

IMPLEMENTACIÓN DE UN DATA MART PARA UN SERVICIO DE DOSIMETRÍA EXTERNA.

Manzano de Armas, J.F.¹, Molina Pérez, D. y Castro Soler, A.

Centro de Protección e Higiene de las Radiaciones, La Habana, Cuba.

RESUMEN

El Laboratorio de Dosimetría Externa del Centro de Protección e Higiene de las Radiaciones cuenta con un sistema automatizado para la gestión de los resultados de los diferentes servicios de dosimetría externa recibidos por los trabajadores ocupacionalmente expuestos a las radiaciones ionizantes de la República de Cuba. Este sistema está diseñado y optimizado para el registro de los datos generados por el servicio y la emisión de los reportes con los resultados dosimétricos para los diferentes periodos de control. Sin embargo, no está diseñado para facilitar la exploración y análisis de los datos desde diferentes perspectivas para la creación de conocimiento y como soporte en la toma de decisiones.

El objetivo del trabajo es el diseño e implementación de un Data Mart para facilitar el almacenamiento, exploración y análisis de la gran cantidad de datos contenidos en el sistema automatizado de gestión. Los Data Mart son almacenes de datos que están centrados en un tema y responden a los intereses de determinada áreas de negocio dentro de una organización. Se caracterizan por disponer de una estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dichas áreas de negocio. Para el diseño del Data Mart se utilizó la metodología HEFESTO. Se cargaron y analizaron los datos mediante el empleo de cubos multidimensionales, utilizando herramientas libres como Mondrian Schema Workbench, Saiku Server y MySQL. Los resultados obtenidos contribuyen a tener una mejor comprensión del desempeño de los servicios de dosimetría externa en el pasado, monitorear las actividades actuales para responder rápidamente a los cambios, mejorar los procesos de los diferentes servicios, obtención de datos para la revisión de los programas de optimización y pronosticar nuevas posibilidades de ampliación de los servicios.

1. INTRODUCCIÓN

Un aspecto importante de los programas de vigilancia radiológica es el mantenimiento de los registros individuales de dosis de los Trabajadores Ocupacionalmente Expuestos (TOE) a las radiaciones ionizantes. Los registros son necesarios para supervisar el cumplimiento de los reglamentos de protección contra las radiaciones; tienen uso médico y legal; permite evaluar las dosis individuales y colectivas recibidas por los trabajadores y posibilitan realizar estudios epidemiológicos y de los efectos sobre la salud. Los servicios de vigilancia radiológica modernos cuentan con un alto grado de automatización, utilizando sistemas integrados que vinculan el mantenimiento de los registros de dosis con el etiquetado, la emisión de dosímetros y su posterior evaluación de dosis.

En nuestro país la vigilancia radiológica individual de la exposición ocupacional a fuentes de radiación externa es realizada por el Centro de Protección e Higiene de las Radiaciones (CPHR). El CPHR es una institución científica perteneciente al Ministerio de Ciencia, Tecnología y Medio Ambiente (CITMA). Tiene la misión de realizar proyectos y servicios competitivos en el campo de la seguridad radiológica y la aplicación de tecnologías que contribuyan a la protección de las personas, el medio ambiente y el desarrollo sostenible del

¹ E-mail del Autor. Manzano@cphr.edu.cu

país [1]. Entre sus funciones se encuentra la de desarrollar, asimilar e implantar técnicas y procedimientos que permitan la evaluación de las dosis de irradiación externa o por contaminación interna que en situaciones normales o accidentales reciben los TOE del país.

Para garantizar el cumplimiento de estas funciones, el CPHR cuenta con un laboratorio provisto del personal calificado y el equipamiento necesario para ofrecer los servicios de control dosimétrico. El Laboratorio de Dosimetría Externa (LDE) atiende todo el servicio de dosimetría personal externa del país y cuenta con dosímetros termoluminiscentes y equipamiento para evaluar las dosis producto de la radiación fotónica [2].

