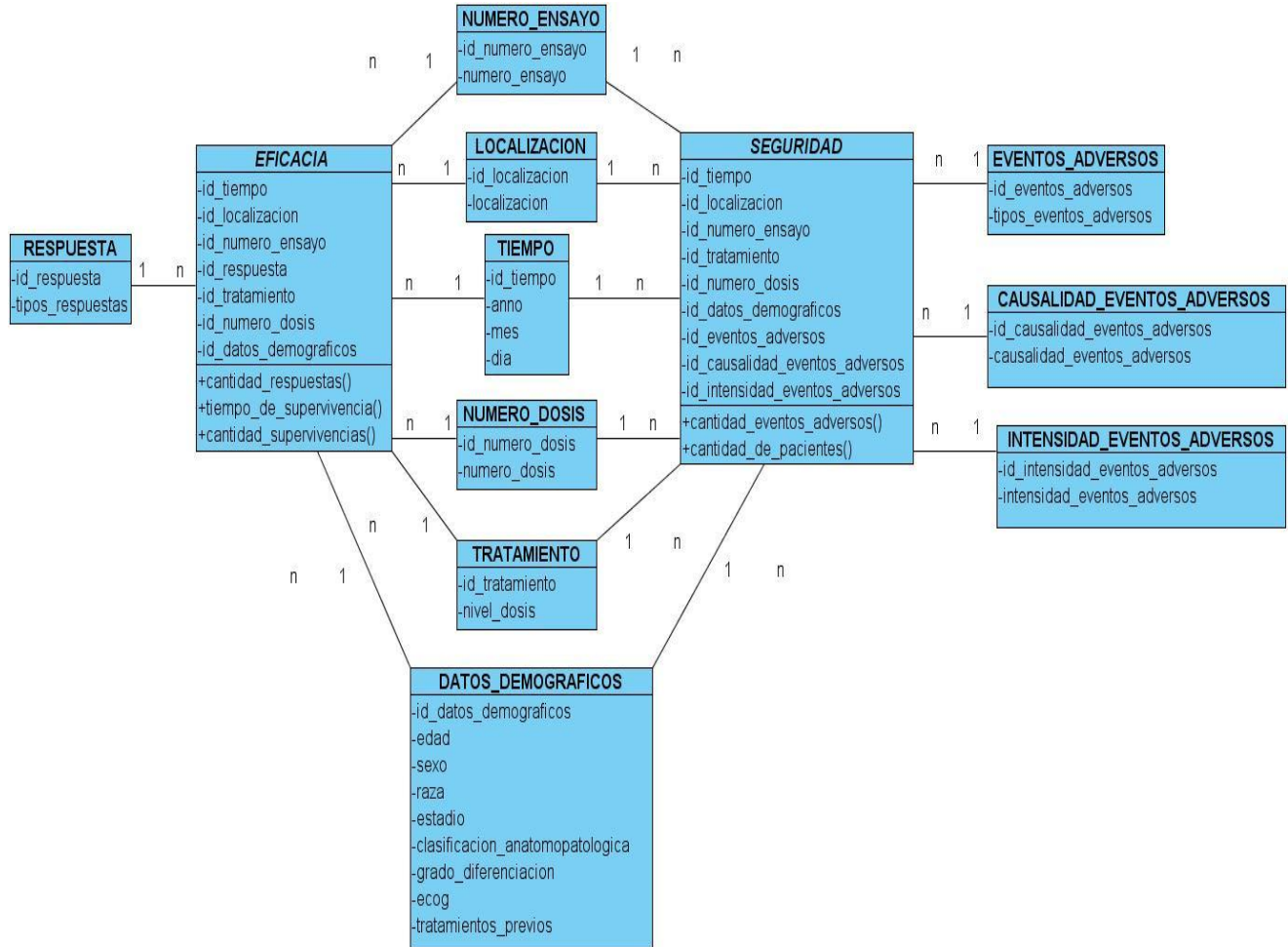


Figura 25: Modelo lógico previo

Capítulo 2: Diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular

Después de analizar el modelo lógico previo se determinó:

- ✓ Unir las tablas EVENTOS_ADVERSOS, CAUSALIDAD_EVENTOS_ADVERSOS e INTENSIDAD_EVENTOS_ADVERSOS en una sola tabla EVENTOS_ADVERSOS pues todas ellas tienen información relacionada a los eventos adversos de un paciente.
- ✓ Unir las tablas NUMERO_DOSIS y TRATAMIENTO en una sola tabla TRATAMIENTO pues las dos tablas tienen información relacionada al tratamiento que reciben los pacientes.
- ✓ Unir las tablas LOCALIZACION y NUMERO_ENSAYO en una sola tabla NUMERO_ENSAYO pues las dos tablas tienen información relacionada a los ensayos clínicos que pertenecen los pacientes.

