

DISEÑO DE DATAMART EN LAS TEMATICAS DE PRODUCCIÓN E INVERSIÓN DE
CIENCIA TECNOLOGIA E INNOVACIÓN EN COLOMBIA.

JINNETH TIQUE ORTIZ

ESCUELA COLOMBIANA DE INGENIERIA JULIO GARAVITO
MAESTRIA GESTION DE INFORMACIÓN
PROYECTO DE GRADO
BOGOTÁ D.C.
2016

DISEÑO DE DATAMART EN LAS TEMATICAS DE PRODUCCIÓN E INVERSIÓN DE
CIENCIA TECNOLOGIA E INNOVACIÓN EN COLOMBIA.

JINNETH TIQUE ORTIZ

PROYECTO DE GRADO

ASESOR DE PROYECTO
OSWALDO CASTILLO NAVETTY

ESCUELA COLOMBIANA DE INGENIERIA JULIO GARAVITO
MAESTRIA GESTION DE INFORMACIÓN
PROYECTO DE GRADO
BOGOTÁ D.C.
2016

CONTENIDO

1	JUSTIFICACIÓN	11
2	OBJETIVOS	13
2.1	OBJETIVO GENERAL	13
2.2	OBJETIVOS ESPECIFICOS	13
3	MARCO TEORICO	14
3.1	INTELIGENCIA DE NEGOCIOS	14
3.2	BODEGAS DE DATOS.....	16
3.2.1	Modelo Multidimensional	18
3.2.2	Arquitectura.....	21
3.3	METODOLOGÍA PARA EL DISEÑO DE UNA BODEGA DE DATOS	22
3.3.1	Especificación de Requerimientos para el Diseño de la Bodega de Datos	23
3.3.2	Fases en el Diseño de una Bodega de Datos.....	24
4	ESTADO DEL ARTE.....	31
4.1	RICYT – RED DE INDICADORES DE CIENCIA Y TECNOLOGÍA IBEROAMERICANA E INTERAMERICANA.....	31
4.2	MINISTERIO DE CIENCIA, TECNOLOGÍA Y TELECOMUNICACIONES DE COSTA RICA.....	32
4.3	MINISTERIO DE EDUCACIÓN DE COLOMBIA.....	33
4.3.1	Sistema Nacional de Información de la Educación Superior – SNIES	34
4.3.2	Graduados Colombia – Observatorio Laboral para la Educación.....	35
5	SISTEMA NACIONAL DE CIENCIA Y TECNOLOGIA E INNOVACIÓN – SNCTI.....	37

5.1	DEPARTAMENTO ADMINISTRATIVO DE CIENCIA, TECNOLOGÍA E INNOVACIÓN – COLCIENCIAS	39
5.2	OBSERVATORIO COLOMBIANO DE CIENCIA Y TECNOLOGIA – OCyT	40
5.2.1	Misión.....	40
5.2.2	Visión	40
5.2.3	Historia	40
6	DISEÑO METODOLÓGICO	43
6.1.1	Fase 1: Gestión del proyecto	43
6.1.2	Fase 2: Análisis	43
6.1.3	Fase 3: Diseño	43
6.1.4	Fase 4: Evaluación y revisión de la entrega	43
6.1.5	Detalle de la Estructura de División de Trabajo - EDT	44
6.1.6	Presupuesto	45
7	MODELO CONCEPTUAL DE DATAMART PARA INVERSIÓN Y PRODUCCIÓN DE CIENCIA TECNOLOGÍA E INNOVACIÓN.....	46
7.1	FUENTES DE INFORMACIÓN	46
7.2	ESCENARIOS DE ANALISIS	48
7.3	DIMENSIONES.....	49
7.4	Tablas de Hechos.....	50
7.4.1	Tabla de Hechos de Inversión.....	50
7.4.2	Tabla de Hechos de Producción.....	53
7.4.3	Tabla de Hechos de Investigadores.....	54
7.5	DISEÑO CONCEPTUAL.....	56

7.6	DISEÑO DE ETL.....	57
7.6.1	Dimensión Persona.....	57
7.6.2	Dimensión Programa Académico.....	58
7.6.3	Dimensión Institución	59
7.6.4	Dimensión Área de Conocimiento.....	60
7.6.5	Dimensión ubicación Geográfica.....	60
7.6.6	Dimensión Tiempo.....	61
7.6.7	Dimensión Revista.....	62
7.6.8	Dimensión Producto.....	62
7.6.9	Dimensión Autor.....	63
7.6.10	Dimensión Objetivo Socio Económico	63
8	APLICACIÓN MODELO CONCEPTUAL DE DATAMART PARA INVERSIÓN Y PRODUCCIÓN DE CIENCIA TECNOLOGÍA E INNOVACIÓN	64
8.1	DIMENSIONES.....	64
8.2	DISEÑO CONCEPTUAL.....	66
8.3	DISEÑO DE ETL.....	67
8.3.1	Dimensión Persona.....	67
8.3.2	Dimensión Tesis.....	70
8.3.3	Dimensión Programa Académico.....	70
8.3.4	Dimensión Institución	71
8.3.5	Dimensión Área de Conocimiento.....	72
8.3.6	Dimensión ubicación Geográfica.....	73
8.3.7	Dimensión Tiempo.....	74

8.3.8	Dimensión Revista.....	75
8.3.9	Dimensión Producto.....	76
8.3.10	Dimensión Autor.....	77
8.3.11	Dimensión Objetivo Socio Económico.....	77
8.4	Tablas de Hechos.....	77
8.4.1	Tabla de Hechos de Inversión.....	77
8.4.2	Tabla de Hechos de Producción.....	80
8.4.3	Tabla de Hechos de Investigadores.....	81
9	DISEÑO DE VISUALIZACIÓN.....	82
10	CONCLUSIONES.....	87
	BIBLIOGRAFIA.....	88
	REFERENCIAS ELECTRONICAS.....	90

LISTA DE IMÁGENES

Imagen 1. Infraestructura de Inteligencia de Negocios	15
Imagen 2. Cubo de datos.....	19
Imagen 3. Arquitectura de Bodegas de Datos.....	21
Imagen 4.Arquitectura de Bus.....	23
Imagen 5. Matriz de la arquitectura de bus	23
Imagen 6. Fases del diseño de bodegas de datos	24
Imagen 7. Pasos Diseño Conceptual Enfoque Analysis-driven	25
Imagen 8. Pasos Diseño Conceptual Enfoque Source-driven.....	26
Imagen 9. Pasos Diseño Conceptual Enfoque Analysis/source-driven	26
Imagen 10. Ejemplo esquema estrella	27
Imagen 11. Ejemplo esquema copo de nieve.....	27
Imagen 12. Ejemplo esquema starflake.....	28
Imagen 13. Ejemplo esquema constelación.....	28
Imagen 14. Estándar modelado de ETL	29
Imagen 15Comparativo de Indicadores RICYT	31
Imagen 16 Selección de país para indicadores por país RICYT	32
Imagen 17 visualización de indicadores Colombia.....	32
Imagen 18 Visualización de Indicadores de MICITT	33
Imagen 19 Evolución del SNIES	34
Imagen 20 Visualización de Estadísticas SNIES	35
Imagen 21 Fuentes de Información y Estructura del Sistema de Información del Observatorio Laboral para la Educación desde el Componente de la Oferta	36
Imagen 22 Visualización de indicadores del Observatorio Laboral Colombiano	37
Imagen 23 Estructura de División de Trabajo – EDT	43
Imagen 24 Fase 1 Gestión de Proyecto.....	44
Imagen 25 Fase 2 Análisis	44

Imagen 26 Fase 3 Diseño.....	44
Imagen 27 Fase 4 Evaluación y Revisión de la Entrega	44
Imagen 28 Diseño Conceptual Genérico.....	56
Imagen 29 Dimensión Persona	57
Imagen 30 Dimensión Programa Académico	58
Imagen 31 Dimensión Institución.....	59
Imagen 32 Dimensión Área de Conocimiento	60
Imagen 33 Dimensión Ubicación Geográfica.....	60
Imagen 34 Dimensión Tiempo	61
Imagen 35 Dimensión Revista	62
Imagen 36 Dimensión Producto	62
Imagen 37 Dimensión Autor	63
Imagen 38 Dimensión Objetivo Socioeconómico.....	63
Imagen 39 Diseño Conceptual Aplicado al OCyT.....	66
Imagen 40 Dimensión Persona I.....	67
Imagen 41 Imagen 30 Dimensión Persona II.....	68
Imagen 42 Imagen 30 Dimensión Persona III	69
Imagen 43 Dimensión Programa Académico	70
Imagen 44 Dimensión Institución.....	71
Imagen 45 Dimensión Áreas del Conocimiento.....	72
Imagen 46 Dimensión Ubicación Geográfica.....	73
Imagen 47 Dimensión Tiempo	74
Imagen 48 Dimensión Revista	75
Imagen 49 Dimensión Producto	76
Imagen 50 Dimensión Autor	77
Imagen 51 Dimensión Objetivo Socioeconómico.....	77
Imagen 52 Diseño de Inicio	82
Imagen 53 Diseño de Tablero de Investigadores	82

Imagen 54 Diseño de Tablero de Investigadores por Departamento83

Imagen 55 Diseño de Tablero de Inversión.....83

Imagen 56 Diseño de Tablero de Inversión por Departamento84

Imagen 57 Diseño de Tablero Producción.....85

Imagen 58 Diseño de Tablero Producción por Departamento86

LISTA DE TABLAS

Tabla 1.Comparación de sistemas OLTP y OLAP	17
Tabla 2 Costos Recurso Humano	45
Tabla 3 Costos Recurso Tecnológico	45
Tabla 4 Descripción de Dimensiones.....	49
Tabla 5 Gasto	50
Tabla 6 Financiamiento.....	52
Tabla 7 Producción.....	53
Tabla 8 Investigadores.....	54
Tabla 5 Gasto	77
Tabla 6 Financiamiento.....	79
Tabla 7 Producción.....	80
Tabla 8 Investigadores.....	81

1 JUSTIFICACIÓN

En Colombia existe el Sistema Nacional de Ciencia Tecnología e Innovación - SNCTI¹ que se compone por diversas instituciones con visiones, misiones y objetivos respaldados por las razones que dieron su origen²; Estas instituciones son: Universidades, institutos públicos de investigación, centros de investigación, institutos tecnológicos, centros de desarrollo tecnológico, organismos de cooperación, ministerios, gobernaciones, y todas aquellas instituciones que desarrollen u originen actividades científicas y de investigación.

Alrededor del SNCTI se genera información de diversa naturaleza mucha de ella responde a preguntas como las siguientes: ¿Quiénes realizan actividades de Ciencia Tecnología e Innovación - CTI?, ¿Qué recursos financieros necesita el desarrollo de actividades en CTI?, ¿Qué producción científica tiene el país?, ¿Quiénes hacen producción científica en el país?, ¿Cuánto se invierte en CTI en Colombia?, ¿Qué formación tienen las personas que desarrollan actividades de CTI?, ¿Con qué capacidades de investigación cuenta Colombia?, ¿Cuánto se invierte anualmente en CTI en Colombia?, entre otras. La respuesta a estas preguntas se obtiene de diferentes instituciones y de fuentes de información diversas debido a que cada actor e institución del SNCTI tiene sus propios sistemas de información y por ende su propia información.

Debido a la diversidad anteriormente descrita, el SNCTI cuenta con información heterogénea, descentralizada y fragmentada en cada una de las temáticas que describe la dinámica del mismo, pero no se cuenta con información, integrada y completa del SNCTI.

El Observatorio Colombiano de Ciencia y Tecnología – OCyT, en cumplimiento de su misión ha suplido esta necesidad de información, puesto que se ha encargado de diseñar, producir, integrar, interpretar y difundir estadísticas e indicadores que permiten conocer la dinámica y el posicionamiento del SNCTI colombiano, esto ha

¹Sistema Nacional de Ciencia Tecnología e Innovación - SNCTI definido por la ley 1286 de 2009, como un sistema abierto del cual forman parte las organizaciones públicas, privadas o mixtas que realicen o promuevan el desarrollo de actividades científicas, tecnológicas y de innovación.

²Monroy S. REV. INNOVAR. Vol. 16, No. 28, JULIO A DICIEMBRE DE 2006. Nuevas políticas y estrategias de articulación del sistema de ciencia, tecnología e innovación colombiano.

sido posible porque ha hecho un ejercicio valioso de recolección, integración y validación de información de las diferentes fuentes y actores.

Resultado de ello anualmente genera el libro de indicadores de CTI en el que se evidencia el comportamiento del SNCTI. Sin embargo, para el OCyT esta integración de fuentes y generación de indicadores es dispendiosa y toma tiempo importante de sus investigadores, lo que ha ocasionado retrasos en actividades propias de investigación.

Así mismo, la generación de indicadores en ocasiones no cumple el criterio de oportunidad que los actores del sistema esperan, es decir, el libro de indicadores de CTI en ocasiones supera el tiempo de expectativa de publicación; esto tiene un impacto nocivo en términos de visibilidad para el OCyT.

Por lo anterior, el SNCTI evidencia la necesidad de crear un sistema de información basado en inteligencia de negocios para brindar indicadores precisos, validos, comprensibles, de calidad y potencialmente útiles a distintos niveles de agregación que cumplan el criterio de oportunidad y sirvan como insumo para estudios, análisis y el diseño de políticas de recursos humanos, producción e inversión en ciencia, tecnología e innovación a nivel nacional y regional.

2 OBJETIVOS

2.1 OBJETIVO GENERAL

Diseñar un datamart³ que soporte el análisis multidimensional de la información de producción e inversión en actividades de Ciencia Tecnología e Innovación – CTI.

2.2 OBJETIVOS ESPECIFICOS

- Identificar los requerimientos frente a la producción e inversión en CTI.
- Realizar el análisis de los requerimientos y establecer las necesidades de información para suplirlos.
- Definir los datos e información de inversión y producción en CTI relevantes (importantes) para el análisis y diseño del datamart.
- Construir el diseño conceptual del modelo multidimensional.
- Establecer los mecanismos de validación e integración de la información relevante para las bases de datos seleccionadas en producción e inversión en CTI.

³Datamart: Es una versión especial de Bodega de datos (data warehouse), contienen una visión de datos operacionales que ayudan a decidir sobre estrategias de negocio basadas en el análisis de tendencias y experiencias pasadas. La diferencia principal es que un datamart es específico para una necesidad o área específica del negocio facilitando el acceso a información relevante.

3 MARCO TEORICO

3.1 INTELIGENCIA DE NEGOCIOS

El término “Inteligencia de Negocios” fue adoptado aproximadamente desde el año 1989, cuando Howard Dresner⁴, un investigador de GartnerGroup, uso el acrónimo “BI”⁵, para referirse al conjunto de métodos y conceptos para mejorar la toma de decisiones en los negocios, usando sistemas basados en hechos.

Según Baglietto *et al*⁶ es el proceso de acceder y explotar áreas específicas de la información, analizando la misma, desarrollando nuevas perspectivas y conocimientos, y finalmente aplicando resultados a soluciones empresariales.

Otros autores, como Mark Ritacco y Astrid Carver⁷ plantean la inteligencia de negocios como un proceso que permite a una organización entender, hacer seguimiento y gestionar el negocio para maximizar el desempeño de la empresa; además de permitir mejorar la eficiencia operativa, construir o fortalecer relaciones más rentables con sus clientes y desarrollar ofertas de productos diferenciados.

En términos generales, en la actualidad, la inteligencia de negocios se puede entender como una disciplina compuesta por un amplio conjunto de tecnologías, aplicaciones y metodologías, que permiten reunir, acceder, transformar y analizar datos, transacciones e información no estructurada, con el propósito de desarrollar el entendimiento necesario para tomar decisiones adecuadas. Sin embargo, dichas técnicas, metodologías y aplicaciones por sí mismas no se consideran Inteligencia de Negocios, si estas no están acompañadas de personas que aprovechen la información y la hagan útil para la toma de decisiones no es inteligencia de negocios.

⁴ GARTNER GROUP. Howard Dresner Biography, disponible en línea en http://www.gartner.com/research/fellows/asset_79427_1175.jsp

⁵ Acrónimo de Business Intelligence en el idioma inglés, que traduce Inteligencia de Negocios. Este acrónimo o sigla es ampliamente adoptado para referirse a este tema.

⁶ BAGLIETTO A., BALLESTEROS A., BARCELÓ M., CORREAS J., FERNANDEZ P., GOMEZ S., *et al.* Hacia una economía del conocimiento. PricewaterhouseCooper. 2001.

⁷ RITACCO M., CARVER A. The business value of business intelligence: a framework to measuring the benefits of business intelligence. Business Objects. 2007.

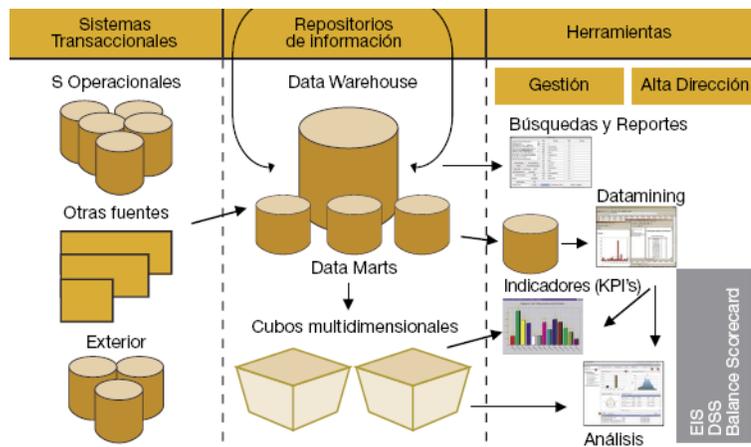


Imagen 1. Infraestructura de Inteligencia de Negocios

Fuente: Boletín de Asesoría Gerencial⁸

Como lo muestra la Imagen 1, la infraestructura de Inteligencia de Negocios se puede entender como un conjunto de capas.

- **Capa de integración de datos:**

Es donde ocurre la extracción, limpieza y transformación de datos, que luego son cargados en la bodega de datos o repositorios de información. Esta capa inicia en los datos y meta-datos de sistemas transaccionales o en información no estructurada.

- **Capa de repositorios de información o bodegas de datos:**

Esta capa consiste en servicios de cubos relacionales y/o OLAP⁹ que permiten a los usuarios de negocio aumentar la percepción en sus respectivas áreas de responsabilidad en la organización.

Para construir correctamente esta capa, es necesario definir requerimientos y preguntas clave del negocio.