El LDE gestiona un elevado volumen de datos pertenecientes a las entidades usuarias del servicio, los TOE, los resultados de la evaluación de las dosis y los resultados de la evaluación de los dosímetros de los diferentes servicios dosimétricos en cada uno de los períodos de control establecidos. Además, para el procesamiento de los dosímetros se utilizan lectores termoluminiscentes conectados a computadoras personales, los cuales cuentan con su propia base de datos desarrollada por el fabricante para almacenar el resultado de las mediciones.

El objetivo de este trabajo es el de diseñar un Data Mart para el almacenamiento y análisis de los datos como apoyo a la toma de decisiones.

Los Data Mart son almacenes de datos pequeños que están centrados en un tema y responden a los intereses de determinadas áreas de negocio dentro de una organización. Se caracteriza por disponer de la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dichas áreas de negocio. Los datos almacenados en ellos pueden ser agrupados, explorados y distribuidos para que un grupo de usuarios realice la explotación de los mismos y los ayude en el proceso de toma de decisiones. Son sistemas orientados a las consultas, donde la frecuencia de cargas de datos es baja y conocida. A partir de ellos se puede hacer uso de herramientas de Procesamiento Analítico en Línea (OLAP).

Existen un conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en los Data Marts. Estas técnicas reciben el nombre de inteligencia empresarial o inteligencia de negocios.

2. MATERIALES Y MÉTODOS

La metodología utilizada para la construcción e implementación del Data Mart se nombra HEFESTO, la cual puede adaptarse a cualquier ciclo de vida de desarrollo de un sistema. Tiene como objetivo entregar una primera implementación del Data Mart que satisfaga parte de las necesidades y expectativas de los usuarios.

Esta metodología se caracteriza por la facilidad en que se distinguen los objetivos y resultados esperados en cada fase. Está basada en los requerimientos del usuario y es capaz de adaptarse rápidamente y con facilidad a cualquier cambio en el negocio. Por otra parte, al estar el usuario final involucrado en cada etapa del desarrollo, reduce la resistencia al cambio. Se distingue por utilizar modelos conceptuales lógicos sencillos de interpretar y analizar. Además, es independiente del tipo de ciclo de vida de desarrollo de un sistema, de las herramientas utilizadas para su implementación y de las estructuras físicas que contenga el

almacén de datos. Puede aplicarse tanto para el desarrollo de un almacén de datos como para un Data Mart.

La metodología está orientada a la construcción de almacenes de datos para Análisis Dimensional (OLAP) y está formada por las siguientes fases:

1. Análisis de requerimientos.
2. Análisis de los sistemas OLTP.
3. Modelo lógico del Data Mart.
4. Procesos de extracción, transformación y carga de datos (ETL)

Análisis de requerimientos

Se identificaron los requerimientos del usuario a través de una serie de preguntas que explican los objetivos de la organización o área de negocio donde se implementará el Data Mart. Las preguntas incluyen variables de análisis que se consideren relevantes ya que permitirán el estudio de la información desde diferentes perspectivas.

Posteriormente se analizaron estas preguntas con el objetivo de identificar los indicadores y perspectivas que se tomarán en cuenta para la construcción del almacén de datos. Finalmente se confeccionó un modelo conceptual donde se podrá visualizar los resultados obtenidos en este paso de la metodología.

En esta etapa se obtuvieron e identificaron las necesidades de información requerida por los especialistas del LDE para obtener una herramienta que de forma eficiente contribuya a la toma de decisiones. Se formularon preguntas complejas sobre el negocio donde se incluyeron variables de análisis que se consideraron relevantes para poder estudiar la información desde diferentes dimensiones. Algunas de las preguntas de negocio formuladas fueron:

1. ¿Cuál es la dosis media, mínima, máxima y colectiva por servicio y práctica ocupacional para cada institución, ubicación geográfica u organismo central en un período de tiempo determinado?
2. ¿Cuántos trabajadores han sobrepasado los límites de dosis por servicio y práctica ocupacional para cada institución, ubicación geográfica u organismo central en un período de tiempo determinado?
3. ¿Cuál es la cantidad de hombres y mujeres que recibieron controles dosimétricos por servicio y práctica ocupacional para cada institución, ubicación geográfica u organismo central en un período de tiempo determinado?
4. ¿Cuál es el tiempo promedio de demora en la recepción y procesamiento de dosímetros por servicios para cada institución, ubicación geográfica u organismo central en un período de tiempo determinado?