Una vez que se obtuvo el modelo lógico previo y se realizaron las uniones pertinentes entre las tablas de dimensiones para eliminar datos redundantes, se obtuvo el siguiente modelo lógico final (ver figura 26):

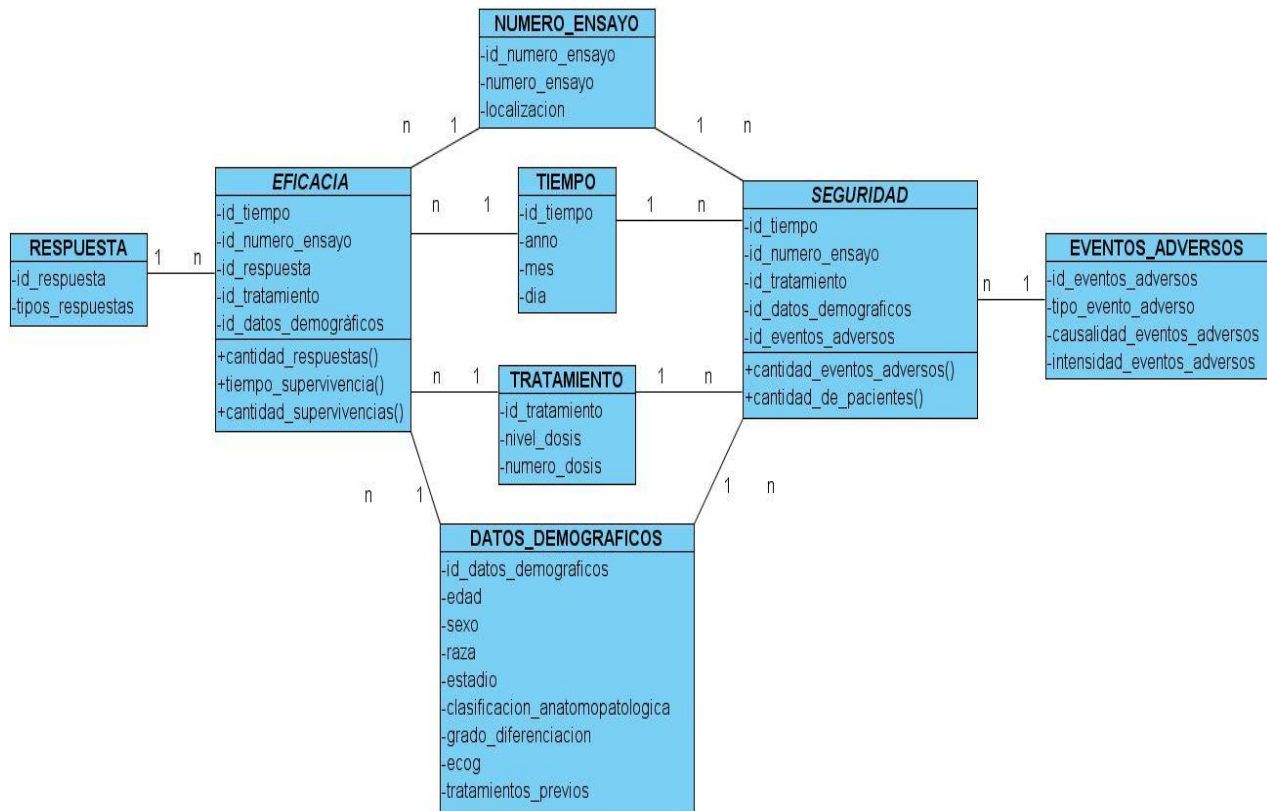


Figura 26: Modelo lógico final

Determinar jerarquías

No se determinó ninguna jerarquía pues no se quiere analizar ningún dato desde un nivel general a otro más detallado y viceversa.

Para realizar la validación del procedimiento propuesto se seleccionó un grupo conformado por siete expertos del Centro de Tecnologías de Datos (DATEC) a los que se les presentó un cuestionario para recoger su opinión sobre los temas referidos a la elaboración y aplicación de dicho procedimiento (ver anexo 3). En el expediente de proyecto se encuentra el documento **Validación del procedimiento**; en este queda reflejado paso a paso cómo se aplicó el método Delphi para la validación de la propuesta. Por los resultados arrojados en la validación se llega a la conclusión que el procedimiento sometido al criterio del panel de expertos resultó Muy Adecuado.