- **Capa de aplicaciones:**

Esta es la capa más visible de inteligencia de negocios, debido a que es la encargada de entregar la información a los usuarios del negocio. Incluye la generación y distribución de reportes programados, capacidades de consulta y

⁸ ESPIÑEIRA, SHELDON Y ASOCIADOS. Boletín de Asesoría Gerencial – La Inteligencia de Negocios. PriceWaterhouseCoopers. Nro 10. 2008.

⁹ Acrónimo referente al término en inglés “On-Line Analytical Processing”, Procesamiento Analítico en Línea

análisis para realizar investigaciones u operaciones analíticas, y análisis gráficos que permita la identificación de tendencias, entre otros.

Esta capa permite a los usuarios del negocio interactuar con la información para obtener nuevas perspectivas y percepciones del negocio, mostrando oportunidades, hechos y variables desconocidas del negocio, para apoyar la toma de decisiones.

- **Capa de presentación web (Portales)**

Los portales basados en web, sin ser necesarios en la infraestructura de inteligencia de negocios, han tomado fuerza como un punto único y personalizado de acceso a la información clave del negocio. La mayoría de proveedores de soluciones de inteligencia de negocios han optado por agregar esta capa en sus productos y servicios.

Esta infraestructura de inteligencia de negocios permite a la organización disponer de información integrada de toda la empresa, empoderando a usuarios del negocio a ser autosuficientes de información.

3.2 BODEGAS DE DATOS

En el entorno de la Inteligencia de Negocios, las bases de datos operacionales, conocidas también como sistemas de procesamiento transaccional en línea (OLTP¹⁰), no cumplen con los requerimientos necesarios para responder a escenarios de análisis de datos; esto debido a que su objetivo no es el análisis sino el acceso rápido de múltiples usuarios a los datos, es decir, el procesamiento de transacciones y el control de concurrencia.

Las bodegas de datos permiten a los usuarios realizar un análisis de datos interactivo como apoyo a sus procesos de toma de decisiones, es por eso que esta solución hace parte de los sistemas de procesamiento analítico en línea (OLAP¹¹). Además de lo anterior, los usuarios pueden usar las bodegas de datos para realizar reportes, minería de datos y análisis estadístico.

La siguiente tabla muestra una comparación de sistemas OLTP y OLAP

¹⁰Siglas en inglés para On Line Transaction Processing

¹¹Siglas en inglés para On Line Analytical Processing

Aspecto	OLTP	OLAP
Tipos de usuarios	Operadores, empleados oficinistas	Gerentes, ejecutivos
Uso	Predictivo, repetitivo	No estructurado, bajo demanda
Contenido de los datos	Actual, datos detallados	Histórico, datos resumidos
Organización de los datos	De acuerdo a las necesidades operacionales	De acuerdo al problema de análisis
Estructura de los datos	Optimizado para transacciones pequeñas	Optimizado para consultas complejas
Frecuencia de acceso	Alta	De media a baja
Tipo de acceso	Lectura, Actualización, Borrado, Inserción	Lectura, Agregación
Número de registros por acceso	Pocos	Muchos
Tiempo de respuesta	Corto	Puede ser largo
Nivel de concurrencia	Alto	Bajo
Uso de bloqueos	Necesario	No es necesario
Frecuencia de actualización	Alta	Ninguna
Redundancia de datos	Baja (tablas normalizadas)	Alta (tablas desnormalizadas)
Modelado de datos	Modelo Entidad- Relación	Modelo multidimensional
Modelado e implementación	Total	Incremental

Tabla 1. Comparación de sistemas OLTP y OLAP

Fuente: Advanced Data Warehouse Design¹²

De acuerdo a la infraestructura de inteligencia de negocios planteada anteriormente, las bodegas de datos o repositorios de información son un componente de la segunda capa.

¹²MALINOWSKI E., ZIMANYI E. Advanced Data Warehouse Design. Springer. 2008.

Según Kimball¹³ “la bodega de datos es una colección de datos en forma de una base de datos, específicamente estructurados para consultas y análisis, que guarda y ordena información que se extrae directamente de los sistemas operacionales (ventas, finanzas, producción, marketing, etc.) y de datos externos”.

Una definición más clásica y ampliamente aceptada, es la que propone Bill Inmon¹⁴ “una bodega de datos es una colección de datos integrados, orientados a temas, no volátiles y variables en el tiempo, organizados para soportar necesidades empresariales”.

Esta última definición establece también las cuatro características más importantes de una bodega de datos:

- **Orientada a un tema:** está orientada a un caso o área del negocio en particular, en vez de concentrarse en las transacciones o actividades de la organización.
- **Integrada:** los datos provienen de diversas fuentes y son integrados para dar una visión holística coherente.
- **No volátil:** los datos son persistentes, por más datos que sean agregados, los existentes no son removidos.
- **Variable en el tiempo:** los datos y la información están asociados a un periodo de tiempo específico, debido a la anterior característica, y teniendo en cuenta que los mismos datos con el tiempo pueden tomar valores diferentes.

3.2.1 Modelo Multidimensional

Como lo explica Riccardo Torlone¹⁵ en un capítulo del libro Multidimensional Databases¹⁶: “es ampliamente reconocido que hay al menos dos nociones específicas que cualquier modelo de datos conceptual para bodegas de datos debe incluir de algún modo: el hecho (o su representación usual, el cubo de datos) y la dimensión.”

Un hecho se puede entender como un suceso representado por un valor, que es sujeto de análisis orientado a decisiones y que es usualmente representado gráficamente por medio de un cubo de datos. Una dimensión corresponde a una perspectiva bajo la cual estos hechos son analizados.

Las bodegas de datos comúnmente hacen uso de un modelo de datos conocido como modelo multidimensional o n-dimensional.

¹³ KIMBALL R., ROSS M. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. Wiley. 2002

¹⁴ INMON W. Building the Data Warehouse. Wiley. 2002.

¹⁵ TORLONE R. Conceptual Multidimensional Models. Università Roma Tre. Disponible en línea en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.6570&rep=rep1&type=pdf>

¹⁶ RAFANELLI M. Multidimensional Databases: Problems and Solutions. Idea Group Publishing. 2003. ISBN: 1-59140-086-4.

Los objetivos del modelado multidimensional son: producir estructuras de datos que sean fáciles de entender por usuarios finales y sobre las cuales sea fácil realizar consultas, y maximizar la eficiencia de las consultas.

Los conceptos básicos del modelado multidimensional son:

- **Hechos (facts):** es la colección de ítems de datos relacionados, compuestos por métricas y datos de contexto.
- **Dimensión:** es una colección de datos que determina el contexto para los hechos.
- **Métrica:** es un atributo numérico de un hecho, representando el desempeño o comportamiento del negocio relativo a la dimensión.

Con las diferentes perspectivas que aportan las dimensiones y los hechos, se puede conformar lo que comúnmente se conoce como un cubo de datos, compuesto por tres dimensiones y una medida.

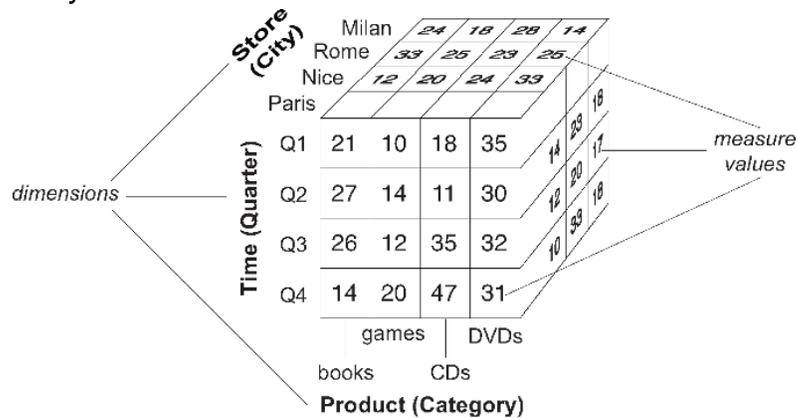


Imagen 2. Cubo de datos

Fuente: Advanced Data Warehouse Design¹⁷

La Imagen 2 muestra un ejemplo de un cubo de datos compuesto por las dimensiones tiempo, almacén y producto, con una medida de cantidades.

¹⁷MALINOWSKI E., ZIMANYI E. Ibid,

3.2.1.1 Jerarquías

Para poder extraer conocimiento estratégico de un cubo, es necesario ver los datos a diferentes niveles de abstracción. Este nivel de abstracción se conoce como granularidad de los datos y está determinado por el nivel de abstracción de cada dimensión.

Las jerarquías permiten realizar un mapeo entre un nivel de abstracción bajo, con conceptos detallados, a un nivel de abstracción más alto, con conceptos generales.

3.2.1.2 Agregación de métricas

Esta agregación ocurre en el momento de subir el nivel de abstracción de un cubo de datos, sumando los datos de las métricas.

Las métricas, según su naturaleza aditiva se pueden clasificar en aditivas, semi-aditivas y no-aditivas.

3.2.1.3 Operaciones OLAP

Estas operaciones explotan y aprovechan las múltiples perspectivas y diferentes niveles de detalle que ofrece el modelo multidimensional y los cubos de datos, generando un entorno de análisis dinámico e interactivo.

Las operaciones OLAP son:

- **Roll-up:** resume métricas detalladas en métricas resumidas.
- **Drill-down:** operación inversa al roll-up, de un nivel resumido pasa a un nivel detallado.
- **Pivot:** rota los ejes del cubo para presentar una presentación alternativa de los datos.
- **Slice:** realiza una selección de solo una dimensión del cubo, resultando en un sub-cubo.
- **Dice:** selecciona dos o más dimensiones definiendo un sub-cubo.
- **Drill - across:** operaciones de consulta que involucran más de un cubo de datos.
- **Drill - through:** permite consultar los datos del cubo en las fuentes de datos de donde se derivaron los datos del cubo.

3.2.2 Arquitectura

La arquitectura de las bodegas de datos se puede separar en tres componentes principales: fuentes de datos, procesamiento o transformación de datos y data marts. Estos componentes a su vez enmarcan los procesos principales de una bodega de datos, que muy someramente se podrían entender como: selección, integración y presentación de datos.

La Imagen 3 muestra la definición de una arquitectura de bodegas de datos altamente aceptada e implementada, que propone una división por capas de los diferentes componentes.

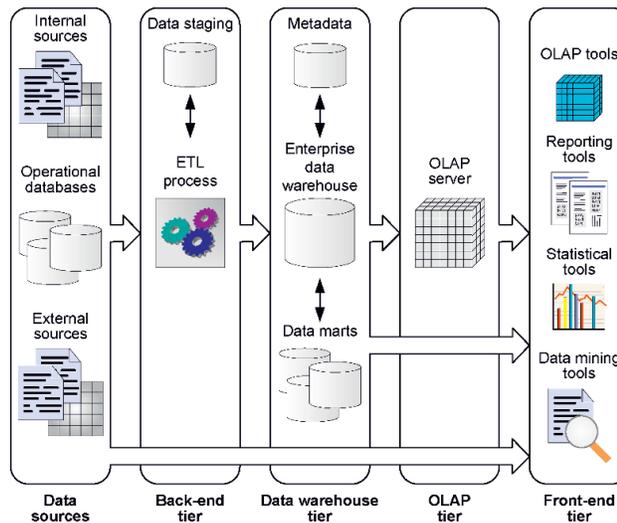


Imagen 3. Arquitectura de Bodegas de Datos

Fuente: Advanced Data Warehouse Design¹⁸

Las fuentes de datos pueden ser heterogéneas en cuanto a la representación de los datos o en los datos mismos. Para traducir modelos de datos heterogéneos en un modelo de datos común, se hace uso de una capa de transformación de datos, que involucra transformaciones que serán aplicadas a los datos fuentes, para finalmente adaptar los datos a la estructura de la bodega de datos.

En el componente de procesamiento o transformación de datos, se puede hacer uso de un área de staging, que es una etapa intermedia entre las fuentes de datos y las bodegas de datos, haciendo uso de ella se está dividiendo el trabajo de transformación en dos fases: en la primera se realiza la integración y en la segunda el resto de trabajo de transformación.

¹⁸MALINOWSKI E., ZIMANYI E. Ibid,

Los Data Marts se proponen como subconjuntos lógicos de la bodega de datos, estos deben ser consistentes en la representación de sus datos para asegurar la robustez de la bodega de datos.

3.3 METODOLOGÍA PARA EL DISEÑO DE UNA BODEGA DE DATOS

Los enfoques para el diseño de bodegas de datos difieren en aspectos como los objetivos, la especificación de requerimientos y las fases del diseño. Existen dos métodos para diseñar las bodegas de datos y los data mart:

- **Diseño Top- down:** Los requerimientos de los usuarios se unen antes del proceso de diseño, y se construye un esquema general para toda la bodega de datos, este esquema debe incluir todos los procesos de cada área del negocio involucrados en el proyecto u organización. Luego se realizan los data mart con las características particulares de cada proceso.
- **Diseño Bottom – up:** Un esquema es realizado para cada data mart, teniendo en cuenta las necesidades descritas por los usuarios de cada área del negocio. Luego se unen los data mart realizados, formando un esquema general de bodega de datos

El uso del enfoque Top – down puede ser costoso en valor y tiempo para la empresa y para el diseñador embarazoso por su tamaño y complejidad. El Bottom – up, permite obtener ganancia en tiempo al ser construido, debido a su tamaño reducido (respecto a una bodega de datos), adicionalmente facilita el proceso de diseño e implementación.

Kimball¹⁹ propone un framework para la integración de los datamart de todas las áreas del negocio, dicho framework recibe el nombre de “data warehouse bus architecture”, en donde las tablas de hechos y las dimensiones compartidas entre los diferentes data mart deben ser conformadas. Una dimensión está conformada cuando es idéntica en todos los data mart; las tablas de hecho son conformadas si tienen la misma semántica, es decir, la misma terminología, granularidad y las unidades de sus medidas en todos los data mart (ver imagen 4 y 5).

¹⁹KIMBALL R., ROSS M. Ibid.

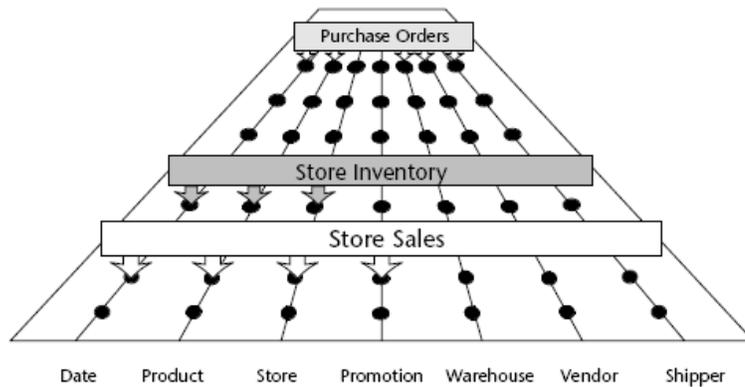


Imagen 4.Arquitectura de Bus

Fuente: The Data Warehouse Toolkit²⁰

BUSINESS PROCESSES	COMMON DIMENSIONS							
	Date	Product	Store	Promotion	Warehouse	Vendor	Contract	Shipper
Retail Sales	X	X	X	X				
Retail Inventory	X	X	X					
Retail Deliveries	X	X	X					
Warehouse Inventory	X	X			X	X		
Warehouse Deliveries	X	X			X	X		
Purchase Orders	X	X			X	X	X	X

Imagen 5. Matriz de la arquitectura de bus

Fuente: The Data Warehouse Toolkit²¹

3.3.1 Especificación de Requerimientos para el Diseño de la Bodega de Datos

Determina qué datos estarán disponibles, cómo serán organizados y qué consultas son de interés. En esta fase se descubren los elementos esenciales del esquema multidimensional, (ej: hechos con las métricas asociadas, dimensiones y jerarquías) que son requeridas para futuros cálculos y manipulaciones de datos.

3.3.1.1 Enfoque dirigido por el usuario (user-driven approach)

El usuario juega un papel fundamental y debe estar activamente involucrado en la definición de hechos y dimensiones relevantes.

²⁰ Ibid.

²¹ Ibid.

3.3.1.2 Enfoque dirigido por el negocio (business-driven approach)

La derivación de la estructura de la bodega de datos parte del análisis de los requerimientos del negocio o de los procesos del negocio. Los elementos que participan en una actividad o servicio del negocio que produce una salida particular, pueden ser considerados como dimensiones en el esquema, y las métricas usadas por los que toman las decisiones para evaluar estas actividades, se pueden considerar métricas de la fact table.

3.3.1.3 Enfoque dirigido a la fuente (source-driven approach)

El esquema se obtiene de analizar los MER de los sistemas fuente. La participación de usuarios no es requerida, aunque pueden especificar requerimientos de información sobre el primer esquema elaborado, seleccionando los ítems de interés.

3.3.1.4 Enfoque combinado

Combina el user-driven o business-driven con el source-driven, teniendo en cuenta que demandan los usuarios o negocio, y que pueden proveer los sistemas fuente.

3.3.2 Fases en el Diseño de una Bodega de Datos

La Imagen 6 muestra las fases del diseño de una bodega de datos, estas pueden ser aplicadas para definir el esquema de los métodos top-down o bottom-up

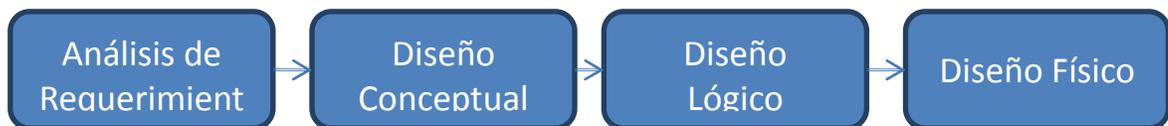


Imagen 6. Fases del diseño de bodegas de datos

3.3.2.1 Análisis de Requerimientos

En esta fase se determina la disponibilidad y organización de los datos, del mismo modo se establecen las consultas de interés. Esta fase lleva a descubrir los elementos esenciales del esquema multidimensional y puede tener los siguientes enfoques:

- **Enfoque dirigido por el análisis (Analysis-driven approach):** Este enfoque dirige el desarrollo del esquema conceptual basado en los requerimientos del usuario o del negocio. Es muy importante identificar los grupos de interés, usuarios, expertos del área de negocio y también el plan empresarial de esta manera se determina el propósito de tener la bodega de datos y las necesidades específicas de análisis.