Se procedió a la descomposición de las preguntas de negocio para descubrir los hechos que se utilizarán y las dimensiones de análisis. Para ello se debe tener en cuenta que los hechos, para que sean realmente efectivos son en general, valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: saldos, promedios, cantidades, sumatorias, fórmulas, etc. [3]

A partir de los hechos y dimensiones obtenidas en el paso anterior se procedió a construir el modelo conceptual. Este modelo permite observar cuál es el alcance del proyecto para posteriormente utilizarlo como base del trabajo. Como se puede observar en la Figura 1, a la izquierda se colocan las dimensiones obtenidas las cuales van unidas a un óvalo que representa las relaciones entre ellas. De esta relación se desprenden los hechos que se encuentran colocados a la derecha. Este modelo permite identificar cuáles serán los resultados que se obtendrán, las variables que se utilizarán para analizarlos y la relación que existe entre ellos.

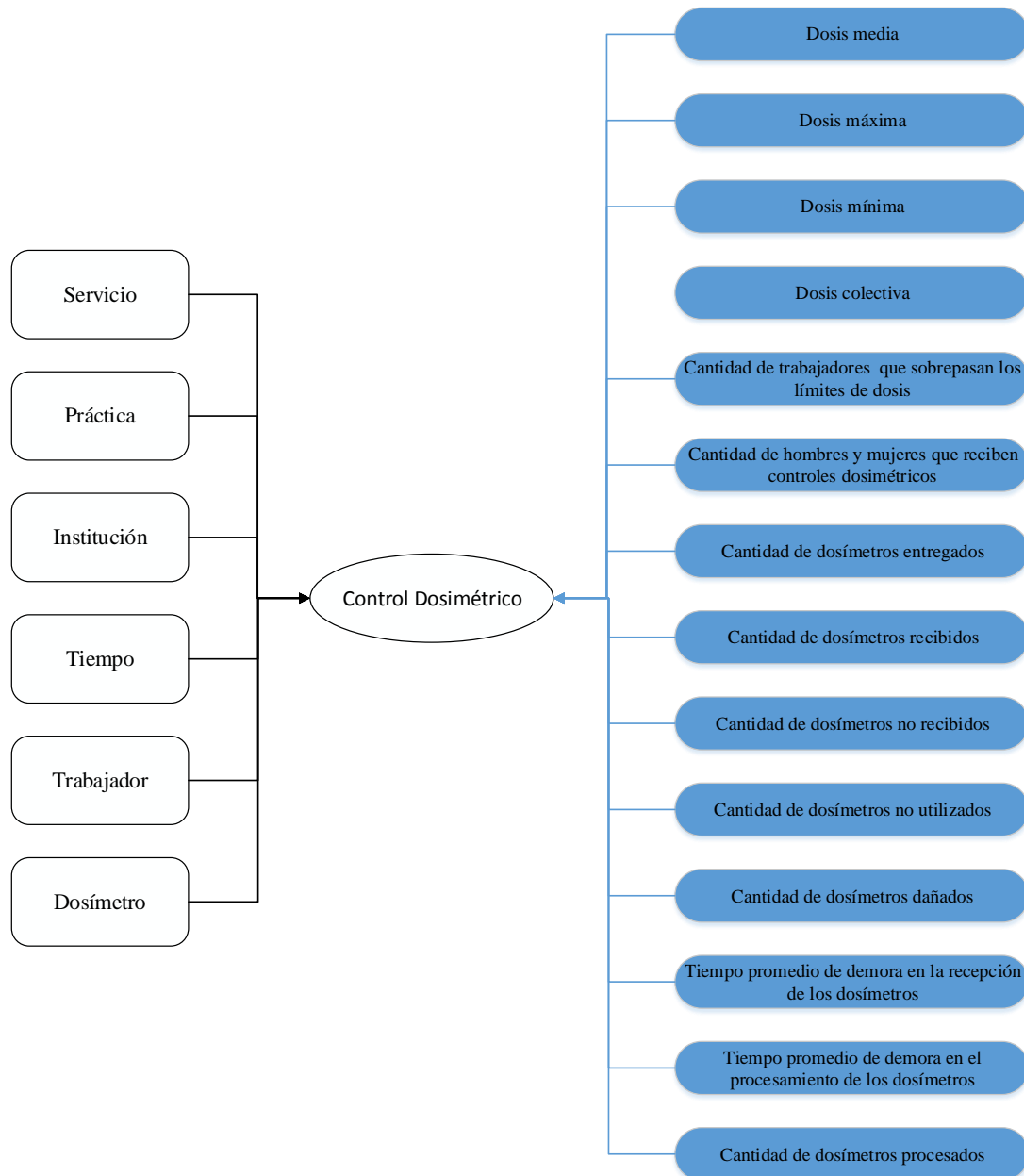


Figura 1. Modelo conceptual del Data Mart

Análisis de los sistemas OLTP

Aunque las bases de datos operacionales y los almacenes de datos tienen en común el hecho de estar formadas por tablas con datos, índices, llaves, vistas y consultas, estas tienen

diferencias significativas. Las bases de datos operacionales están diseñadas y optimizadas para registrar datos, son sistemas de Procesamiento de Transacciones en Línea (OLTP, On-Line Transaction Processing), donde cada transacción debe ser registrada lo más rápidamente posible. Sin embargo, los almacenes de datos están diseñados y optimizados para facilitar la consulta y el análisis de los datos para el soporte en la toma de decisiones. Son sistemas de Procesamiento Analítico en Línea (OLAP, On-Line Analytical Processing) que contienen datos que no se modifican y que pueden ser consultados y analizados más eficientemente que las bases de datos operacionales.

En esta etapa se procedió a examinar el sistema OLTP utilizado por el laboratorio, el cual contienen la información requerida para el diseño del almacén de datos. Se identificaron sus características para que se puedan determinar las correspondencias entre el modelo conceptual confeccionado en la etapa anterior y las fuentes de datos. También se determinaron los indicadores y las fórmulas necesarias para calcularlos.