2.4 Conclusiones

En este capítulo se propuso un procedimiento para la etapa de diseño de un Datawarehouse, se explicó paso a paso la propuesta y se aplicó a uno de los productos sobre el cual se realizan los Ensayos Clínicos en el Centro de Inmunología Molecular. Luego de haber aplicado este procedimiento, se obtuvieron los siguientes resultados:

- ✓ Se identificaron dos objetivos decisionales a partir de los objetivos estratégicos de la institución.
- ✓ Se identificaron 22 reglas del negocio.
- ✓ Se identificaron diez necesidades fundamentales de la organización de acuerdo a los objetivos estratégicos.
- ✓ Se identificaron cinco indicadores y diez perspectivas de análisis a partir de las necesidades fundamentales de la organización.
- ✓ Se diseñaron dos modelos conceptuales a partir de los indicadores y las perspectivas identificados.
- ✓ Se realizaron diez diccionarios de datos a partir del estudio realizado a las bases de datos proporcionadas por los especialistas de la institución, con el objetivo de establecer la correspondencia de las variables con los datos fuentes.
- ✓ Se definió la granularidad a nivel de paciente.
- ✓ Se decidió emplear como tipología de esquema constelación de hechos.
- ✓ Se modelaron diez tablas de dimensiones a partir de las perspectivas identificadas.
- ✓ Se modelaron dos tablas de hechos a partir de los procesos fundamentales.

Capítulo 2: Diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular

- ✓ Se diseñó un modelo lógico previo luego de haber establecido las uniones correspondientes entre las tablas de dimensiones y sus relaciones con las tablas de hechos.
- ✓ Se diseñó un modelo lógico final con el objetivo de eliminar la redundancia de datos entre las tablas de dimensiones.
- ✓ No se determinó ninguna jerarquía.
- ✓ Se validó el procedimiento propuesto a través del método Delphi el cual resultó **Muy Adecuado** luego de sometido al criterio del panel de expertos.

CAPÍTULO 3: EVALUACIÓN DEL DISEÑO REALIZADO

3.1 Introducción

En este capítulo, se elabora y aplica una lista de chequeo con el objetivo de evaluar el diseño del DWH realizado.

3.2 ¿Por qué lista de chequeo?

Debido a la gran evolución tecnológica existente en el mundo del desarrollo de software, se ha hecho necesaria la creación de mecanismos para asegurar la calidad y obtener el mejoramiento continuo de los productos. Desde este enfoque, el equipo de calidad de la Universidad de las Ciencias Informáticas ha establecido estándares para definir su propio modelo de aseguramiento de la calidad de software. Uno de los mecanismos más utilizados para la evaluación y control de los productos que se desarrollan en la universidad es la lista de chequeo; documento que tiene un conjunto de parámetros a medir sobre un aspecto determinado. Anotando que algunos parámetros de dicho documento se ajustaron a las necesidades de evaluación para la presente investigación y teniendo en cuenta que en la bibliografía consultada no se encontró ninguna forma de evaluación, se determinó confeccionar una lista de chequeo para evaluar el diseño del Datawarehouse realizado para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular.

3.2.1 ¿Qué es una lista de chequeo?

La lista de chequeo es un documento que tiene un conjunto de parámetros a medir sobre un aspecto determinado para verificar el grado de cumplimiento de determinadas reglas. Es un instrumento de medición y evaluación que consiste básicamente 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. No existe un modelo estándar para la confección de una lista de chequeo; existen diversos formatos en dependencia del negocio sobre el cual se esté trabajando, sin que esto quiera decir que se pueden tomar una de estas y aplicarla a un negocio similar.

3.3 Elaboración y evaluación de la lista de chequeo

Para elaborar la lista de chequeo, se tuvieron en cuenta los elementos de validación que no deben faltar una vez que se realice el diseño del DWH; estos elementos se encuentran referenciados en el Epígrafe

1.5.3 del Capítulo 1 de la presente investigación. La lista de chequeo contiene diferentes indicadores a evaluar los cuales se encuentran distribuidos en tres secciones fundamentales:

- ✓ **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 diseño.
- ✓ **Semántica del documento:** contempla todos los indicadores a evaluar respecto a la ortografía, redacción y otros aspectos de forma y estilo.

Elementos que forman parte de la estructura de la lista de chequeo:

- ✓ **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**; estos últimos dependen de los elementos de validación definidos para la etapa de diseño de construcción de un Datawarehouse. A un elemento de validación puede responder uno o varios indicadores.
- ✓ **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.
- ✓ **N.P. (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**, la cual presenta la siguiente estructura:

- ✓ **No.:** es un número consecutivo e indica la cantidad de no conformidades identificadas.
- ✓ **Elemento de evaluación:** se refiere a un número que identifica al elemento de evaluación para el cual se corresponden los indicadores identificados.
- ✓ **No Conformidad (NC):** especifica la no conformidad a la que se refiere.

- ✓ **Fase correspondiente:** especifica la fase del procedimiento a la que corresponde la no conformidad encontrada.
- ✓ **Significación:** especifica si la no conformidad es o no significativa, dependiendo si el indicador es o no crítico.
- ✓ **Recomendación:** especifica si la no conformidad es una recomendación, es decir que no es de obligatorio cumplimiento que se solucione por parte de los diseñadores.
- ✓ **Estado NC:** especifica el estado de solución en que se encuentra la no conformidad, puede ser: **pendiente** o **solucionada**.
- ✓ **Respuesta del equipo de desarrollo:** si es necesario se especifica la respuesta que le da el equipo de desarrollo a la no conformidad.