- **Enfoque dirigido por la fuente (Source-driven approach):** El esquema de la bodega de datos se realiza analizando profundamente los archivos y las bases de datos fuente, para descubrir elementos que puedan representar hechos con sus métricas asociadas, y dimensiones con sus jerarquías.
- **Enfoque “combinado” (Analysis/Source-driven approach):** Combina ambos enfoques, que pueden ser usados en paralelo, donde el resultado de cada actividad es la identificación de elementos para el esquema multidimensional inicial.

3.3.2.2 *Diseño Conceptual*

El diseño conceptual cuenta con los elementos necesarios para el esquema inicial, no depende del enfoque escogido en el análisis de requerimientos.

- **Enfoque dirigido por el análisis (Analysis-driven approach):** Se verifica que los datos requeridos por los usuarios existen en los sistemas fuente antes de desarrollar los esquemas lógico y físico. Durante esta verificación, se describe en detalle las transformaciones requeridas, entre los datos en los sistemas fuente y los elementos en el esquema de la bodega de datos. En la Imagen 7 se pueden observar los pasos de este enfoque.



Imagen 7. Pasos Diseño Conceptual Enfoque Analysis-driven

- **Enfoque dirigido por la fuente (Source-driven approach):** Debido a que no todos los hechos (facts) son de interés para el propósito del soporte a decisiones, se deben identificar que hechos son importantes apoyados en los usuarios. El esquema inicial de la bodega de datos es modificado hasta llegar a una versión final aceptada por los usuarios. En la Imagen 8 se pueden observar los pasos de este enfoque.



Imagen 8. Pasos Diseño Conceptual Enfoque Source-driven

- **Enfoque combinado (Analysis/Source-driven approach):** Combina ambos enfoques, que pueden ser usados en paralelo. En la Imagen 9 se pueden observar los pasos de este enfoque.

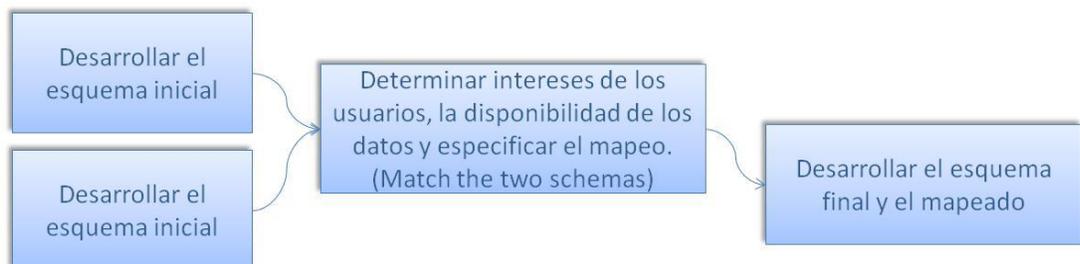


Imagen 9. Pasos Diseño Conceptual Enfoque Analysis/source-driven

3.3.2.3 Diseño Lógico

En el diseño lógico se deben tener en cuenta dos aspectos: la transformación del esquema conceptual multidimensional, y la especificación del proceso ETL considerando los mapeos y transformaciones indicadas en la fase anterior.

- **Transformación del esquema conceptual multidimensional:** Es necesario aplicar reglas de mapeo sobre el esquema conceptual resultante para obtener el esquema lógico. La elección entre un modelo relacional y un modelo de objetos depende en las necesidades de la aplicación. Cuando se hace esta elección, es importante considerar el diseño a nivel físico y las características específicas del DBMS objetivo. Existen diferentes esquemas tales como:

Esquema en Estrella: En un esquema en estrella, solo hay una tabla de hechos central y un conjunto de tablas dimensión, una para cada dimensión, existen restricciones de integridad referencial entre la tabla de hechos y cada dimensión. En este esquema puede contener redundancia. La imagen 10 muestra un ejemplo de este esquema.

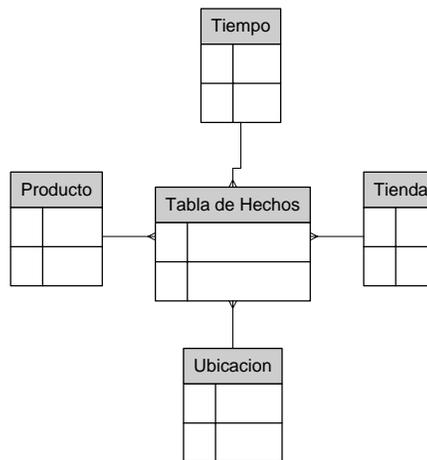


Imagen 10. Ejemplo esquema estrella

Esquema Copo de Nieve: Se evita la redundancia en el esquema copo de nieve normalizando las dimensiones. Una dimensión está representada por varias tablas relacionadas por restricciones de integridad referencial, también se relacionan con la tabla de hechos y las dimensiones en máximo nivel de detalle. Una desventaja de este esquema es el rendimiento debido a que debe realizar joins para recorrer las jerarquías. La imagen 11 muestra un ejemplo de este esquema.

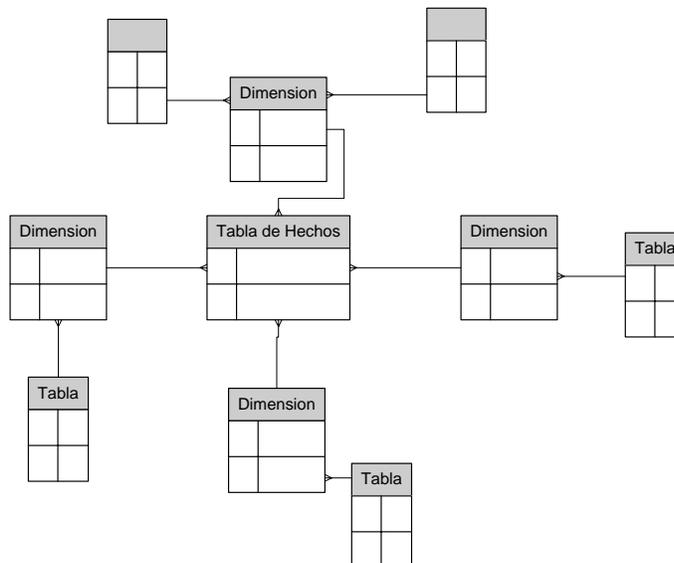


Imagen 11. Ejemplo esquema copo de nieve

Esquema Starflake: Es una combinación de los esquemas en estrella y copo de nieve, en donde algunas de las dimensiones son normalizadas mientras otras no. La imagen 12 muestra un ejemplo de este esquema.

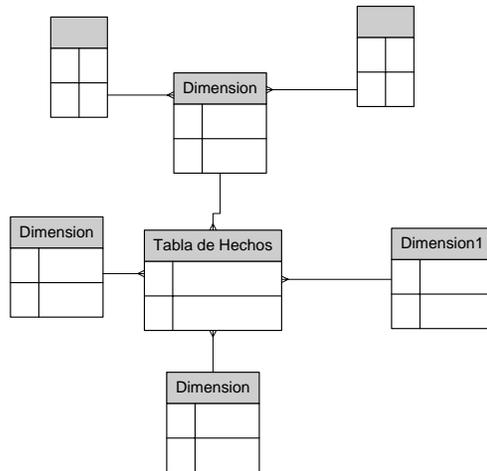


Imagen 12. Ejemplo esquema starflake

Esquema Constelación: Este esquema tiene varias tablas de hecho que comparten las dimensiones. La imagen 13 muestra un ejemplo de este esquema.

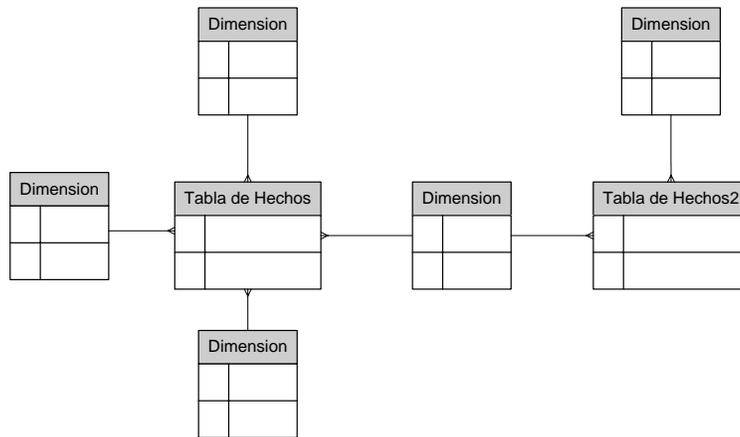


Imagen 13. Ejemplo esquema constelación

- **Especificación del proceso ETL:**

Los ETL, son los procesos de Extracción, Transformación y Carga, cuyo objetivo es transferir datos desde las aplicaciones de producción y múltiples fuentes a los DWH. La construcción del ETL es exigente, puesto que tienen que estar unidos a las necesidades del negocio, formatos y deficiencias de los datos orígenes y entregar los datos de manera eficiente y confiable a la bodega de datos.

El primer paso para modelar un ETL es la Planeación y el diseño, para ello se deben registrar todas las realidades y necesidades del negocio, tales como:

- Las necesidades de la empresa.
- Realidades de los datos y otras fuentes de datos
- Cumplimiento de los requerimientos de usuario y de seguridad.
- Integración de datos y datos de latencia
- Interfaces de usuario

Luego de tener claras las necesidades y realidades del negocio se debe iniciar la definición de la arquitectura para el ETL; este paso incluye decisiones como, manejo de excepciones, flujo de datos, calidad, recuperación de información o del proceso en caso de fallos y seguridad de la información.

Para realizar el modelado de un ETL se aconseja usar el estándar mostrado en la Imagen 14, la lógica del proceso ETL debe ser desarrollada antes de ser implementado.

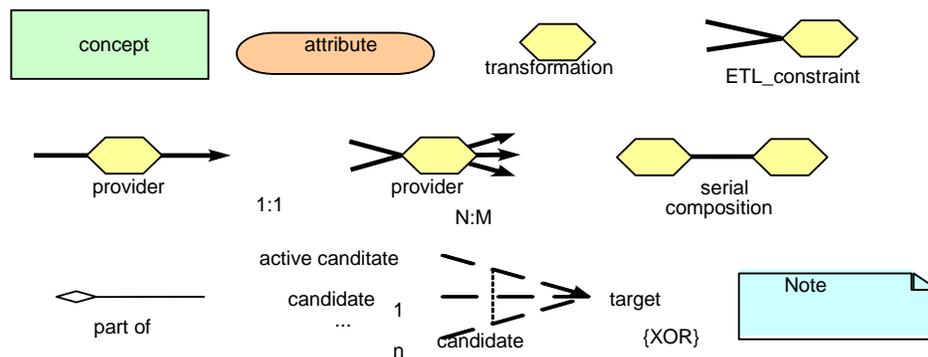


Imagen 14. Estándar modelado de ETL

La implementación del sistema de ETL se debe realizar teniendo en cuenta el hardware, software, prácticas de codificación y documentación y los controles de calidad. Finalmente, se realiza la administración del ETL en donde se debe evaluar y ajustar el rendimiento, los procedimientos de reversión, actualización y modelo de propagación.

Un ETL debe asegurar la existencia de toda la información de las fuentes de datos en la bodega de datos, asegurar la calidad, consistencia, precisión, integridad y eliminar la duplicación de datos. Para garantizar la confiabilidad de la bodega de datos.

3.3.2.4 Diseño Físico

En el diseño físico se deben tener en cuenta dos aspectos:

- **Aplicación del esquema de bodega de datos:** Se deben considerar las decisiones del diseño lógico y las consultas o solicitudes descritas en la especificación de requerimientos. La esencia del diseño físico es contribuir a mejorar el acceso a los datos, el rendimiento en las consultas, el mantenimiento de la bodega de datos y el proceso de carga de datos.
- **Aplicación de procesos ETL:** Todos los procesos de extracción, transformación y carga deben ser definidos de manera detallada y estar listos para ser ejecutados, no solo para la carga inicial de la bodega de datos, sino también para las actualizaciones posteriores. Para este último es recomendable realizarlo de tal forma que solo se carguen los datos nuevos y los que hayan cambiado desde la última actualización.

De acuerdo a la forma como se almacene la bodega de datos se define el enfoque del modelo para la aplicación multidimensional:

- **Relational OLAP – ROLAP:** Servidores de datos en bases de datos relacionales y apoyo de extensiones SQL, con métodos de acceso para modelos de datos multidimensionales y operaciones relacionadas. En estos sistemas los datos multidimensionales se implementan como tablas relacionales organizadas en estructuras de especializadas como los esquemas de estrella, starflake y constelación.
- **Multidimensional OLAP – MOLAP:** Servidores que almacenan directamente estructuras de datos multidimensionales y aplican operaciones OLAP sobre las estructuras de datos. Los sistemas MOLAP tienen capacidad de almacenamiento inferior a los sistemas OLAP, sin embargo su rendimiento es superior cuando se realizan consultas de datos multidimensionales.
- **OLAP Híbrido – HOLAP:** Servidores que combinan la tecnología MOLAP y ROLAP, aprovechando la capacidad de almacenamiento ROLAP y las capacidades de procesamiento MOLAP.

4 ESTADO DEL ARTE

4.1 RICYT – RED DE INDICADORES DE CIENCIA Y TECNOLOGÍA IBEROAMERICANA E INTERAMERICANA

La red de indicadores de Ciencia y Tecnología – Iberoamericana e Interamericana – RICYT, nació a partir del primer taller Iberoamericano sobre Indicadores de Ciencia y Tecnología, realizado en Argentina en 1994. Su puesta en marcha se realizó en abril de 1995. El objetivo principal de la RICYT es *promover el desarrollo de instrumentos para la medición y análisis de la ciencia y la tecnología en Iberoamérica, en un marco de cooperación internacional, con el fin de profundizar en su conocimiento como instrumento político para la toma de decisiones*. En su página web <http://www.ricyt.org/indicadores> tiene una sección de indicadores en donde se pueden visualizar tablas de los diferentes indicadores en comparativos o por países.

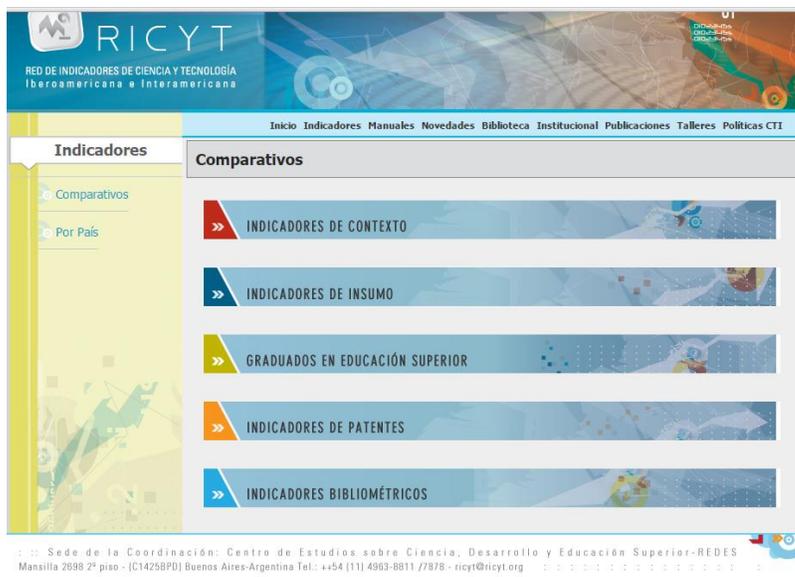


Imagen 15 Comparativo de Indicadores RICYT



Imagen 16 Selección de país para indicadores por país RICYT

Red de Indicadores de Ciencia y Tecnología -Iberoamericana e Interamericana - RICYT
[Descargar Excel](#) / [Download Excel](#)

Colombia												C											
Indicador	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	U
Población																							
millones de personas	34.96	35.60	36.40	37.12	37.64	38.54	38.07	38.64	39.20	39.75	40.28	40.80	41.32	41.84	42.36	42.88	43.40	43.92	44.45	44.97	45.50	46.04	n
Población económicamente activa (PEA)																							
millones de personas	15.22	15.49	15.76	16.10	16.32	16.83	16.52	16.69	16.92	17.10	17.59	19.14	19.11	19.75	19.38	19.71	18.79	19.78	19.67	21.69	22.16	23.31	n
PBI US\$																							
PBI millones de US\$	40274.20	42514.28	49259.30	55825.93	70152.31	80525.43	97147.06	106671.13	98443.85	86186.01	99898.53	90205.57	97818.09	94684.58	117074.86	146520.13	162773.60	207520.15	244056.73	233821.67	266398.24	333164.80	C
PBI en PPC																							
PBI millones de US\$ expresados en PPC	126236.40	133626.95	143582.58	150356.78	139511.29	151853.70	181403.12	190761.11	193978.79	168508.99	236377.15	246759.67	257649.22	305637.54	330762.31	357437.99	393101.48	431380.30	455406.34	466490.69	489829.48	529965.36	C

Notas:
 - PPC: Corresponde a Paridad de Poder de Compra. Las estimaciones en Paridad de Poder de Compra fueron obtenidas aplicando los factores de conversión del Banco Mundial sobre la información en moneda local provista por cada país.
 - PPC: Paridad de Poder de Compra.
 - Las estimaciones en dólares fueron obtenidas aplicando los datos de Tipo de Cambio del Banco Mundial sobre la información en moneda local, provista por el país.

Gasto en Ciencia y Tecnología (US\$)												C											
Indicador	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	U
ACT (millones de dólares corrientes)						441.92	546.90	561.40			307.20	170.81	298.67	392.03	519.00	684.44	690.50	994.49	1252.69	1122.16	1434.36	1689.73	C
I+D (millones de dólares corrientes)						236.39	290.69	291.91			105.92	107.48	113.44	170.79	206.28	255.72	281.02	411.99	516.84	488.38	601.90	727.66	R

Notas:
 - América Latina y el Caribe: Los datos son estimados.
 - Iberoamérica: Los datos son estimados.
 - ACT: Corresponde a Actividades Científicas y Tecnológicas.

Gasto en Ciencia y Tecnología (PPC)												C											
Indicador	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	U

Imagen 17 visualización de indicadores Colombia

4.2 MINISTERIO DE CIENCIA, TECNOLOGÍA Y TELECOMUNICACIONES DE COSTA RICA

El departamento de Indicadores del ministerio de Ciencia Tecnología y Telecomunicaciones de Costa Rica – MICITT realiza recolección de información primaria sobre inversión en I+D, Recursos Humanos I+D, TICS en empresas, Actividades de Innovación; esta información es presentada en su página web <http://indicadores.micit.go.cr/dashboards/7953/inversion-en-act-e-id-2011/> a través de gráficas y tablas visualmente entendibles y amenas; no es una herramienta de

inteligencia de negocios. Sin embargo, provee los indicadores de Costa Rica en materia de Ciencia, Tecnología e Innovación.