Posteriormente se examinaron y seleccionaron los campos que contendrá cada perspectiva, ya que será a través de éstos por los que se manipularán y filtrarán los indicadores. Finalmente se ampliará el modelo conceptual con los campos o atributos seleccionados.

En este caso se examina el OLTP utilizado por el LDE para la gestión del servicio con el objetivo de identificar las correspondencias entre el modelo conceptual obtenido anteriormente y la fuente de datos.

Uno de los aspectos más importantes en el diseño de un almacén de datos es el tema de la granularidad y se refiere al nivel de detalle o sumalizaciones de las unidades de datos en un Data Mart.

Modelo lógico del Data Mart

En esta etapa se confeccionó el modelo lógico de la estructura del Data Mart, teniendo como base el modelo conceptual que ya ha sido creado anteriormente. Se selecciona el esquema que se utilizará para contener la estructura del almacén de datos, que mejor se adapta a los requerimientos y necesidades de usuario.

Se diseñaron las tablas de dimensiones que formarán parte del almacén de datos a partir de las dimensiones definidas en el modelo conceptual. También se definieron las tablas que contienen los hechos, a través de los cuales se constituyen los indicadores de estudio. Finalmente se realizarán las uniones correspondientes entre las tablas de dimensiones y las tablas de hechos.

Para la implantación del Data Mart se utilizó MySQL, un popular gestor de base de datos de uso libre, empleado por muchas empresas para la gestión de sus datos y como soporte a herramientas de inteligencia empresarial. Ha sido probada en aplicaciones que requieren de una alta demanda de transacciones y como almacén de datos con volúmenes de almacenamiento del orden de los terabyte [4].

El modelo lógico seleccionado fue el de estrella. Este modelo se caracteriza por una tabla central de hechos con los datos necesarios para el análisis que se encuentra rodeada de las

tablas de las dimensiones. Las tablas de dimensiones tendrán siempre una clave primaria simple, mientras que en la tabla de hechos, la clave principal estará compuesta por las claves principales de las tablas dimensionales. En la Figura 2 se puede apreciar el modelo lógico obtenido.