Evaluación del diseño

Se aborta la revisión del diseño revisado 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 no conformidades de una revisión a otra.

Se evalúa de regular la calidad del diseño revisado si no cumple los criterios para ser abortado y:

- ✓ 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.

El diseño 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.

3.4 Evaluación de la etapa de diseño a través de la lista de chequeo

Lista de chequeo para evaluar el diseño realizado

Estructura del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
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)	0		0	
Indicadores definidos por la etapa					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
	1. ¿Se realizó un análisis profundo del área de la organización, responsable de la utilización del DWH?	0		0	
	2. ¿Se realizó un análisis profundo de las fuentes de datos que poblarán el DWH?	0		0	
	3. ¿Las personas que realizarán la estructuración del DWH están capacitadas para eso?	0		0	
crítico	4. ¿Existe una total correspondencia entre el análisis del negocio realizado y los objetivos estratégicos de la institución?	0		0	

	5. ¿Las necesidades de la organización se encuentran correctamente especificadas; es decir: completas, consistentes, no ambiguas, precisas, verificables, modificables y que puedan ser leídas?	0		0	
crítico	6. ¿Las reglas del negocio identificadas regulan algún aspecto del negocio?	0		0	
	7. ¿Los indicadores son valores numéricos y representan lo que se desea analizar concretamente?	0		0	
	8. ¿Quedó explícito cómo se calculará cada indicador?	0		0	
	9. ¿Las perspectivas se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de resolver las necesidades planteadas?	0		0	
	10. ¿Se definieron las correspondencias que existen entre las variables de las tablas, bases de datos o modelos determinados con sus perspectivas?	1		13	Se encontraron campos de variables sin definir
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?	0		0	
crítico	12. ¿La granularidad definida representa el nivel de detalle con el que se desea almacenar la información sobre el negocio analizado?	0		0	

crítico	13. ¿La granularidad definida permite tener mayores posibilidades analíticas?	0		0	
	14. ¿La tipología de esquema seleccionada es la que mejor se adapta a los requerimientos y necesidades del usuario?	0		0	
	15. ¿Se diseñaron correctamente las tablas de dimensiones?	0		0	
	16. ¿Se diseñaron correctamente las tablas de hechos?	0		0	
	17. ¿Se realizaron las uniones correspondientes entre las tablas de dimensiones y las tablas de hechos?	0		0	
	18. ¿La jerarquía planteada tiene una relación lógica “padre-hijo”?		X		
crítico	19. ¿Permite el modelo lógico tener una primera representación de la estructura de los datos?	0		0	
Semántica del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
crítico	1. ¿Se han identificado errores ortográficos en los entregables o en los modelos diseñados?	0		0	
crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?	0		0	
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?	0		0	

Tabla 7: Lista de chequeo

En el siguiente gráfico de barras aparece el comportamiento de los indicadores definidos para la lista de chequeo elaborada. De forma general se identificaron 23 indicadores, de ellos nueve críticos y luego de aplicada la lista de chequeo se generó una no conformidad de uno de los indicadores definidos por la etapa, es válido aclarar que este indicador no fue considerado como crítico (ver figura 26).

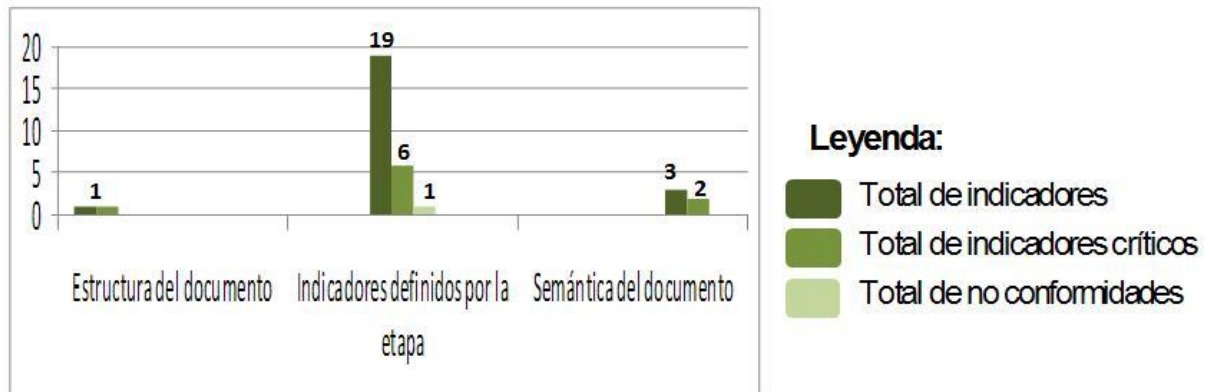


Figura 27: Comportamiento de los indicadores por secciones

Con el objetivo de darle solución al indicador evaluado con dificultades de la sección: **Indicadores definidos por la etapa**, se creó la **tabla de no conformidades** que aparece a continuación (ver tabla 8):

No.	Elemento de evaluación	NC	Fase correspondiente	Significación	Recomendación	Estado NC	Respuesta del equipo de desarrollo
1	4	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.