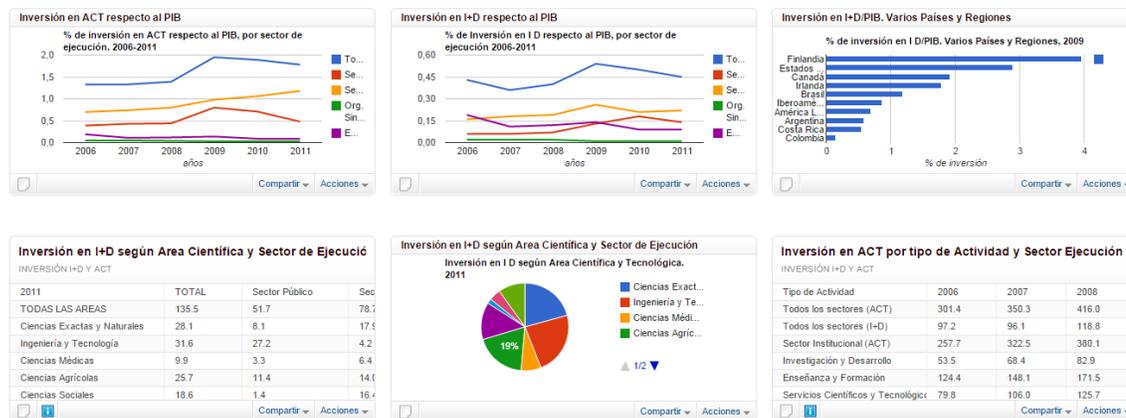


Imagen 18 Visualización de Indicadores de MICITT

4.3 MINISTERIO DE EDUCACIÓN DE COLOMBIA

El Ministerio de Educación Nacional fue creado mediante la ley 7ª de agosto 25 de 1886, con el fin de preparar, proponer y dictar las normas para la organización y fortalecimiento de la gestión del sector educativo, formular la política nacional para el fomento de la Educación Superior. Así mismo, dirigir el Sistema Nacional de Información Educativa, entre otros propósitos.

A partir de la promulgación de la Ley 30 de 1992 se da inicio a la formalización de la información de la Educación Superior en Colombia mediante la creación de un sistema de información con el objetivo principal de convertirse en el medio de divulgación de las características, la calidad y la cantidad de las instituciones y programas en el país.

El Instituto Colombiano para el Fomento de la Educación Superior emprendió las acciones pertinentes para formular el diseño y llegara la puesta en marcha del Sistema Nacional de Información de la Educación Superior - SNIES, el cual surtió diferentes etapas de evaluación y validación por parte de expertos. En mayo de 2002 se culminó un primer rediseño que obedeció a las necesidades del sector y su dinámica propia.²²

²²http://redes.colombiaaprende.edu.co/ntg/men/pdf/Documento_Metodologico_SNIES_2012.pdf



Imagen 19 Evolución del SNIES

4.3.1 Sistema Nacional de Información de la Educación Superior – SNIES

El Sistema Nacional de Información de la Educación Superior - SNIES es un sistema de información creado para responder a las necesidades de información sobre la educación superior en Colombia.

Esta fuente de información consolida y suministra datos, estadísticas e indicadores relevantes, confiables sobre las instituciones y programas académicos aprobados por el Ministerio de Educación Nacional.

Es así como el Centro Andino de Altos Estudios - CANDANE, responsable de la formación estadística del Departamento Administrativo Nacional de Estadísticas (DANE), ha otorgado al SNIES la máxima certificación que se confiere 'Grado A', al concluir que el proceso de gestión de la información del sistema en el nivel institucional, regional y nacional, cumple con los requisitos de calidad, confiabilidad y oportunidad.

Dicha certificación fue concedida por el proceso que incluye el diseño del software, la captura, la consistencia y la validación de las cifras.

Asimismo en la Conferencia Regional de Educación Superior (CRES 2008) que se llevó a cabo en Cartagena de Indias, el Instituto de Educación Superior para América Latina y el Caribe (IESALC) exaltó al SNIES como uno de los mejores sistemas de América Latina y se convirtió en modelo para la construcción del Mapa de la Educación Superior de América Latina y el Caribe - MESALC.



Imagen 20 Visualización de Estadísticas SNIES

4.3.2 Graduados Colombia – Observatorio Laboral para la Educación

El Observatorio Laboral para la Educación es un sistema de información que brinda información para analizar la pertinencia de la educación a partir del seguimiento a los graduados y su empleabilidad en el mercado laboral. Pretende contribuir al mejoramiento de la calidad de los programas académicos ofrecidos en Colombia.

Desde el año 2005, hace seguimiento a los graduados de la educación superior, y más recientemente de aquellos certificados en competencias laborales, para conocer cómo se insertan y son acogidos por el mercado laboral colombiano.

Ofrece información estadística sobre el nivel de formación académica de los egresados, sus aportes a seguridad social y los salarios promedio que reciben inicialmente. Además, entrega un panorama sobre cuánto tiempo les toma conseguir empleo, las ciudades en las que trabajan y la demanda de egresados que tienen en el mercado laboral una mayor o menor acogida.

El observatorio Laboral cuenta con dos componentes de información: Uno de oferta y otro de demanda.

- Desde el componente de la oferta: El sistema de información se fundamenta en la integración entre las bases de datos de graduados provenientes del Sistema Nacional de Información de Educación Superior

(www.mineduacion.gov.co/snies) (SNIES), las bases de datos de certificados del Sistema de Información de Educación para el Trabajo (www.mineduacion.gov.co/siet) (SIET) y las bases de datos del Ministerio de Salud y Protección Social y la Unidad de Gestión Pensional y Parafiscales (UGPP), entidad adscrita al Ministerio de Hacienda y Crédito Público.

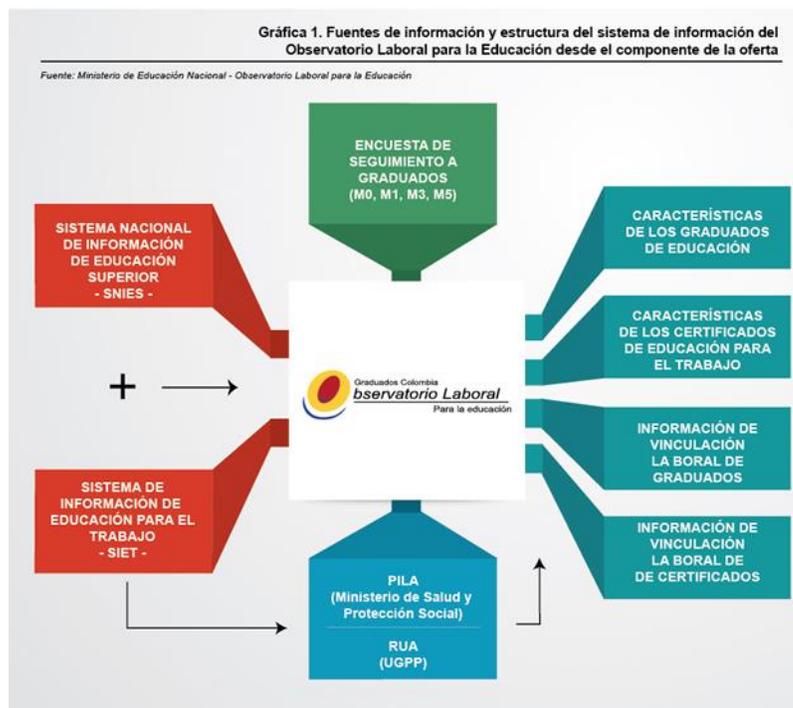


Imagen 21 Fuentes de Información y Estructura del Sistema de Información del Observatorio Laboral para la Educación desde el Componente de la Oferta

El componente de demanda, está dividido en dos direcciones. En primer lugar, se realiza una encuesta a empleadores, gracias a la cual es posible tener información acerca de cuáles son las competencias que deben tener sus empleados, cuál es el nivel de satisfacción del desempeño de los recién graduados al interior de sus empresas y cómo es su contratación. En segundo lugar, se realizan estudios que dan cuenta del capital humano requerido desde los diferentes sectores productivos. El Observatorio Laboral participa en el análisis de estudios de prospectiva laboral cualitativa sectorial con el fin de identificar cuáles serán los perfiles ocupacionales, competencias y necesidades de formación claves para las distintas regiones y sectores económicos de Colombia en el futuro.

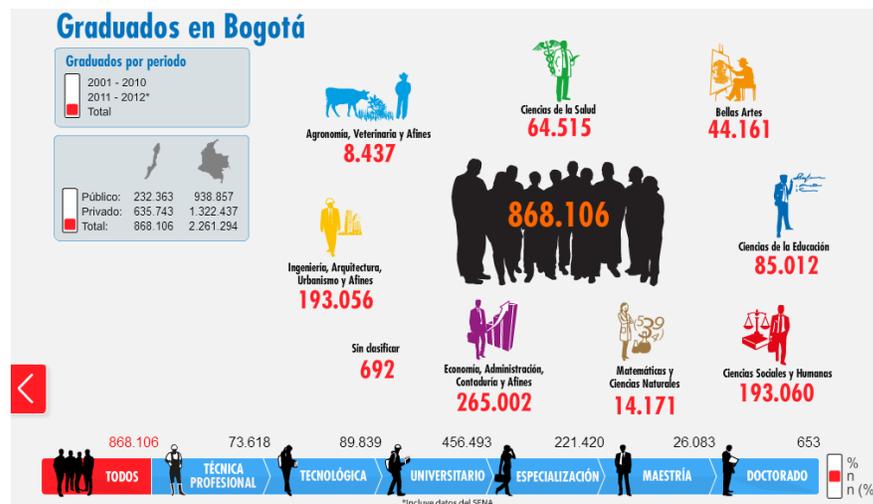


Imagen 22 Visualización de indicadores del Observatorio Laboral Colombiano

5 SISTEMA NACIONAL DE CIENCIA Y TECNOLOGIA E INNOVACIÓN – SNCTI

En Colombia, el Sistema Nacional de Ciencia y Tecnología fue creado por la ley 29 de 1990, sin embargo para llegar a ser lo que es hoy en día ha pasado por un proceso desde el año 1968, a continuación se relata una síntesis de la creación del sistema.

En el periodo comprendido desde el año 1940 hasta el año 1968, Colombia se ve influenciada por organismos internacionales como la Organización de Estados Americanos - OEA, el Banco Interamericano de Desarrollo- BID y la Agencia Internacional para el Desarrollo – AID para diseñar e implementar políticas de desarrollo tales como: reforma agraria, educativa y del estado.

Del mismo modo se inició la creación de institutos estatales descentralizados de educación e investigación, entre ellos: Icetex, Instituto de Investigaciones Tecnológicas, Instituto Colombiano Agropecuario - ICA, Servicio Nacional de Aprendizaje - SENA, Instituto Colombiano de la Reforma Agraria - Incora, Instituto de Asuntos Nucleares.

En el año 1968 bajo el decreto 2869 se crean el Consejo Nacional de Ciencia y Tecnología y el Fondo Colombiano de Investigaciones Científicas y Proyectos Especiales “Francisco José de Caldas” conocido como COLCIENCIAS.

En el año 1987 se realiza el primer Foro Internacional sobre Política de Ciencia y Tecnología.

El año 1988 el decreto 595 declara este año como el año Nacional de la Ciencia y la Tecnología y el decreto 1600 integra la misión de ciencia y tecnología y señala sus funciones.

En el año 1990 con la ley 29 se define la Política Nacional de Ciencia y Tecnología, el decreto 585 establece la creación del Consejo Nacional de Ciencia y Tecnología, se organiza el Sistema de Ciencia y Tecnología - SNCyT y de las Comisiones Regionales de Ciencia y Tecnología; este mismo decreto adscribe al Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología Francisco José de Caldas- Colciencias al Departamento Nacional de Planeación – DNP.

En el año 1992 se expide la ley 6a en donde se establecen Incentivos Tributarios a la Ciencia y Tecnología.

El decreto 2739 del año 1994 aprueba el primer Conpes de Ciencia y Tecnología, en este mismo año se crea la Comisión Nacional de Doctorados y Maestrías.

En el año 1999 se Crea el Observatorio Colombiano de Ciencia y Tecnología – OCyT. “El OCyT es una institución del Sistema Nacional de Ciencia y Tecnología - SNCTI dedicada a producir conocimiento sobre la dinámica y el posicionamiento del sistema mediante el diseño, producción, integración, interpretación y difusión de estadísticas e indicadores, para orientar y evaluar las políticas y la acción de los diversos actores del SNCTI. Los beneficiarios del OCyT son los miembros de la comunidad científica nacional y, en especial, las autoridades que tienen la responsabilidad de adoptar las estrategias y de implementar y evaluar las políticas que guían las actividades en los campos de la ciencia, la tecnología y la innovación, al igual que las universidades, centros de investigación y desarrollo Tecnológico y empresas que llevan a cabo dichas tareas y en general, los actores del SNCTI”²³.

En el periodo 2000 – 2002 se crea el documento Conpes 3080 en donde se establece la Política de Ciencia y Tecnología, en el año 2001 se crea el Programa

²³ http://docs.politicascsti.net/reportes/CO_SI.pdf

de Prospectiva Tecnológica y se expide la Ley 643 del Fondo de Investigación en Salud. En el año 2002 se conforman las Agendas Regionales de Ciencia y Tecnología. Y se hace el lanzamiento Plataforma ScienTI (2002).

En el año 2003 mediante la ley del Plan Nacional de Desarrollo se Incorporan recursos a la Ciencia y Tecnología haciendo uso de la Ley 344 de 1996.

En el año 2005 se hace la reforma de los Programas Nacionales de Ciencia, Tecnología e Innovación, hacia Áreas de Conocimiento.

En el año 2008 se hace reevaluación de la estrategia CNCyT y se Promulga el documento: “Colombia Construye y Siembra Futuro. Política Nacional de Fomento a la investigación y la innovación”, por Colciencias.

En el año 2009 la ley 1286 transforma a Colciencias en Departamento Administrativo de Ciencia, Tecnología e Innovación, crea el Sistema Nacional de Ciencia, Tecnología e Innovación - SNCTI y dicta otras disposiciones sobre estas materias.

5.1 DEPARTAMENTO ADMINISTRATIVO DE CIENCIA, TECNOLOGÍA E INNOVACIÓN – COLCIENCIAS

Colciencias es el ente gubernamental que promueve las políticas públicas para fomentar la CT+I en Colombia. Sus actividades implican concertar políticas de fomento a la producción de conocimientos, construir capacidades para CT+I, y propiciar la circulación para el desarrollo integral del país y el bienestar de los colombianos.

Así mismo, coordina el Sistema Nacional de Ciencia, Tecnología e Innovación-SNCTI, crea sinergias e interacciones para que Colombia cuente con una cultura científica, tecnológica e innovadora. Y define los programas estratégicos para el desarrollo del país en pro del conocimiento producido por los investigadores e innovadores del país, aunando esfuerzos y aprovechando alianzas estratégicas y cooperación internacional.

Colciencias en su iniciativa Estado de la ciencia en Colombia crea una herramienta al público en donde se pretende visualizar la información estadística

correspondiente a la Convocatoria 693 de 2014 y Convocatoria 737 de 2015 distribuida en 6 pestañas.

5.2 OBSERVATORIO COLOMBIANO DE CIENCIA Y TECNOLOGIA – OCyT

“La asociación se denomina Observatorio de Ciencia y Tecnología y podrá usar indistintamente el nombre abreviado de El Observatorio o la sigla OCyT. Es una asociación civil de participación mixta y de carácter privado, sin ánimo de lucro, con patrimonio propio organizada bajo las leyes colombianas dentro del marco de la Constitución Política y las normas de Ciencia y Tecnología y regida por ellas, en especial por las regulaciones previstas para las corporaciones en el Código Civil y por sus Estatutos”²⁴.

El OCyT fue constituido el 13 de agosto de 1999 e inscrito ante la Cámara de Comercio de Bogotá el 19 de octubre de 1999 bajo el número 00026411, del libro I de las entidades sin ánimo de lucro.

5.2.1 Misión

El OCyT es una institución del Sistema Nacional de Ciencia, Tecnología e Innovación (SNCTI) dedicada a producir conocimiento sobre la dinámica y el posicionamiento del sistema mediante el diseño, producción, integración, interpretación y difusión de estadísticas e indicadores, para orientar y evaluar las políticas y la acción de los diversos actores del SNCTI.

5.2.2 Visión

Seremos la institución líder en Colombia, referente nacional e internacional a partir de un trabajo en red, en la producción, interpretación y difusión de estadísticas e indicadores de CTI, y en la definición de los modelos y metodologías para el manejo de información requerida por el SNCTI.

5.2.3 Historia

²⁴ Tomado de <http://ocyt.org.co/es-es/informacion-institucional>

En los primeros años de la década de 1990 nace la Red Iberoamericana de Indicadores de Ciencia y Tecnología (RICyT), auspiciada por la OEA y el programa Cyted. Para ese entonces, los países de la región manifestaban la necesidad de contar con indicadores de CyT confiables, entre ellos Colombia. En el año de 1995 se realiza en el país una reunión preparatoria para el Taller Iberoamericano llamada “Hacia la Construcción de un Observatorio de CyT” promovida por Colciencias y el DNP. El Segundo Taller Iberoamericano sobre Indicadores de CyT (Colciencias, RICyT, OEA), como tal, se lleva a cabo en Cartagena en 1996.

Dándole continuidad a estos talleres, Colciencias contrata un estudio de factibilidad técnica y financiera y hace un pre-lanzamiento del observatorio nacional en Medellín en julio 1998. Sin embargo su formalización (suscripción del acta de constitución) se da en agosto de 1999). Claramente la creación del OCyT fue promovida por Colciencias, con el propósito de apoyar los procesos de toma de decisiones en materia de investigación e innovación a través del diseño, producción y difusión de indicadores de ciencia, tecnología e innovación (CTI). La misión del OCyT en el marco del Sistema Nacional de Ciencia, Tecnología e Innovación (SNCTI) es fundamental para el seguimiento de su dinámica, pues consiste en brindar información y conocimiento que permita a los principales actores del Sistema, la toma de decisiones y la definición de políticas y estrategias. Colciencias delegó la producción de indicadores en el Observatorio, otorgándole autonomía e independencia.