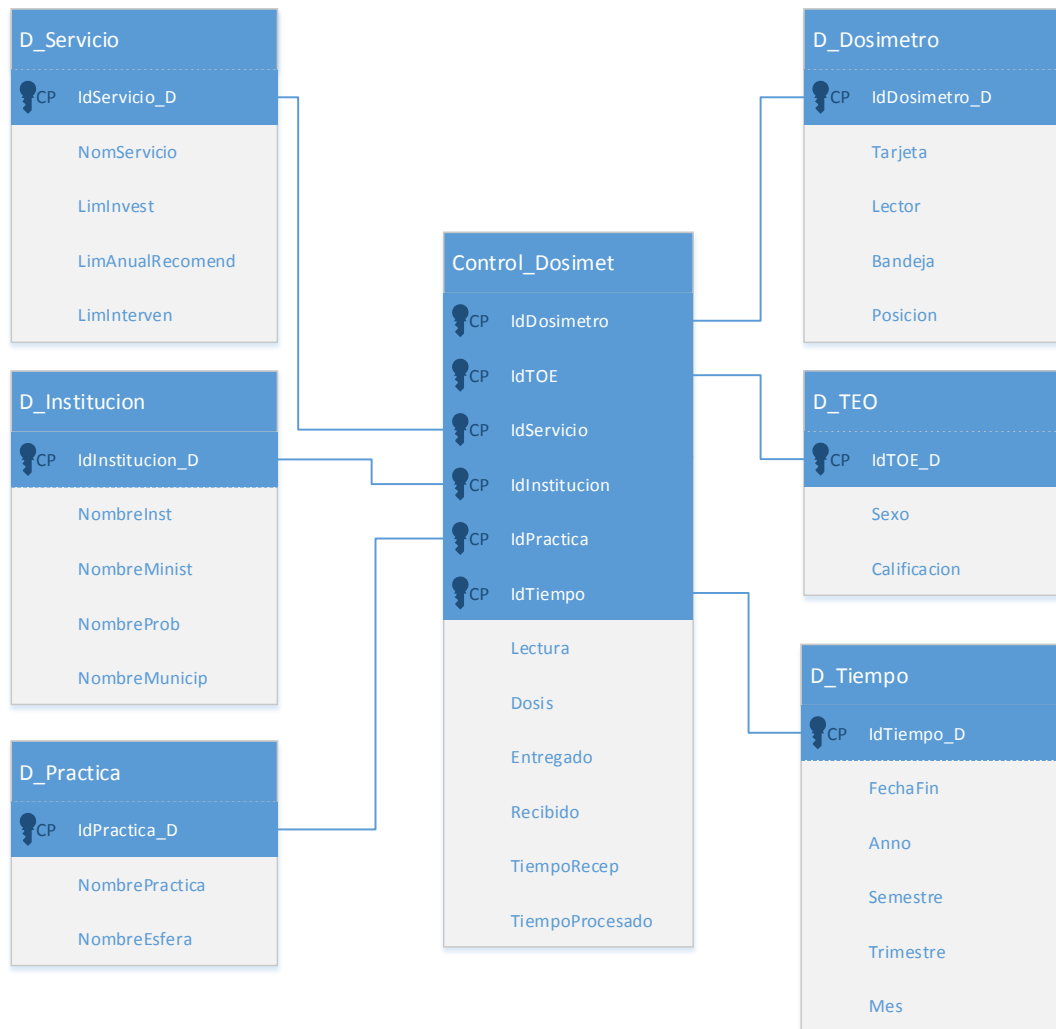


Figura 2. Modelo lógico del Data Mart

Procesos de extracción, transformación y carga de datos (ETL)

En esta etapa se debe realizar el análisis, la definición y el desarrollo de los procesos necesarios para la extracción, transformación y carga de los datos desde los sistemas de origen hasta el almacén de datos. La metodología tiene en cuenta que existen varios software que facilitan esta tarea, por lo que se concentra en el diseño de las sentencias SQL que contendrán los datos que serán de interés.