Tabla 8: Tabla de no conformidades

Se realizaron dos iteraciones en la revisión del diseño del DWH a través de la lista de chequeo propuesta; donde se identificó un indicador con dificultades del cual se generó una NC. 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, no existieron problemas con los formatos de las plantillas, no se encontraron errores ortográficos en los documentos revisados y la NC generada fue solucionada.

3.5 Conclusiones

En este capítulo se elaboró y aplicó una lista de chequeo con el objetivo de evaluar el diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular. Luego de evaluado el diseño a través de la lista de chequeo, se obtuvieron los siguientes resultados:

- ✓ Se identificaron de forma general 23 indicadores necesarios e imprescindibles para la evaluación final de esta etapa de diseño distribuidos en tres secciones: **Estructura del documento**, **Indicadores definidos en la etapa**, **Semántica del documento**; de ellos nueve indicadores con peso crítico.
- ✓ Se evaluó un indicador no crítico con dificultades en la sección **Indicadores definidos por la etapa**, lo cual no afectó la evaluación final de la etapa de diseño.
- ✓ La evaluación final fue de **Bien**, resultado que demuestra la calidad del procedimiento y diseño realizado.

CONCLUSIONES

El estudio de la etapa de diseño para la construcción de un Datawarehouse como parte de su proceso de desarrollo, proporcionó la elaboración del presente trabajo, cuyo resultado fue el “Diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular”. Dicho diseño contribuyó a la integración de los datos de la institución y arrojó los siguientes resultados:

- ✓ Se elaboró un procedimiento para la etapa de diseño de un Datawarehouse con el objetivo de integrar los datos de los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular.
- ✓ Se aplicó el procedimiento para la etapa de diseño de un Datawarehouse logrando la integración de los datos de los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular.
- ✓ Se elaboró una lista de chequeo para evaluar el diseño realizado.
- ✓ Se evaluó el diseño realizado a partir de la aplicación de la lista de chequeo.

El Trabajo de Diploma marcó un punto de partida para las otras dos etapas que forman parte del proceso de construcción de un Datawarehouse, en aras de facilitar su trabajo con la entrega de modelos y artefactos necesarios para su desarrollo. Además de disminuir los errores y redundancias en los datos que poblarán el Datawarehouse de los Ensayos Clínicos en el Centro de Inmunología Molecular.

RECOMENDACIONES

- ✓ Utilizar la presente investigación como base de referencia cuando se quiera realizar el diseño de un Datawarehouse.
- ✓ Realizar la extracción, transformación y carga de los datos, basándose en el diseño propuesto, para obtener datos estandarizados que sirvan para su posterior análisis y toma de decisiones, acorde a las necesidades del Centro de Inmunología Molecular.
- ✓ Utilizar el procedimiento propuesto para guiar la etapa de diseño 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 necesiten una base consistente durante esta etapa.
- ✓ Utilizar la propuesta de evaluación del diseño del Datawarehouse a través de la lista de chequeo en otros proyectos.

REFERENCIAS BIBLIOGRÁFICAS

- [1] CHAN, Dra. Margaret. Cáncer. Organización Mundial de la Salud [en línea] (2009). [Consulta: 23 de noviembre 2009]. Disponible en: <<http://www.who.int/topics/cancer/es/>>
- [2] ALFONSO, Carmen. El problema del cáncer, ¿cómo podemos enfrentarlo? Trabajadores.CU [en línea] (2006). [Consulta: 25 de noviembre 2009]. Disponible en: <http://www.trabajadores.cu/materiales_especiales/suplementos/salud/enfermedades-cronicas-no-trasmisibles-1/el-problema-del-cancer-bfcomo-podemos-enfrentarlo>
- [3] CENTRO DE INMUNOLOGÍA MOLECULAR. Información General [en línea] (2009). [Consulta: 25 de noviembre 2009]. Disponible en: <<http://www.cim.sld.cu/>>
- [4] MINISTERIO DE SALUD PÚBLICA. Centro para el control estatal de la calidad de los medicamentos. Regulación No. 45-2007 [en línea] (2009). [Consulta: 28 de noviembre 2009]. Disponible en: <www.bvv.sld.cu/download.php?url=regulaciones/124413435046.doc>
- [5] CENTRO NACIONAL COORDINADOR DE ENSAYOS CLÍNICOS (CENCEC). Experiencias [en línea] (2009). [Consulta: 10 de diciembre 2009]. Disponible en: <<http://www.cencec.sld.cu/pgs/resultados.htm>>
- [6] LÓPEZ DÍAZ, Aidacelys y RODRÍGUEZ GARCÍA, Lucía. Sistema de Manejo de Datos de Ensayos Clínicos. Tesis (Ingeniero en Ciencias Informáticas) Ciudad Habana. Universidad de las Ciencias Informáticas, 2007.
- [7] SUÁREZ MARTÍNEZ, G. y MOREJÓN ZAYAS, O. M. Informe periódico de Seguridad Nimotuzumab. CIMAB. Ciudad de la Habana, Cuba: s.n., 2010. pág. 34.
- [8] HERNÁNDEZ LÓPEZ, Ing. Asnioby. Almacenes de Datos aplicada a la Seguridad Ciudadana. Tesis (Máster en Informática Aplicada) Ciudad Habana. Universidad de las Ciencias Informáticas, 2009.
- [9] SQL MAX CONNECTIONS. Data Warehousing [en línea] (2009). [Consulta: 12 de diciembre 2009]. Disponible en: <<http://www.sqlmax.com/dataw1.asp>>
- [10] WREMBEL, Robert y CONCILIA, Christian. El modelo de Hechos dimensionales [en línea] (2010). [Consulta: 9 de marzo 2010]. Disponible en: <<http://www.monografias.com/trabajos57/data-warehouse-sql/data-warehouse-sql2.shtml>>