El Observatorio Colombiano de Ciencia y Tecnología (OCyT) se constituyó en una asociación civil de participación mixta y de carácter privado, sin ánimo de lucro, con patrimonio propio. Tiene cuatro tipos de miembros: promotores, adjuntos, vinculados, y aliados estratégicos que se diferencian por los montos del aporte al patrimonio que hicieron al momento de su vinculación al OCyT, donde los aliados estratégicos no aportaron. Las entidades que se asociaron con Colciencias fueron 30, en el momento de la fundación, entre las cuales se encuentran el Departamento Nacional de Planeación, 17 universidades públicas y privadas, centros de investigación, gremios, entidades promotoras del desarrollo científico-tecnológico, y una gobernación departamental.

El modelo de producción de indicadores escogido por el país fue el que se conoce como externalización vía la creación de una organización, en este caso el Observatorio. De los diferentes modelos identificados por RémiBarré, cada uno tiene fortalezas y debilidades. En el caso del modelo observatorio hay ventajas en

el análisis de las necesidades de los usuarios de la información, la concepción de los indicadores, el desarrollo de metodologías y técnicas, y la producción de indicadores e interpretación de los mismos. La recolección de datos no se considera una de las fortalezas de este modelo, y es lo que ha guiado la acción del OCyT a lo largo de su existencia: hacer uso de las bases de datos y sistemas de información existentes, y solo recolectar información cuando no exista otra fuente.

A continuación los nombres y períodos de administración de los Directores Ejecutivos del Observatorio:

- Gonzalo Ordóñez Matamoros, agosto 1999 - febrero 2003
- José Luis Villaveces Cardoso, marzo 2003 - octubre 2005
- Jorge Lucio Álvarez (encargado), noviembre - diciembre 2005
- Rafael Hurtado Heredia, enero 2006 - junio 2007
- Comité Administrativo, julio 2007 - abril 2008
- Mónica Salazar Acosta, mayo 2008 - febrero 2015

6 DISEÑO METODOLÓGICO

El proyecto se desarrollará en cuatro fases descritas de manera general en la imagen 15 por medio de una Estructura de División de Trabajo (EDT) o también conocida en sus siglas en ingles WBS, en la imagen se aprecia el primer nivel de la EDT, que a su vez muestra las fases del proyecto.

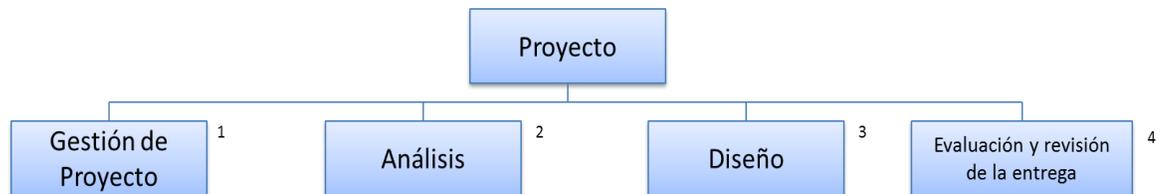


Imagen 23 Estructura de División de Trabajo – EDT

6.1.1 Fase 1: Gestión del proyecto

Esta fase se refiere a todo el proceso de planeación y gestión del proyecto en general. Su entregable son los documentos del proyecto, donde se encuentran detalles sobre líneas bases y el respectivo control del proyecto, además de información general sobre el proyecto.

6.1.2 Fase 2: Análisis

Contiene todas las actividades correspondientes al análisis a realizar según la metodología abordada de diseño de bodega de datos, expuesta en el marco referencial. Su entregable son los requerimientos y escenarios de negocio con un respectivo planteamiento de dimensiones, métricas y otras consideraciones que serán la entrada para el proceso de diseño.

6.1.3 Fase 3: Diseño

Las actividades de esta fase conforman el proceso de diseño de la bodega de datos, donde sus entregables son los diferentes diagramas de diseño conceptual, además de diseños y consideraciones de ETL y el diseño de staging área y de ser el caso, algunos diseños de la aplicación analítica.

6.1.4 Fase 4: Evaluación y revisión de la entrega

Esta fase contempla las actividades para la entrega y aprobación de finalización exitosa del proyecto, además de la documentación respectiva del proyecto. El entregable de esta fase es la aprobación de finalización con la firma del acta de entrega.

6.1.5 Detalle de la Estructura de División de Trabajo - EDT

Las imágenes 16 a 19 detallan cada una de las fases del proyecto según la EDT planteada.



Imagen 24 Fase 1 Gestión de Proyecto



Imagen 25 Fase 2 Análisis

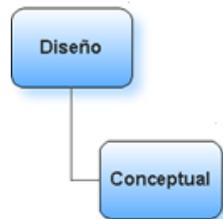


Imagen 26 Fase 3 Diseño



Imagen 27 Fase 4 Evaluación y Revisión de la Entrega

6.1.6 Presupuesto

La solución de BI para producción e inversión en Ciencia y Tecnología con una duración de seis (6) meses se estima con los siguientes costos:

Tabla 2 Costos Recurso Humano

<i>Descripción</i>	<i>Tiempo de Dedicación %</i>	<i>Valor mensual*</i>	<i>Valor Proyecto *</i>
Coordinador de Gestión de la Información.	30	2.100.000	12.600.000
Experto o Líder de área(s)	30	2.100.000	12.600.000
Asistente de Investigación	50	750.000	4.500.000
Especialista Bases de datos	100	4.500.000	27.000.000
Analista de Bases de Datos	100	3.000.000	18.000.000
Analista de Sistemas Junior	40	800.000	4.800.000
Técnico de Sistemas	40	720.000	4.320.000
Costo Total		13.970.000	83.820.000

* valores expresados en pesos colombianos

Tabla 3 Costos Recurso Tecnológico

<i>Tipo de Recurso</i>	<i>Etapas</i>	<i>Descripción</i>	<i>Costo mensual*</i>	<i>Valor Proyecto*</i>
Software	Capa de Back – End	PostgreSQL versión 9.3	0****	0
		PentahoDataIntegration	0****	0
	Capa de Bodega de Datos	PostgreSQL versión 9.3	0****	0
		PentahoDataIntegration	0****	0
	Capa OLAP	Mondrian – Pentaho	0****	0
	Capa Front – End	Quilkview	0**	0
Business Analytics& BI		0***	0	
Hardware	Capa de Back – End	Server Bases de Datos – Mantenimiento	166.667	1.000.000
	Capa de Bodega de Datos	Server Aplicaciones - Mantenimiento	166.667	1.000.000
		Server Aplicaciones Públicas – Mantenimiento		
	Capa OLAP			
	Capa Front – End			
Capacitación	Capacitación estimada en tiempo		1.166.667	7.000.000
Total Costo Recurso Tecnológico			1.500.000	9.000.000

* Valor expresado en pesos colombianos

** En versión de escritorio para versión Enterprise tiene costo adicional que debería asumirse por todas las áreas del OCyT.
*** En versión de código abierto, para versión Enterprise tiene costo adicional que debería asumirse por todas las áreas del OCyT.
**** software de código abierto

7 MODELO CONCEPTUAL DE DATAMART PARA INVERSIÓN Y PRODUCCIÓN DE CIENCIA TECNOLOGÍA E INNOVACIÓN

Al realizar el modelo conceptual para el datamart de producción e inversión de ciencia tecnología e innovación, se tuvo en cuenta la información contenida en el informe anual de indicadores de ciencia y tecnología²⁵ producido por el Observatorio Colombiano de Ciencia y Tecnología – OCyT, puesto que es el instrumento más completo para establecer fuentes de información del SNCTI.

7.1 FUENTES DE INFORMACIÓN

- ScienTI es la plataforma de información administrada por Colciencias para recoger información del SNCTI, cuenta con servicios y aplicaciones para registrar información sobre investigadores, grupos de investigación e instituciones entre otras, y tiene por objeto contribuir a la gestión de la actividad científica, tecnológica y de innovación del país.
 - CvLAC es la plataforma de información adscrita a Colciencias y que se encarga de recopilar información sobre hoja de vida del personal que realiza actividades de CTI. Desde 2002 registra permanentemente datos de la hoja de vida de los individuos que hacen parte del SNCTI.
 - GrupLAC Es un aplicativo que recopila datos y estadísticas sobre la trayectoria de los grupos de investigación en Colombia. En esta fuente de información se integran datos sobre la actuación de los integrantes, la producción científica y tecnológica, proyectos de investigación en el marco del grupo de investigación.

²⁵Versión en línea en <http://ocyt.org.co/es-es/InformeAnualIndicadores>

- **SNIES:** El Sistema Nacional de Información de la Educación Superior (SNIES), es la fuente oficial de información de la educación superior que consolida y suministra datos, estadísticas e indicadores relevantes del sector educativo en Colombia. El objetivo de este sistema es mantener y divulgar información confiable, oportuna y relevante de las instituciones y de los programas de educación superior aprobados por el Ministerio de Educación Nacional Colombiano.
- **OLE:** Observatorio Laboral para la Educación es un sistema de información especializado perteneciente al Ministerio de Educación Nacional en él se encuentra información acerca de oferta de graduados, perfil de los graduados, ingreso mensual promedio y vinculación al sector formal de la economía, tiene como objetivo hacer seguimiento a los graduados del país y su empleabilidad en el mercado laboral colombiano.
- **SCOPUS:** Es una base de datos bibliográfica de resúmenes y citas de artículos de revistas científicas. Cubre documentos de más de 5.000 editores internacionales, incluyendo la cobertura de 16.500 revistas revisadas por pares de las diferentes áreas del conocimiento.
- **WEB OF SCIENCE:** es un servicio en línea de información científica, suministrado por Thomson Reuters. Proporciona información de bases de datos de citas de artículos de revistas científicas, libros y otros tipos de material impreso que abarcan todas las áreas del conocimiento académico.
- **SCIELO:** (Scientific Electronic Library Online o Biblioteca Científica Electrónica en Línea) es un proyecto de biblioteca electrónica, iniciativa de la Fundación para el Apoyo a la Investigación del Estado de São Paulo, Brasil (Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP) y del Centro Latinoamericano y del Caribe de Información en Ciencias de la Salud (BIREME), que permite la publicación electrónica de ediciones completas de las revistas científicas.
- **PUBLINDEX:** Es el sistema nacional de indexación y homologación de revistas especializadas en CT+I, es un sistema colombiano para la clasificación, actualización y certificación de publicaciones científicas y tecnológicas.

- REDALYC: Es la red de revistas científicas de América Latina y el Caribe, España y Portugal
- EDIT: Es la Encuesta de Desarrollo e Innovación Tecnológica realizada por el DANE bianual tiene como objetivo caracterizar la dinámica de innovación realizada por las empresas del país, así como analizar la utilización de los instrumentos públicos de apoyo
- Información de Inversión: Los valores de inversión deben ser los correspondientes a la ejecución realizada por la(s) institución(es), es decir, el valor financiado y el total del gasto.

Para calcular la inversión en Ciencia y Tecnología, se deben tener en cuenta el gasto y la financiación de las actividades como Investigación y Desarrollo (I+D), enseñanza y formación científica y técnica, servicios científicos y tecnológicos, administración y actividades de apoyo. Así como las actividades estrechamente relacionadas con producción, aplicación y difusión de conocimientos científicos y tecnológicos.

7.2 ESCENARIOS DE ANALISIS

El repositorio de información y los informes construidos deben permitir observar la información referente a la inversión, producción, formación académica, y la vinculación a grupos de investigación de las personas ocupadas en actividades de investigación, en esta medida el repositorio de información de inversión y producción deben dar respuesta a los siguientes interrogantes:

1. ¿Cuánto dinero invierte Colombia en CTI?
2. ¿Cuánto dinero invierte una institución en investigación?
3. ¿Con el dinero que ha invertido la institución cuántos y qué productos de investigación ha realizado la institución?
4. ¿Cuál es la distribución de los productos de investigación realizados en Colombia? (áreas de conocimiento, departamento, instituciones, etc.)
5. ¿Cuál es la relación entre investigadores, inversión y producción bibliográfica en Colombia?
6. ¿Quiénes realizan actividades de Ciencia Tecnología e Innovación - CTI?
7. ¿Qué recursos financieros necesita el desarrollo de actividades en CTI?

8. ¿Qué producción científica tiene el país?
9. ¿Quiénes hacen producción científica en el país?
10. ¿Cuánto se invierte en CTI en Colombia?
11. ¿Qué formación tienen las personas que desarrollan actividades de CTI?
12. ¿Con qué capacidades de investigación cuenta Colombia?
13. ¿Cuánto se invierte anualmente en CTI en Colombia?

7.3 DIMENSIONES

Tabla 4 Descripción de Dimensiones

<i>Dimensión</i>	<i>Descripción</i>	<i>Tipo</i>
Objetivo Socio Económico	Esta dimensión representa el objetivo socioeconómico de las empresas en donde se invierte.	Conformed
Gasto	Esta dimensión representa las variables de gasto de las instituciones en CTI	Junk
Financiamiento	Esta dimensión representa las variables del financiamiento de las instituciones en CTI	Junk
Persona	Esta dimensión representa la información de las personas perteneciente a la base de datos de investigadores	Conformed
Programa	Esta dimensión representa la información de programas académicos utilizados para identificar las profesiones y nivel de estudio de los investigadores	Conformed
Área Conocimiento	Esta dimensión representa las áreas del conocimiento en las que se ubica un investigador por su estudio o por la aplicación de su producción bibliográfica	Conformed - Role-Playing
Institución	Esta dimensión representa las instituciones que pertenecen al sistema nacional de ciencia y tecnología – SNCTI, el tipo de institución.	Conformed

Dimensión	Descripción	Tipo
Ubicación Geográfica – Entidad Territorial	Esta dimensión representa información de ubicación de las instituciones del SNCTI	Conformed - Role-Playing
Tiempo	Esta dimensión representa la Información sobre periodos de tiempo.	Date and Time – Conformed - Role-Playing
Autor	Esta dimensión representa la información de los autores de la producción bibliográfica ²⁶	Conformed
Producto	Esta dimensión representa la producción bibliográfica del país	Conformed
Revista	Esta dimensión representa la información de las revistas en donde está contenida la información de artículos y documentos de la dimensión producto	Conformed

7.4 Tablas de Hechos

7.4.1 Tabla de Hechos de Inversión

7.4.1.1 Indicadores

Tabla 5 Gasto

Nombre	Gasto en CTI
Variables	1. Valor Total del Gasto interno o intramuros <ul style="list-style-type: none"> a. Gastos Corrientes <ul style="list-style-type: none"> • Total anual de salarios del personal de CTI • Total anual de remuneraciones del personal de CTI • Total anual de primas, vacaciones, contribuciones a fondos de pensiones, pagos de seguridad social, impuestos salariales, etc. • Total anual de compra de materiales, suministros y equipos de apoyo a las actividades ACTI que no hacen parte de los gastos de capital

²⁶ Para este diseño no se almacena la información del autor dentro de la dimensión persona debido a que esto representa un proceso de depuración y verificación que no hace parte del alcance de este proyecto

	b. Gastos de Capital <ul style="list-style-type: none"> Total del capital fijo utilizado en programas de ACTI 2. Valor Total Gasto Extramuros 3. Tipo de Recursos 4. Tipo de Institución 5. Tipo de Actividad 6. Objetivo Socioeconómico	
Fuente de Información	Información de Gasto de la(s) Institución(es)	
Cálculo	Suma de total de gasto en Intramuros y Extramuros Por año	
Frecuencia	Anual	
Desagregaciones	Temática	<ul style="list-style-type: none"> Tipo de Recursos Tipo de Institución Tipo de Actividad Objetivo Socioeconómico
	Geográfica	<ul style="list-style-type: none"> Departamental
	Específica	<ul style="list-style-type: none"> Institucional
Tratamiento de las variables	Se debe calcular en pesos constantes	
Limitaciones del indicador	Se debe contar con la información real de la(s) institución(es) la información no es publica	

Tabla 6 Financiamiento

Nombre	Financiamiento en CTI	
Variables	<ol style="list-style-type: none"> 1. Total de recursos propios 2. Total de recursos públicos <ul style="list-style-type: none"> • Total de recursos de la nación • Total de recursos territoriales • Total de recursos por regalías 3. Recursos Privados <ul style="list-style-type: none"> • Total de recursos por asociaciones • Total de recursos empresariales • Total de recursos internacionales • Total de recursos de otras entidades del conocimiento 4. Tipo de Recursos 5. Tipo de Institución 6. Tipo de Actividad 7. Objetivo Socioeconómico 	
Fuente de Información	Información de financiación de la(s) Institución(es)	
Cálculo	Suma de total de la financiación por año	
Frecuencia	Anual	
Desagregaciones	Temática	<ul style="list-style-type: none"> • Tipo de Recursos • Tipo de Institución • Tipo de Actividad • Objetivo Socioeconómico
	Geográfica	<ul style="list-style-type: none"> • Departamental
	Específica	<ul style="list-style-type: none"> • Institucional
Tratamiento de las variables	Se debe calcular en pesos constantes	
Limitaciones del indicador	Se debe contar con la información real de la(s) institución(es) la información no es publica	

7.4.2 Tabla de Hechos de Producción

7.4.2.1 Indicadores

Tabla 7 Producción

Nombre	Financiamiento en CTI	
Variables	<ol style="list-style-type: none"> 1. Total de producción bibliográfica de autores vinculados a instituciones colombianas Web of Science 2. Total de producción bibliográfica de autores vinculados a instituciones colombianas Scopus 3. Total de producción bibliográfica de autores vinculados a instituciones colombianas SciELO 4. Tipo de Documento 5. Índice Documental 6. Área de Conocimiento OCDE 7. Disciplina 8. Coautoría 	
Fuente de Información	Scopus, Web of Science, SciELO	
Cálculo	Conteo de productos por año	
Frecuencia	Anual	
Desagregaciones	Temática	<ul style="list-style-type: none"> • Tipo de Documento • Índice Documental • Área de Conocimiento OCDE • Disciplina • Coautoría
	Geográfica	<ul style="list-style-type: none"> • Departamental • Países de Latinoamérica
	Específica	
Tratamiento de las variables	<ul style="list-style-type: none"> • Se debe tener en cuenta la vinculación del autor a la institución. • Se debe tener en cuenta que el documento puede estar escrito en idiomas diferentes y ser el mismo estos documentos deben contar una sola vez 	
Limitaciones del indicador		

7.4.3 Tabla de Hechos de Investigadores

7.4.3.1 Indicadores

Tabla 8 Investigadores

Nombre	Investigadores	
Variables	<ol style="list-style-type: none"> 1. Investigadores 2. Sexo 3. Fecha de Nacimiento 4. Nivel de Formación 5. Área de conocimiento OCDE 6. Institución 7. Tipo Institución 8. Entidad Territorial 	
Fuente de Información	Base de datos de investigadores de las institución(es)	
Cálculo	Conteo de investigadores por año	
Frecuencia	Anual	
Desagregaciones	Temática	<ul style="list-style-type: none"> • Área de conocimiento OCDE • Tipo de Institución • Máximo Nivel de Formación
	Geográfica	<ul style="list-style-type: none"> • Departamental
	Específica	<ul style="list-style-type: none"> • Sexo • Rango de Edad
Tratamiento de las variables	<ul style="list-style-type: none"> • Fecha de Nacimiento: Se extrae el año de nacimiento del investigador y se calcula la edad del mismo de acuerdo al año de consulta, luego se agrupan en los siguientes rangos: <ul style="list-style-type: none"> • 0 a 20 • 21 a 30 • 31 a 40 • 41 a 50 • 51 a 60 • > 60 • Máximo Nivel de Formación: Se consulta el historial de formación del investigador y se selecciona el máximo nivel de formación. • Tipo Institución: Se establece de la institución que avala al grupo de investigación al que pertenece el investigador. 	