Se realizó la limpieza de los datos para evitar valores faltantes y anómalos, también tuvo cuenta cual es la información que se desea almacenar estableciendo condiciones adicionales y restricciones. Al finalizar la carga de los datos se establecieron las políticas de actualización y la frecuencia de carga de los mismos.

Con el objetivo de diseñar y probar los cubos de datos que darán respuesta a las necesidades de información identificadas por los especialistas del LDE se realizó una carga inicial con datos pertenecientes al sistema de información Dosis, utilizado desde el año 2000 por el LDE. Se prepararon y cargaron datos correspondientes a los años del 2001 al 2007, los cuales se encontraban almacenados en el sistema OLTP utilizado por el LDE. Los Cubos Multidimensionales tienen la función de representar los datos planos que se encuentran almacenados en filas y columnas, en una matriz de N dimensiones. Mediante la creación y consulta de los mismos se puede tener acceso a los datos almacenados en el Data Mart para explorarlos, analizar las relaciones existentes entre ellos y obtener resultados.

Entre los objetos que se pueden encontrar en un Cubo Multidimensional están los siguientes:

- **Atributos:** Son los campos o criterios de análisis pertenecientes a las tablas de dimensiones, dentro de un Cubo Multidimensional son los ejes del mismo.
- **Indicadores:** Los indicadores son sumalizaciones efectuadas sobre algún hecho o expresiones pertenecientes a una tabla de hechos. El valor que estos tomen estará condicionado por los atributos o jerarquías que se utilicen para analizarlos.
- **Jerarquías:** Representa una relación lógica entre dos o más atributos pertenecientes a un Cubo Multidimensional. Mediante ellas se pueden analizar los datos desde su nivel más general al más detallado y viceversa, al desplazarse por los diferentes niveles.

Las consultas ejecutadas sobre algún Cubo Multidimensional previamente definido son respondidas con gran eficiencia, minimizando al máximo el tiempo que se hubiese incurrido en realizar dicha consulta sobre una base de datos transaccional.

Para el diseño de los Cubos Multidimensionales se utilizó la herramienta Mondrian Schema Workbench [5], la cual forma parte de la suite de inteligencia de negocios Pentaho CE. Mediante esta herramienta es posible construir un esquema que define a una base de datos multidimensional. Este esquema contiene un modelo lógico consistente en cubos con sus atributos, indicadores y jerarquías y es un mapa del modelo físico del Data Mart obtenido.

El esquema se representa en un fichero XML (eXtensible Markup Language) y contiene los elementos necesarios para construir consultas con el lenguaje MDX (MultiDimensional eXpressions).

Los Cubos Multidimensionales diseñados con el objetivo de explorar y analizar los datos contenidos en el Data Mart fueron los siguientes:

- Cubo Multidimensional Dosis.
- Cubo Multidimensional Control Dosimétrico.
- Cubo Multidimensional Trabajadores.

El Cubo Multidimensional Dosis posibilita analizar la información relacionada con las dosis recibidas por los TOE. Como indicadores se definen las dosis máxima, mínima, media y colectiva.

El Cubo Multidimensional Control Dosimétrico analiza la información relacionada con el desempeño de los servicios dosimétricos realizados por el LDE. Como indicadores se definen la cantidad de dosímetros entregados, recibidos y procesados. Además, el tiempo de

recepción y procesamiento de los dosímetros. En el caso del Cubo Multidimensional Trabajadores se analiza la información relacionada con el volumen de trabajadores que reciben los servicios dosimétricos. Como indicadores se define la cantidad de TOE.

3. RESULTADOS Y DISCUSIÓN

Con el objetivo de obtener información del Data Mart diseñado para el LDE se utilizó la herramienta Saiku-Server [6]. Se trata de una herramienta ROLAP (Relational On Line Analytic Processing), se caracterizan por implementar la organización física de los datos sobre tecnología relacional y disponen de facilidades para mejorar el rendimiento de las consultas.