- [11] KLE, Transforming Knowledge into action! BI en la práctica. Artículos de BI en la práctica [en línea] (2010). [Consulta: 9 de marzo 2010]. Disponible en: <<http://www.siskle.com/spanish/articulo04.html>>
- [12] RICARDO DARIO, Ing. Bernabeu. Data Warehousing: Investigación y sistematización de conceptos. Hefesto: Metodología propia para la construcción de un Datawarehouse. Córdoba, Argentina, 2009.
- [13] HUAMANTUMBA, Rayner. Datamart paso a paso [en línea] (2010). [Consulta: 7 de febrero 2010]. Disponible en: <<https://www.u-cursos.cl/ingenieria/2008/1/IN830/1/.../11313>>
- [14] TRUJILLO, Juan; SOLER, Emilio; ZUBCOFF, José; MAZON, José Norberto; GLORIO, Octavio y PARDILLO, Jesús. Desarrollo de almacenes de datos dirigido por modelos.
- [15] LUJAN MORA, Sergio. Data Warehouse Design with UML Tesis (Doctorado). España, Universidad de Alicante, 2005.
- [16] ESPINOSA MILLA, Roberto. El Rincón del BI. Descubriendo el Business Intelligence [en línea] (2010). [Consulta: 7 de febrero 2010]. Disponible en: <<http://churriwifi.wordpress.com/2009/12/05/5-fases-en-la-implantacion-de-un-sistema-dw-metodologia-para-la-construccion-de-un-dw>>
- [17] GUISAO CARTAGENA, Carlos Mauricio y PADILLA URZOLA, León. Bodega de datos (Data warehouse) [en línea] (2005). [Consulta: 9 de marzo 2010]. Disponible en: <<http://www.monografias.com/trabajos24/bodega-de-datos/bodega-de-datos.shtml>>
- [18] KIMBALL, Ralph y ROSS, Margy. The Data Warehouse Lifecycle Toolkit. 2a. ed. Canadá: Wiley Publishing, Inc, 2002. ISBN: 0471200247.
- [19] GONZÁLEZ HERNÁNDEZ, Marlien y DE LAOSA SOCARRÁS, Jenely. Sistema de Manejo de Datos de Ensayos Clínicos Cubano. Módulo Validación: Diseño del submódulo “Derivación de las variables del Cuaderno de Recogida de Datos”. Tesis (Ingeniero en Ciencias Informáticas) Ciudad Habana. Universidad de las Ciencias Informáticas, 2008.
- [20] LAS BITÁCORAS GWAI. Captura de Requisitos [en línea] (2007). [Consulta: 9 de marzo 2010]. Disponible en: <http://jcgm.ei.uvigo.es/jcmoreno/?page_id=2>