	<ul style="list-style-type: none">• Ubicación geográfica: Se establece del departamento en donde se ubica la institución a la que pertenece el investigador.
Limitaciones del indicador	

7.5 DISEÑO CONCEPTUAL

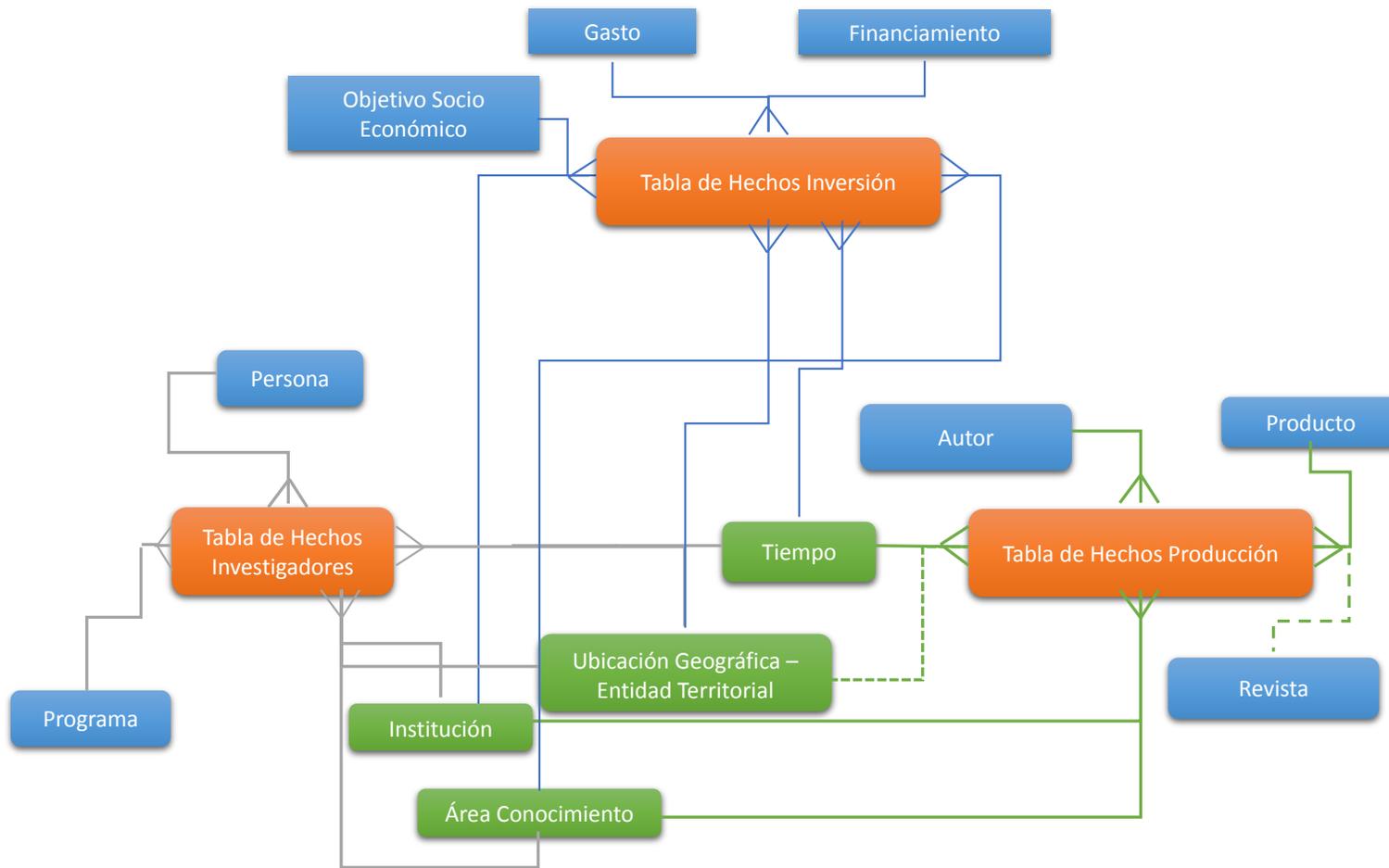


Imagen 28 Diseño Conceptual Genérico

7.6 DISEÑO DE ETL

7.6.1 Dimensión Persona

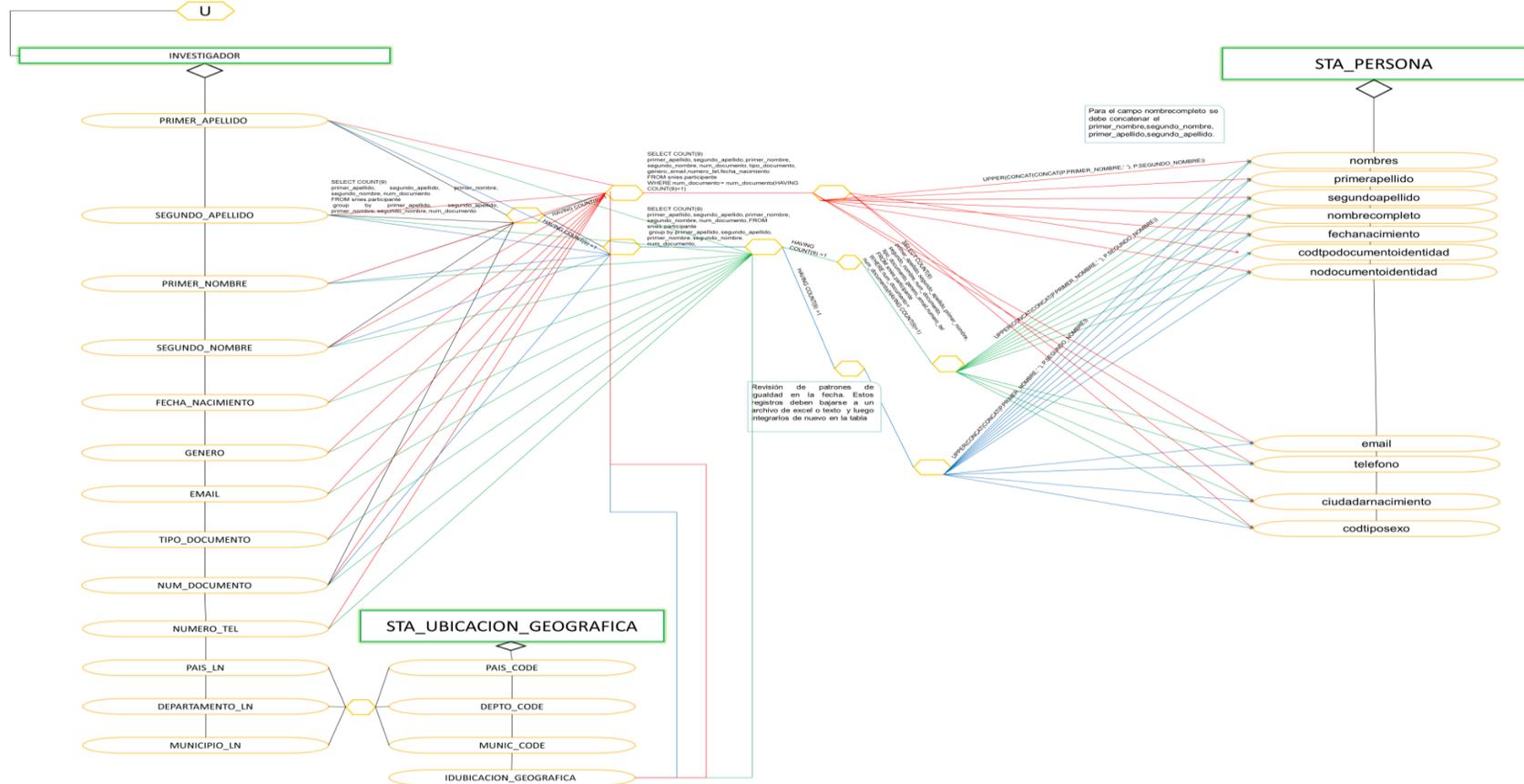


Imagen 29 Dimensión Persona

7.6.2 Dimensión Programa Académico

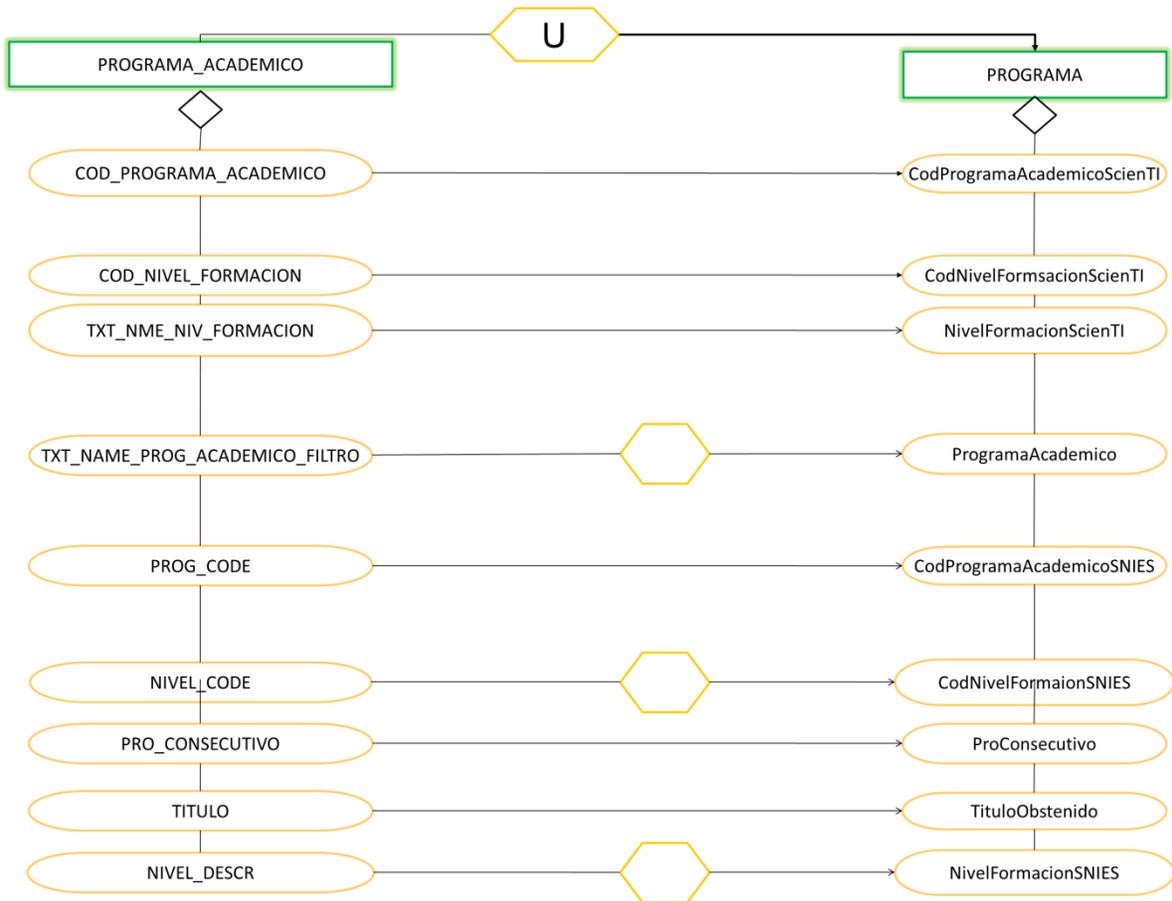


Imagen 30 Dimensión Programa Académico

7.6.3 Dimensión Institución

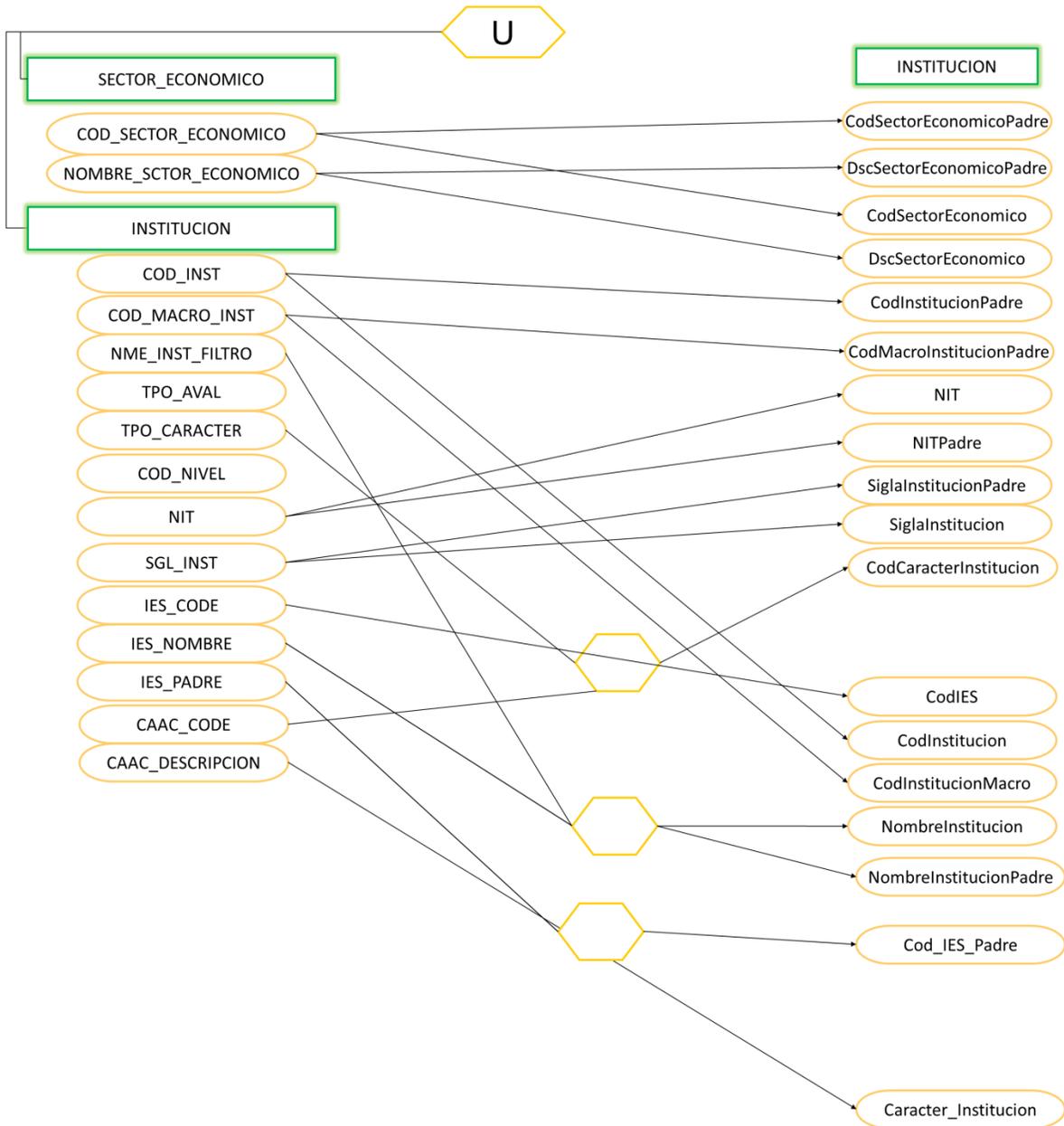


Imagen 31 Dimensión Institución

7.6.4 Dimensión Área de Conocimiento

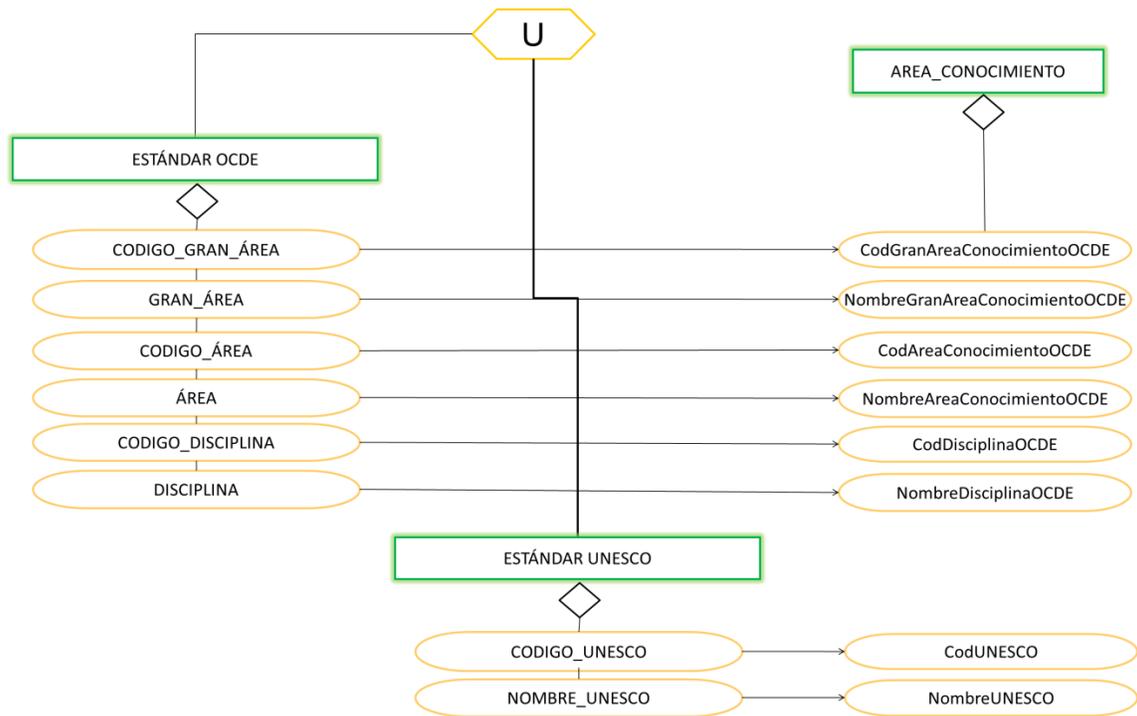


Imagen 32 Dimensión Área de Conocimiento

7.6.5 Dimensión ubicación Geográfica

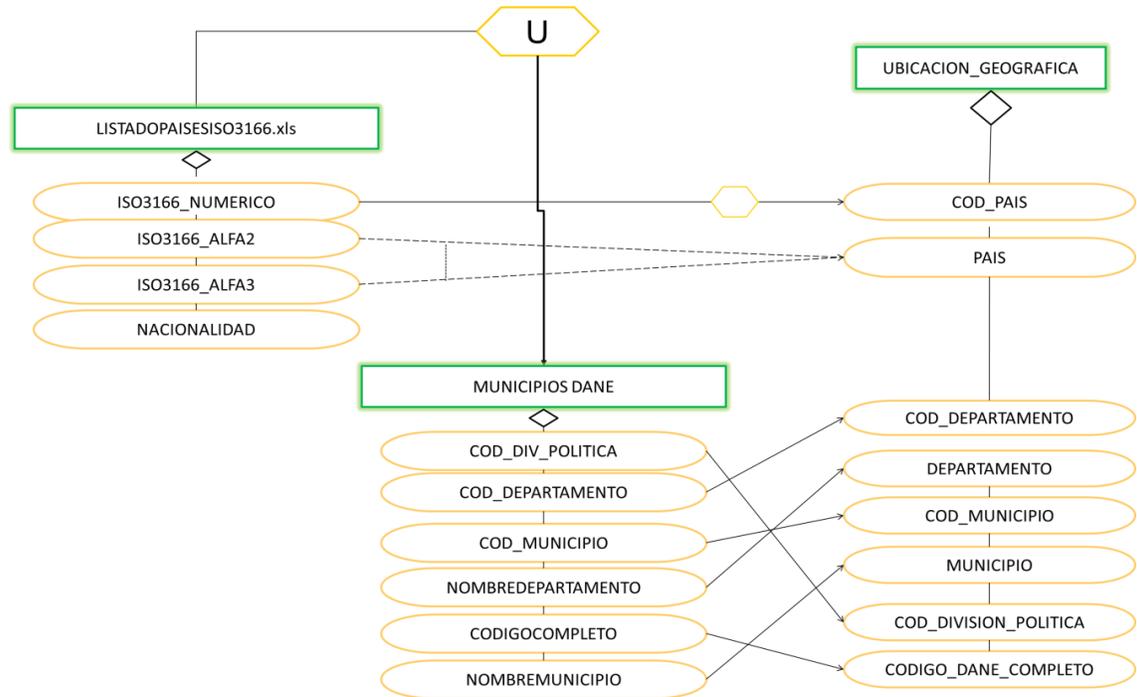


Imagen 33 Dimensión Ubicación Geográfica

7.6.6 Dimensión Tiempo

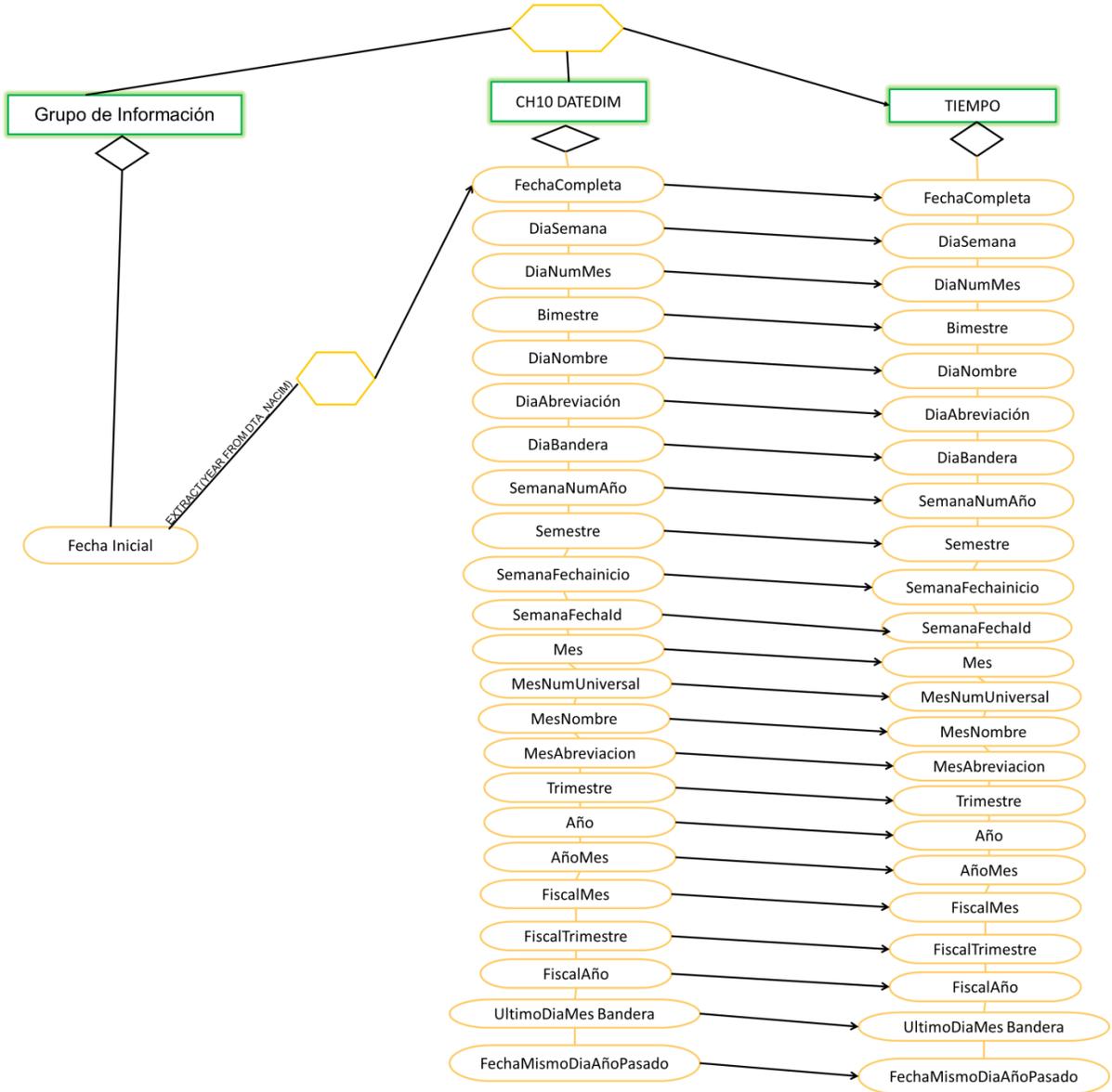


Imagen 34 Dimensión Tiempo

7.6.7 Dimensión Revista

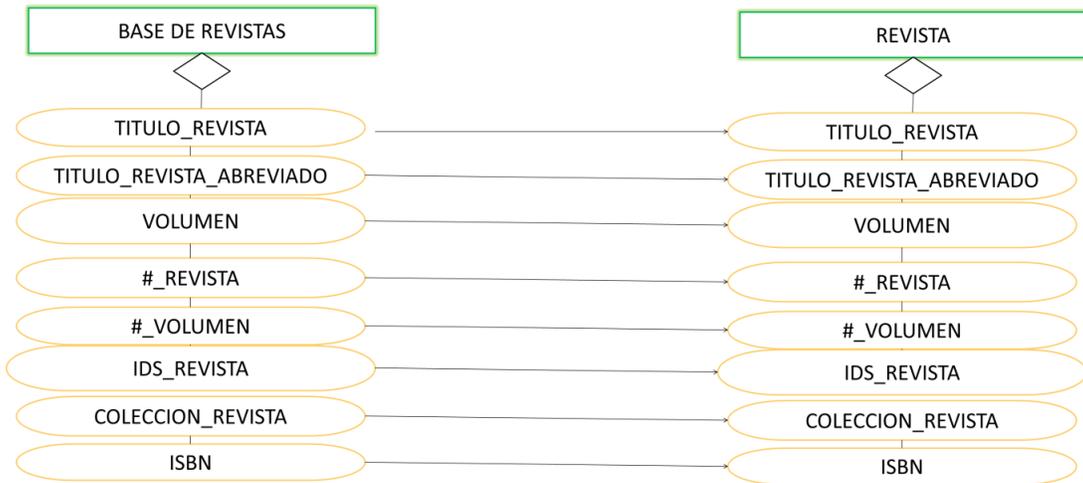


Imagen 35 Dimensión Revista

7.6.8 Dimensión Producto

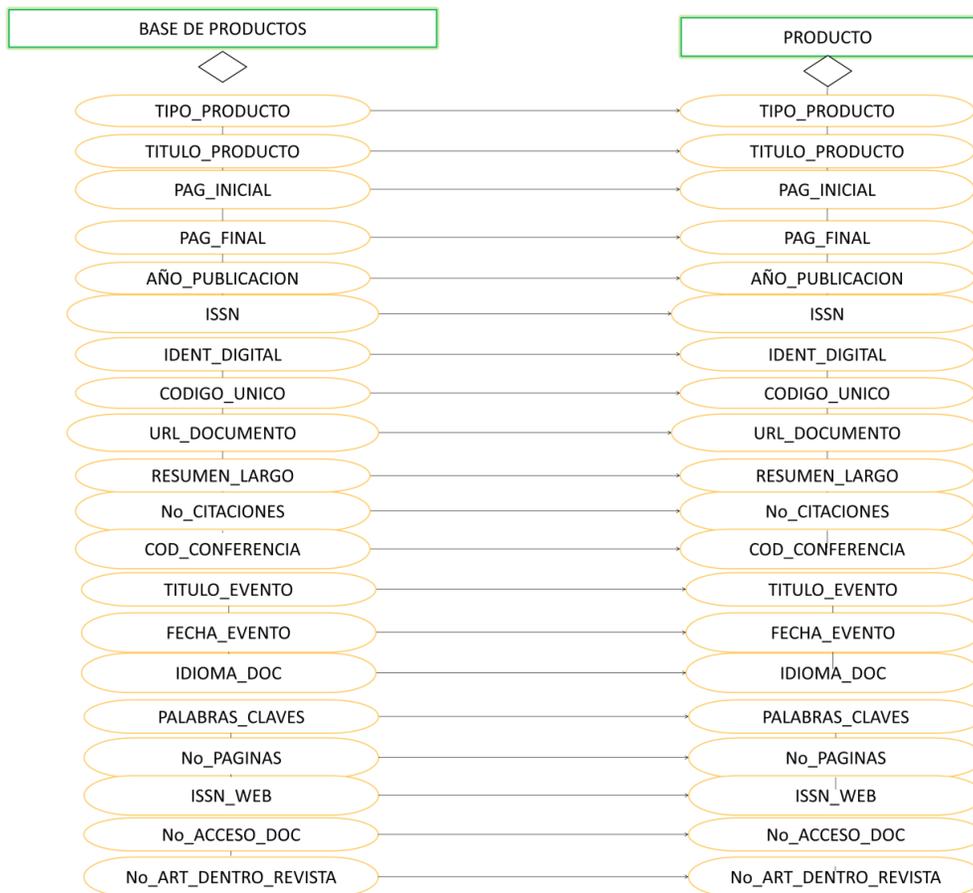


Imagen 36 Dimensión Producto

7.6.9 Dimensión Autor



Imagen 37 Dimensión Autor

7.6.10 Dimensión Objetivo Socio Económico



Imagen 38 Dimensión Objetivo Socioeconómico

8 APLICACIÓN MODELO CONCEPTUAL DE DATAMART PARA INVERSIÓN Y PRODUCCIÓN DE CIENCIA TECNOLOGÍA E INNOVACIÓN

8.1 DIMENSIONES

<i>Dimensión</i>	<i>Descripción</i>	<i>Tipo</i>
Objetivo Socio Económico	Esta dimensión representa el objetivo socioeconómico de las empresas en donde se invierte.	Conformed
Gasto	Esta dimensión representa las variables del gasto de las instituciones en CTI	Junk
Financiamiento	Esta dimensión representa las variables del financiamiento de las instituciones en CTI	Junk
Persona	Esta dimensión representa la información de las personas perteneciente a la base de datos de investigadores	Conformed