Los Cubos Multidimensionales se generan dinámicamente en el momento en que se realizan las diferentes consultas, haciendo que su manejo sea transparente para el usuario final. Para este proceso el usuario selecciona los indicadores, atributos y jerarquías y de forma dinámica se crea y calcula el Cubo Multidimensional que dará respuesta a la consulta ejecutada. La visualización de los resultados se realiza a través de un navegador web y permite que un usuario sin conocimientos de lenguaje MDX pueda diseñar las consultas de forma visual. Los resultados pueden obtenerse en forma de tablas o mediante diferentes tipos de gráficos como se muestra en la Figura 3.

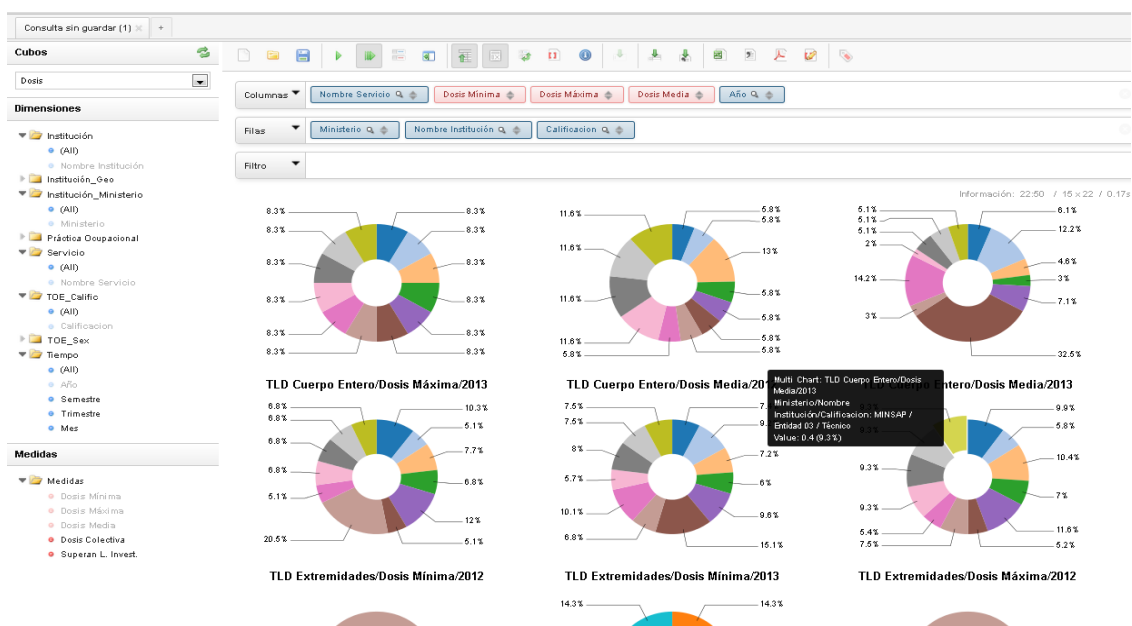


Figura 3. Representación gráfica de los resultados obtenidos

A partir de una serie de consultas diseñadas con la herramienta Saiku-Server se les da respuesta a las preguntas obtenidas en la etapa de diseño del Data Mart, analizando los datos almacenados desde diferentes perspectivas. A modo de ejemplo, en la Tabla 1 se muestran los resultados obtenidos al realizar una consulta al cubo multidimensional “Dosis” para conocer la dosis máxima y media por servicio y calificación técnica para cada ubicación geográfica y Ministerio en el año 2007.

Tabla 1. Información obtenida de una consulta al Cubo Multidimensional Dosis

Nombre Servicio	Ministerio	Provincia	Calificación	2007	
				Dosis Máxima	Dosis Media
TLD Cuerpo Entero	MINSAP	La Habana	Médico	1.72	0.017
			Personal de Enfermería	1.6	0.024
			Técnico en Medicina Nuclear	0.18	0.005
			Técnico en Radioterapia	0	0
			Técnico Rayos X	3.22	0.013
		Matanzas	Médico	0	0
			Técnico en Radioterapia	0	0
		Stgo de Cuba	Médico	0	0
			Personal de Enfermería	0	0
			Técnico en Medicina Nuclear	0	0
			Técnico en Radioterapia	0.35	0.004
			Técnico Rayos X	0.1	0.002
TLD Extremidades	MINSAP	La Habana	Médico	18.45	1.232
			Personal de Enfermería	18.03	1.165
			Técnico en Medicina Nuclear	99.33	3.134
			Técnico Rayos X	7.88	2.909
		Stgo de Cuba	Técnico en Medicina Nuclear	6.37	1.252
			Técnico en Radioterapia	19.94	1.736

En la Tabla 2 se muestran los resultados obtenidos al realizar una consulta al cubo multidimensional “Trabajadores” para conocer la cantidad de hombres y mujeres por servicios a partir del año 2001 hasta el 2007.