BIBLIOGRAFÍA

1. ALFONSO, Carmen. El problema del cáncer, ¿cómo podemos enfrentarlo? Trabajadores.CU [en línea] (2006). Disponible en:
<http://www.trabajadores.cu/materiales_especiales/suplementos/salud/enfermedades-chronicas-no-trasmisibles-1/el-problema-del-cancer-bfcomo-podemos-enfrentarlo>
2. CENTRO DE INMUNOLOGÍA MOLECULAR. Información General [en línea] (2009). Disponible en:
<<http://www.cim.sld.cu/>>
3. CENTRO NACIONAL COORDINADOR DE ENSAYOS CLÍNICOS (CENCEC). Experiencias [en línea] (2009). Disponible en: <<http://www.cencec.sld.cu/pgs/resultados.htm>>
4. CHAN, Dra. Margaret. Cáncer. Organización Mundial de la Salud [en línea] (2009). Disponible en:
<<http://www.who.int/topics/cancer/es/>>
5. ESPINOSA MILLA, Roberto. El Rincón del BI. Descubriendo el Business Intelligence [en línea] (2010). Disponible en: <<http://churriwifi.wordpress.com/2009/12/05/5-fases-en-la-implantacion-de-un-sistema-dw-metodologia-para-la-construccion-de-un-dw>>
6. GONZÁLEZ HERNÁNDEZ, Marlien y DE LAOSA SOCARRÁS, Jenely. Sistema de Manejo de Datos de Ensayos Clínicos Cubano. Módulo Validación: Diseño del submódulo “Derivación de las variables del Cuaderno de Recogida de Datos”. Tesis (Ingeniero en Ciencias Informáticas) Ciudad Habana. Universidad de las Ciencias Informáticas, 2008.
7. GUISAO CARTAGENA, Carlos Mauricio y PADILLA URZOLA, León. Bodega de datos (Data warehouse) [en línea] (2005). Disponible en: <<http://www.monografias.com/trabajos24/bodega-de-datos/bodega-de-datos.shtml>>
8. HERNÁNDEZ LÓPEZ, Ing. Asnioby. Almacenes de Datos aplicada a la Seguridad Ciudadana. Tesis (Máster en Informática Aplicada) Ciudad Habana. Universidad de las Ciencias Informáticas, 2009.
9. HERNÁNDEZ RAMÍREZ, Ing. Martha Denia, SOTOLONGO LEÓN, Ing. Anthony Rafael. Introducción a la Inteligencia de Negocios. Conferencia 1. Universidad de las Ciencias Informáticas. Ciudad Habana, 2009.
10. HUAMANTUMBA, Rayner. Datamart paso a paso [en línea] (2010). Disponible en:
<<https://www.u-cursos.cl/ingenieria/2008/1/IN830/1/.../11313>>

11. KIMBALL, Ralph y ROSS, Margy. The Data Warehouse Lifecycle Toolkit. 2a. ed. Canadá: Wiley Publishing, Inc, 2002. ISBN: 0471200247.
12. KLE, Transforming Knowledge into action! BI en la práctica. Artículos de BI en la práctica [en línea] (2010). Disponible en: <<http://www.siskle.com/spanish/articulo04.html>>
13. LAS BITÁCORAS GWAI. Captura de Requisitos [en línea] (2007). Disponible en: <http://jcgem.ei.uvigo.es/jcmoreno/?page_id=2>
14. LÓPEZ DÍAZ, Aidacelys y RODRÍGUEZ GARCÍA, Lucía. Sistema de Manejo de Datos de Ensayos Clínicos. Tesis (Ingeniero en Ciencias Informáticas) Ciudad Habana. Universidad de las Ciencias Informáticas, 2007.
15. LUJAN MORA, Sergio. Data Warehouse Design with UML Tesis (Doctorado). España, Universidad de Alicante, 2005.
16. MINISTERIO DE SALUD PÚBLICA. Centro para el control estatal de la calidad de los medicamentos. Regulación No. 45-2007 [en línea] (2009). Disponible en: <www.bvv.sld.cu/download.php?url=regulaciones/124413435046.doc>
17. MOREJÓN ZAYAS, MSc. Olga Mary, SUÁREZ MARTÍNEZ, Dra. Giselle. Informe Periódico de Seguridad. Nimotuzumab. Centro de Inmunología Molecular. Ciudad Habana, 2010.
18. RICARDO DARIO, Ing. Bernabeu. Data Warehousing: Investigación y sistematización de conceptos. Hefesto: Metodología propia para la construcción de un Datawarehouse. Córdoba, Argentina, 2009.
19. SQL MAX CONNECTIONS. Data Warehousing [en línea] (2009). Disponible en: <<http://www.sqlmax.com/dataw1.asp>>
20. SUÁREZ MARTÍNEZ, G. y MOREJÓN ZAYAS, O. M. Informe periódico de Seguridad Nimotuzumab. CIMAB. Ciudad de la Habana, Cuba: s.n., 2010. pág. 34.
21. TRUJILLO, Juan, SOLER, Emilio, ZUBCOFF, José, MAZON, José Norberto, GLORIO, Octavio y PARDILLO, Jesús. Desarrollo de almacenes de datos dirigido por modelos.
22. WREMBEL, Robert y CONCILIA, Christian. El modelo de Hechos dimensionales [en línea] (2010). Disponible en: <<http://www.monografias.com/trabajos57/data-warehouse-sql/data-warehouse-sql2.shtml>>