Programa	Esta dimensión representa la información de programas académicos utilizados para identificar las profesiones y nivel de estudio de los investigadores	Conformed
Área Conocimiento	Esta dimensión representa las áreas del conocimiento en las que se ubica un investigador por su estudio o por la aplicación de su producción bibliográfica	Conformed - Role-Playing
Institución	Esta dimensión representa las instituciones que pertenecen al sistema nacional de ciencia y tecnología – SNCTI, el tipo de institución.	Conformed

Dimensión	Descripción	Tipo
Ubicación Geográfica – Entidad Territorial	Esta dimensión representa información de ubicación de las instituciones del SNCTI	Conformed - Role-Playing
Tiempo	Esta dimensión representa la Información sobre periodos de tiempo.	Date and Time – Conformed - Role-Playing
Autor	Esta dimensión representa la información de los autores de la producción bibliográfica ²⁷	Conformed
Producto	Esta dimensión representa la producción bibliográfica del país	Conformed
Revista	Esta dimensión representa la información de las revistas en donde está contenida la información de artículos y documentos de la dimensión producto	Conformed
Tesis	Esta dimensión representa la realización o no de tesis por parte de la persona asociada	Junk
Actividad	Esta dimensión representa el estado del investigador, entendiéndose como una etiqueta que define si un investigador está activo o inactivo de acuerdo a una serie de reglas establecidas	Conformed

²⁷ Para este diseño no se almacena la información del autor dentro de la dimensión persona debido a que esto representa un proceso de depuración y verificación que no hace parte del alcance de este proyecto

8.2 DISEÑO CONCEPTUAL

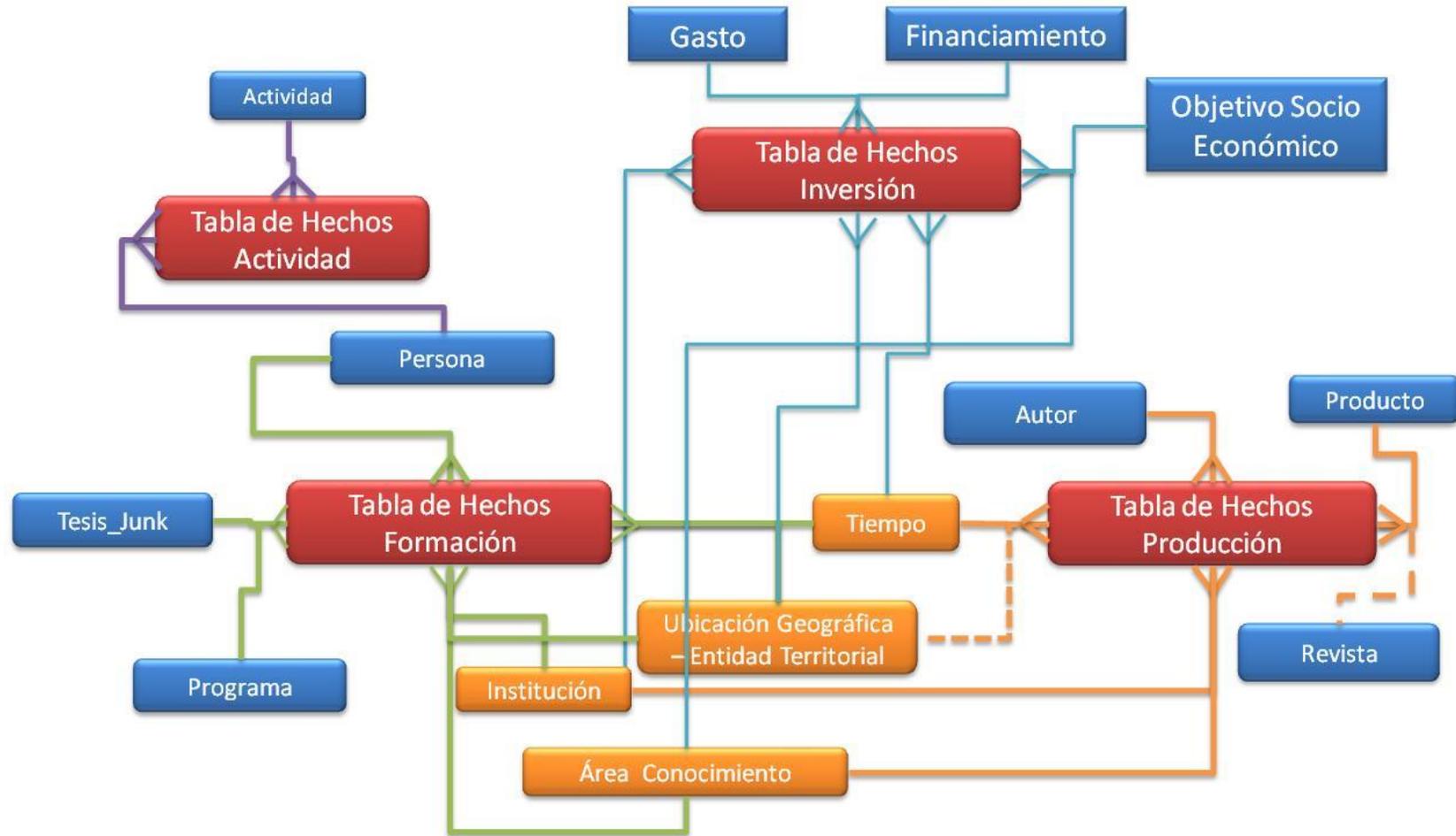


Imagen 39 Diseño Conceptual Aplicado al OCyT

8.3 DISEÑO DE ETL

8.3.1 Dimensión Persona

Parte 1: carga de información SNIES a Staging.