Tabla 2. Información obtenida de una consulta al cubo multidimensional Trabajadores

Nombre Servicio	Cantidad TOE													
	2001		2002		2003		2004		2005		2006		2007	
	F	M	F	M	F	M	F	M	F	M	F	M	F	M
Dosimetría Filmica	267	668	4	21										
TLD Cristalino	3	4	5	1	4	2	4	2	5	9	4	4	5	15
TLD Cuerpo Entero	545	731	2,144	2,776	2,625	3,287	2,873	3,420	3,206	3,836	3,837	4,508	4,215	5,008
TLD Extremidades	131	188	131	247	122	248	119	229	137	282	184	336	186	357

En la Tabla 3 se muestran los resultados obtenidos al realizar una consulta al cubo multidimensional “Control Dosimétrico” para conocer la cantidad de dosímetros entregados, recibidos y por ciento de no recibidos de las provincias de Holguín y la Habana por servicio y esfera de aplicación para el año 2007.

Tabla 3. Información obtenida de una consulta al cubo multidimensional Control Dosimétrico

		2007					
		Holguín			La Habana		
Nombre Servicio	Esfera	Entregados	Recibidos	% No Recibido	Entregados	Recibidos	% No Recibido
TLD Cuerpo Entero	Defensa	36	24	33.33	423	283	33.10
	Industria	142	108	23.94	303	238	21.45
	Investigación y Docencia				512	408	20.31
	Medicina	5,698	4,306	24.43	14,698	10,628	27.69
	Otra	297	240	19.19	1,676	1,341	19.99
TLD Extremidades	Defensa	36	24	33.33	382	254	33.51
	Industria	28	20	28.57			
	Investigación y Docencia				183	137	25.14
	Medicina	277	217	21.66	1,009	804	20.32
	Otra	107	85	20.56	844	687	18.60

4. CONCLUSIONES

Con la implementación de un Data Mart se logra almacenar una elevada cantidad de datos relevantes para el LDE a partir del sistema OLTP utilizado para la gestión de los servicios dosimétricos. Se utilizó la metodología HEFESTO y se cargaron datos históricos del servicio para realizar una gran variedad de consultas mediante los Cubos Multidimensionales diseñados y se obtuvieron resultados que sirven de apoyo a la toma de decisiones.

Los resultados obtenidos contribuyen a tener una mejor comprensión del desempeño de los servicios de dosimetría externa en el pasado, monitorear las actividades actuales para responder rápidamente a los cambios, mejorar los procesos de los diferentes servicios, obtención de datos para la revisión de los programas de optimización y pronosticar nuevas posibilidades de ampliación de los servicios. Tanto para el diseño como para la implantación del Data Mart se utilizaron herramientas libres, lo que abarata considerablemente la implementación de este tipo de tecnología en las organizaciones.

5. REFERENCIAS

1. “Centro de Protección e Higiene de las Radiaciones,” <http://www.cphr.edu.cu> (2014)
2. Pernas Salomón, R., & Molina Perez, D. “Pruebas tipo al sistema automático de dosimetría termoluminiscente del CPHR”, *Nucleus* **37**, pp 25-29 (2005).
3. Bernabeu, R.D. *HEFESTO v2.0: Metodología para la Construcción de un Data Warehouse*. Cordoba, Argentina (2010)
4. MySQL. *Enterprise Data Warehousing with MySQL*. MySQL Business White Paper (2007)
5. “Mondrian” <http://sourceforge.net/projects/mondrian/files/schema%20workbench/> (2014)
6. “Saiku Community”, <http://community.meteorite.bi/> (2014)