ANEXOS

Anexo # 1: Evaluación de la capacidad funcional

Escala de ECOG		Escala de Karnofsky	
Grado	Descripciones	Porcentaje	Descripciones
0	Capaz de llevar a cabo una actividad física normal sin restricciones.	100	Normal. No presenta síntomas o signos de enfermedad.
		90	Capaz de una actividad normal, ligeros síntomas o signos de enfermedad.
1	Sintomático, pero ambulatorio. Restricción en actividades físicas vigorosas, pero ambulatorio y capaz de hacer trabajos ligeros o de naturaleza sedentaria.	80	Actividad normal con esfuerzo. Algunos síntomas o signos de enfermedad.
		70	Puede cuidar de sí mismo. Incapaz de desarrollar una actividad o trabajo activo normales.
2	En cama menos del 50 % del tiempo. Ambulatorio y capaz de valerse por sí mismo, pero incapaz de trabajar. Más del 50 % del tiempo fuera de la cama.	60	Precisa ocasionalmente asistencia, pero es capaz de atender por sí mismo a la mayor parte de sus propias necesidades.
		50	Requiere asistencia y frecuentes cuidados médicos.
3	Capaz de realizar sus cuidados personales, pero más del 50 % del tiempo confinado a la cama o silla.	40	Incapacidad. Encamado. Requiere asistencia y cuidados especiales.
		30	Grave incapacidad. Estado severo. La muerte no es inminente. Requiere hospitalización.
4	Completamente incapaz de realizar ningún esfuerzo, confinado totalmente a la cama.	20	Estado grave. Intenso tratamiento de sostén. Requiere hospitalización Requiere hospitalización.
		10	Estado muy grave (moribundo). Proceso fatal que progresa rápidamente.
5	Muerto	0	Muerto

Anexo # 2: Listado de Eventos Adversos

Eventos Adversos					
1	Nauseas	34	Induración	67	Cianosis peribucal
2	Vómitos	35	Escalofríos	68	Debilidad, dolor tórax
3	Diarrea	36	Mialgias	69	Rubicundez
4	Hematuria	37	Artralgias	70	Neumonía
5	Disnea	38	Urticarias	71	Molestias al tragar
6	Fiebre	39	Alopecia	72	Sedimento urinario patológico
7	Temblores	40	Estomatitis	73	Neutrofilea
8	Alergia	41	Esofagitis	74	Fatiga
9	Sensación de calor	42	Cistitis	75	Sangramiento bucal
10	Otros	43	Cardiotoxicidad	76	Crisis vagal
11	Radiodermatitis húmeda	44	Infertilidad	77	Edema lengua
12	Radiodermatitis seca	45	Anemia	78	Molestias en la lengua
13	Disfagia	46	Aumento creatinina	79	Somnolencia
14	Mucositis	47	Disminución proteínas total	80	Dolor oído
15	Laringitis	48	Leucocitosis	81	Asco
16	Dolores óseos	49	Leucopenia	82	Sensación de plenitud gástrica
17	Broncoespasmo	50	Sequedad bucal	83	Isquemia arterial
18	Shock anafiláctico	51	Disgeusia	84	Epigastralgia
19	Hipertensión	52	Sepsis	85	Gastritis y esofagitis
20	Hipotensión	53	Pérdida de peso	86	Alta bilirrubina
21	Cefalea	54	Odinofagia	87	Alta glicemia
22	Tos seca	55	Sangrado	88	Hipoglicemia
23	Mareo	56	C.Focal	89	Hipoproteinemia
24	Convulsiones	57	Anorexia	90	Cólico nefrítico
25	Depresión	58	Decaimiento	91	Secreción bronquial
26	TGP	59	Insomnio	92	Neutrofilia
27	L.Trope	60	TGO	93	Dolor de garganta
28	D.E.I.	61	Acidez		
29	Radiode	62	Edema angioneurótico		
30	Embotam	63	Trombocitopenia		
31	Astenia	64	Disfonía por radiaciones		
32	Dolor sitio de inyección	65	Carcinogénesis o leucemogenesis		
33	Eritema local	66	Anafilaxis		

Anexo # 3: Encuesta a expertos 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, actividades así como los artefactos de entrada y salida que se propusieron son correctos, 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 para la etapa de diseño de un Datawarehouse para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular.

Valore el grado de factibilidad de de cada pregunta de acuerdo a la siguiente escala:

- ✓ Muy Adecuado (C1)
- ✓ Bastante Adecuado (C2)
- ✓ Adecuado (C3)
- ✓ Poco Adecuado (C4)
- ✓ No adecuado (C5)

Preguntas	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 la integración de los datos de los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular es:					
2. Se proponen 3 sub procesos fundamentales para lograr la integración de los datos de los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular. Categorice cada uno de ellos:					
2.1 Elaboración de la propuesta					
2.2 Aplicación de la propuesta					

2.3 Calidad de los resultados obtenidos					
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:					
6. La evaluación del modelo conceptual y el diccionario de datos en la etapa desarrollo del modelo conceptual del procedimiento es:					
7. La correspondencia que existe entre la elaboración y aplicación del procedimiento propuesto es:					
8. En sentido general emita su criterio acerca de los aportes que pueda tener la aplicación del procedimiento para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular a la hora de adicionar la información relacionada con otros productos al almacén de datos.					