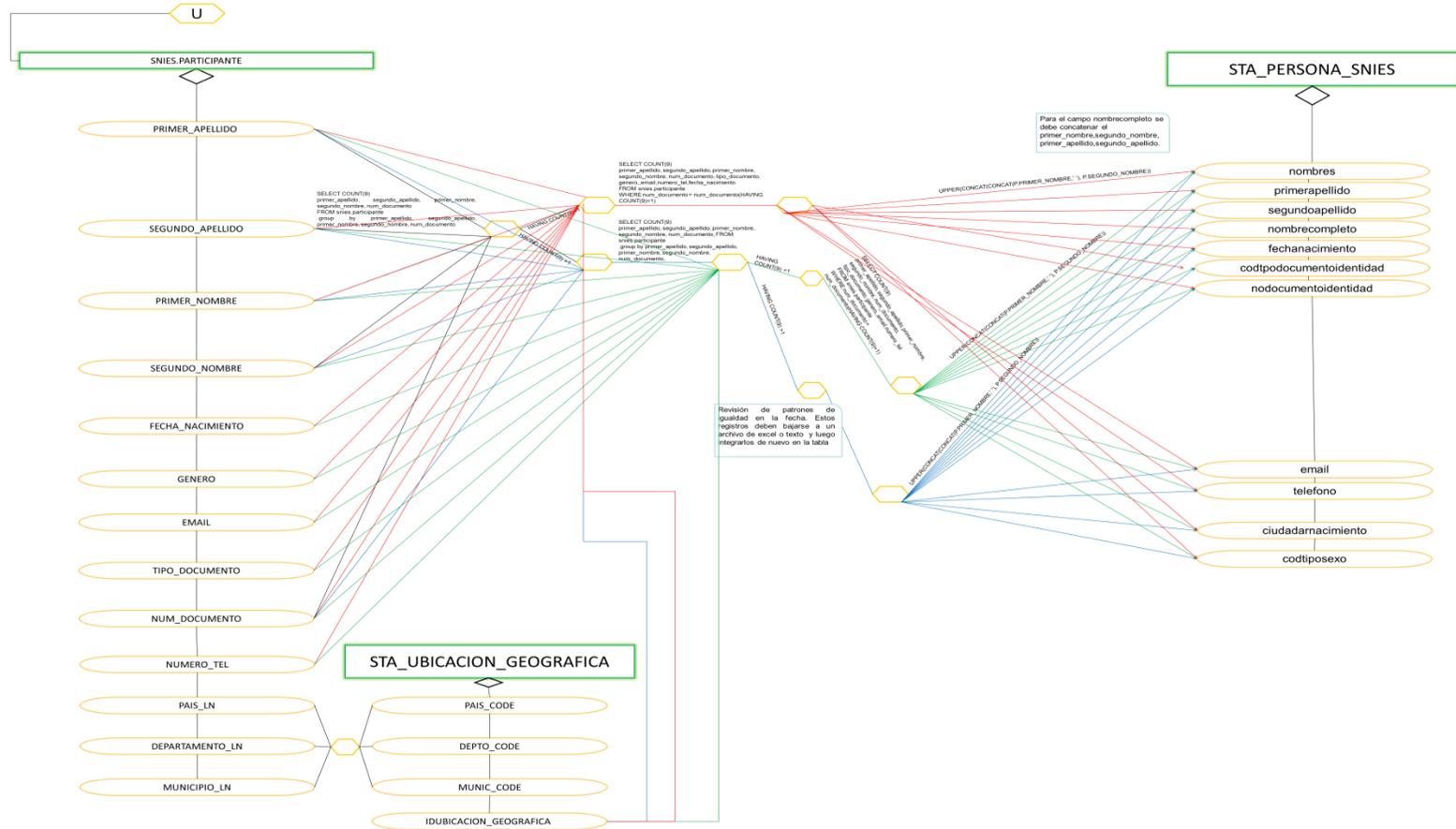


Imagen 40 Dimensión Persona I

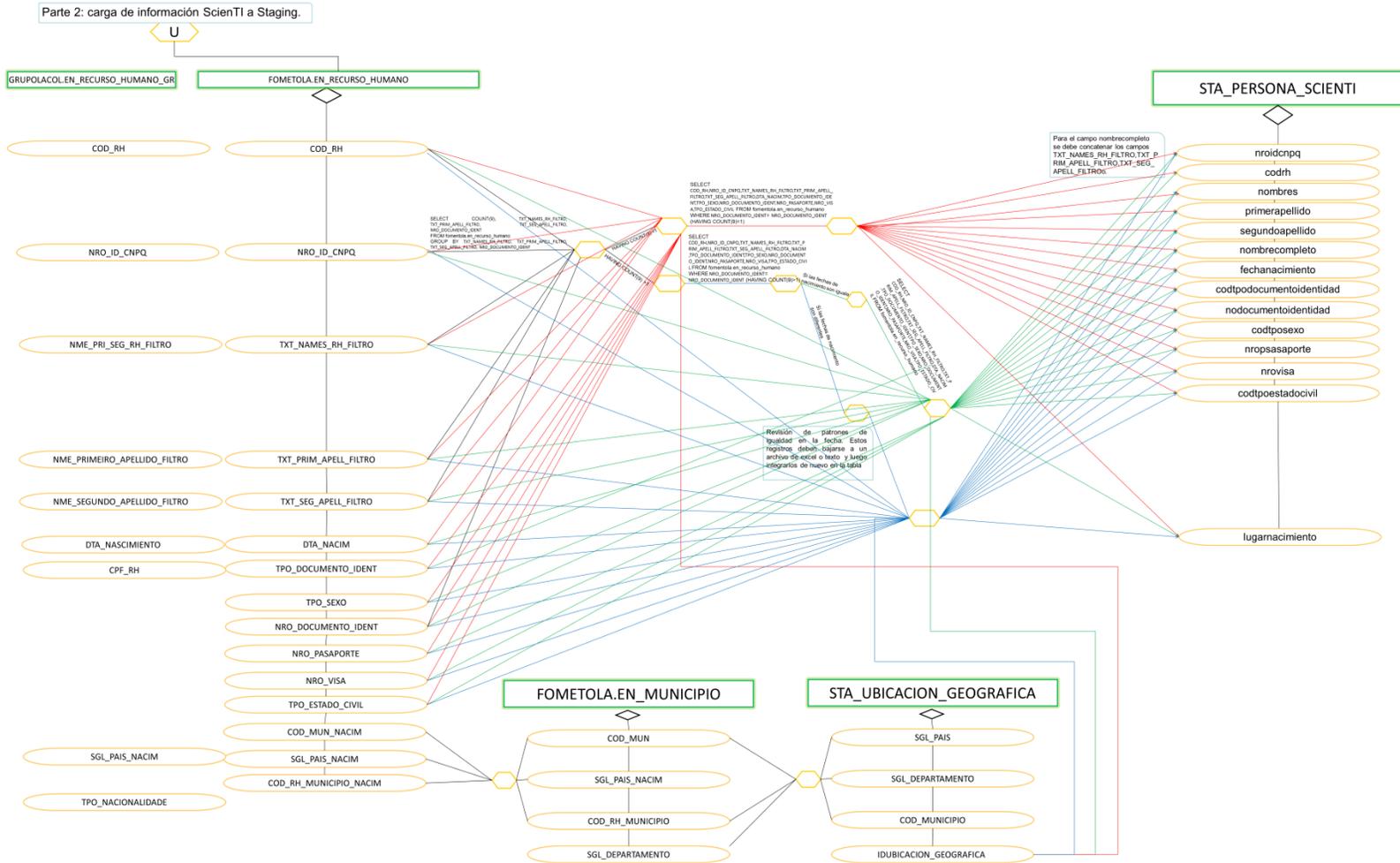
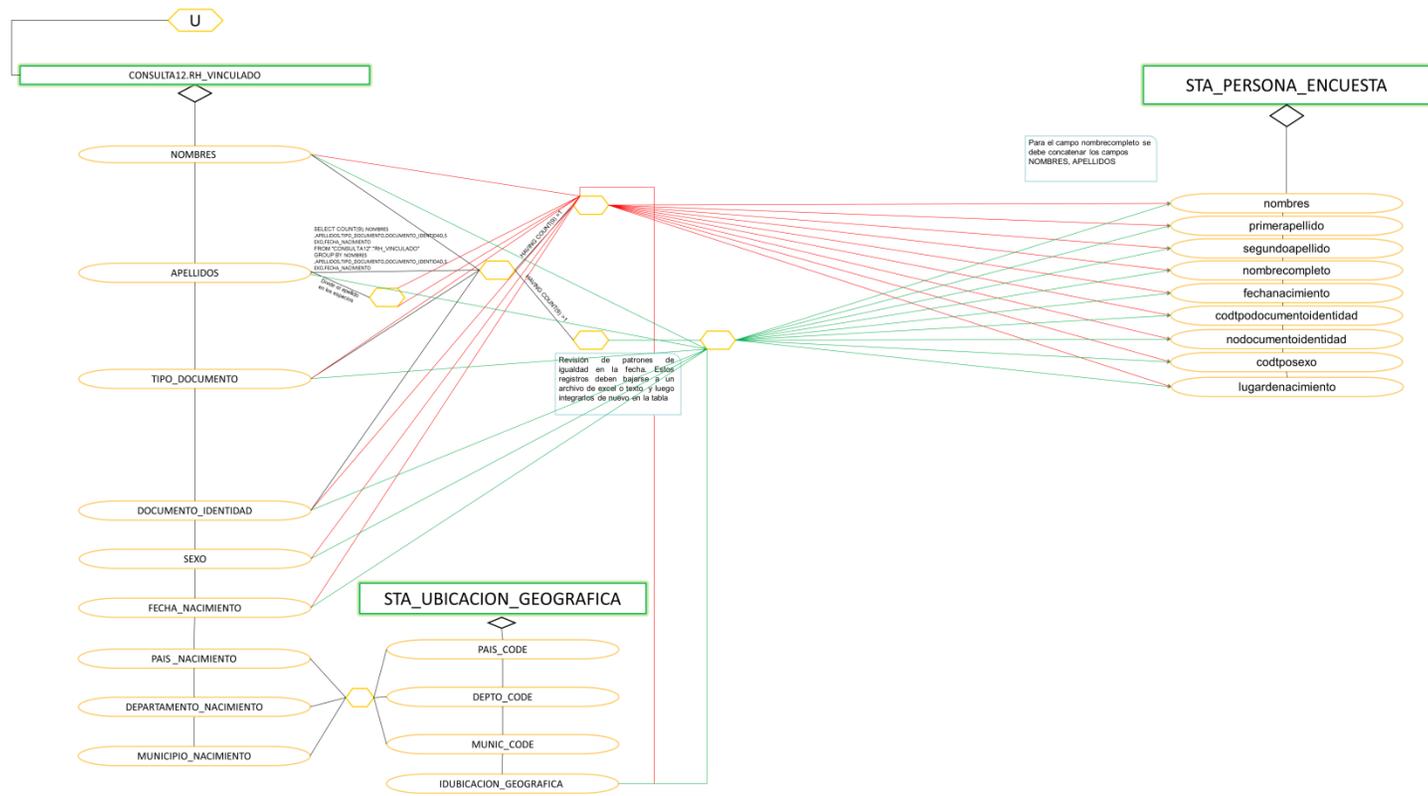


Imagen 41 Imagen 30 Dimensión Persona II



Parte 3: carga de información Encuesta a Staging.

Imagen 42 Imagen 30 Dimensión Persona III

8.3.2 Dimensión Tesis

8.3.3 Dimensión Programa Académico

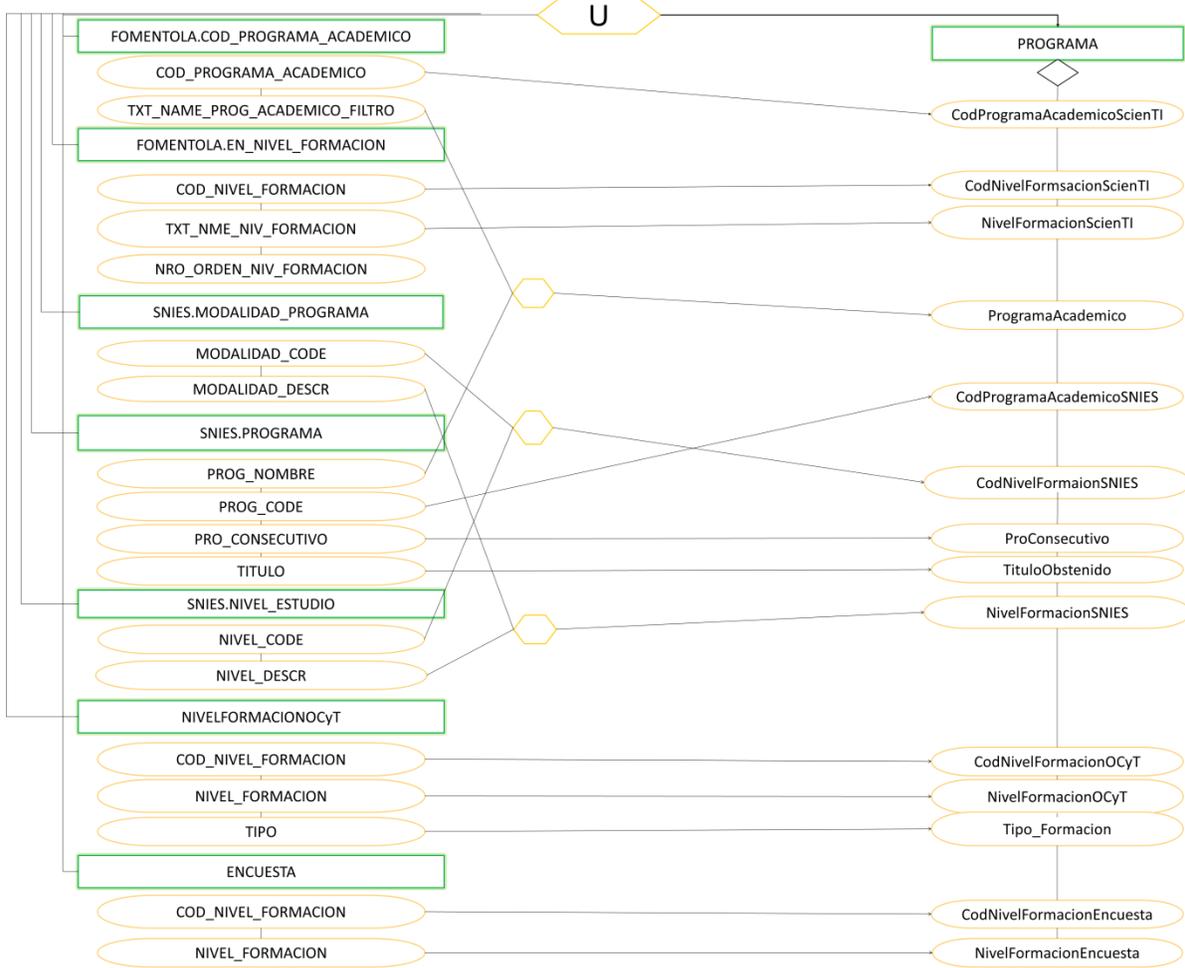


Imagen 43 Dimensión Programa Académico

8.3.4 Dimensión Institución

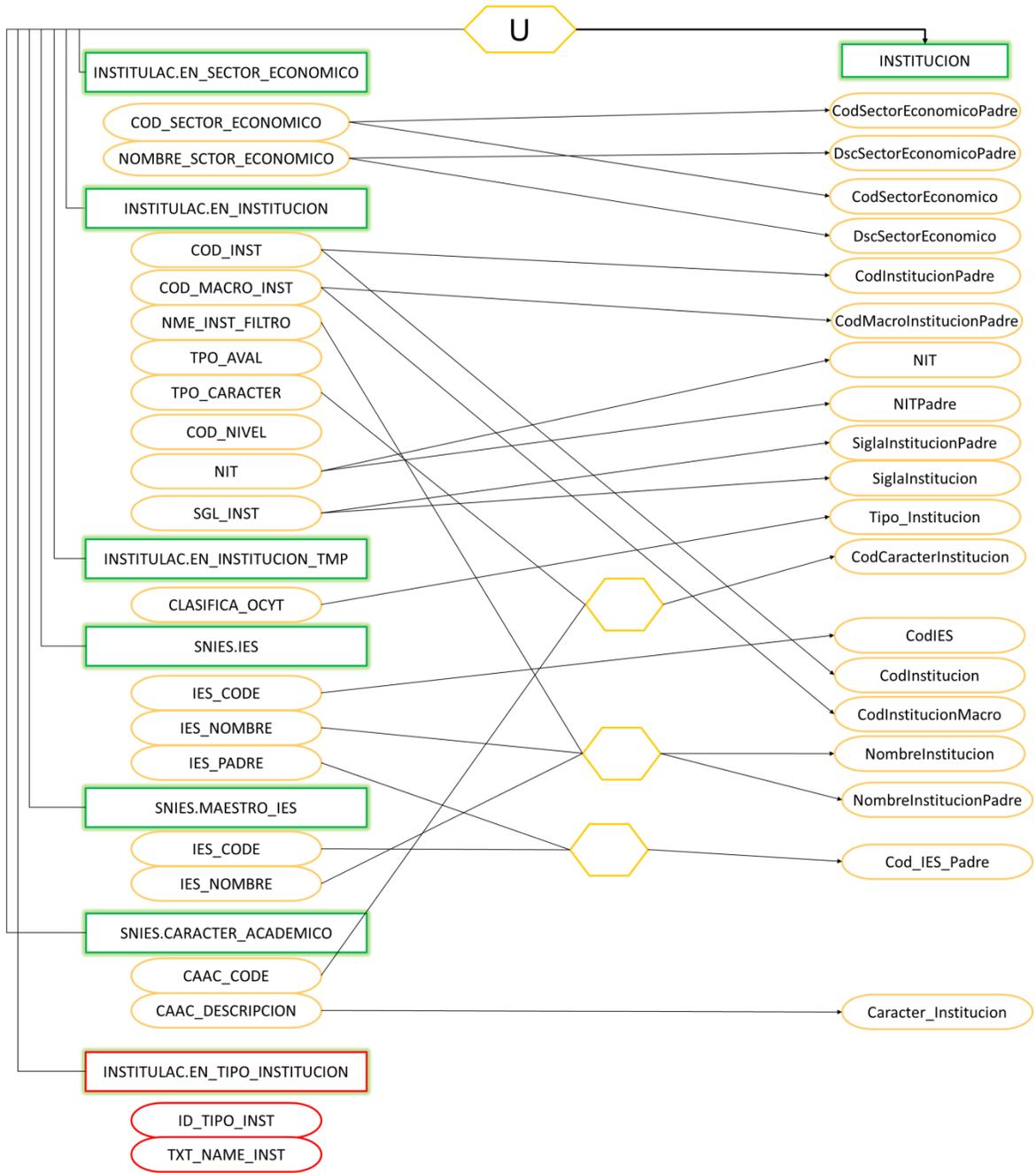


Imagen 44 Dimensión Institución

8.3.5 Dimensión Área de Conocimiento

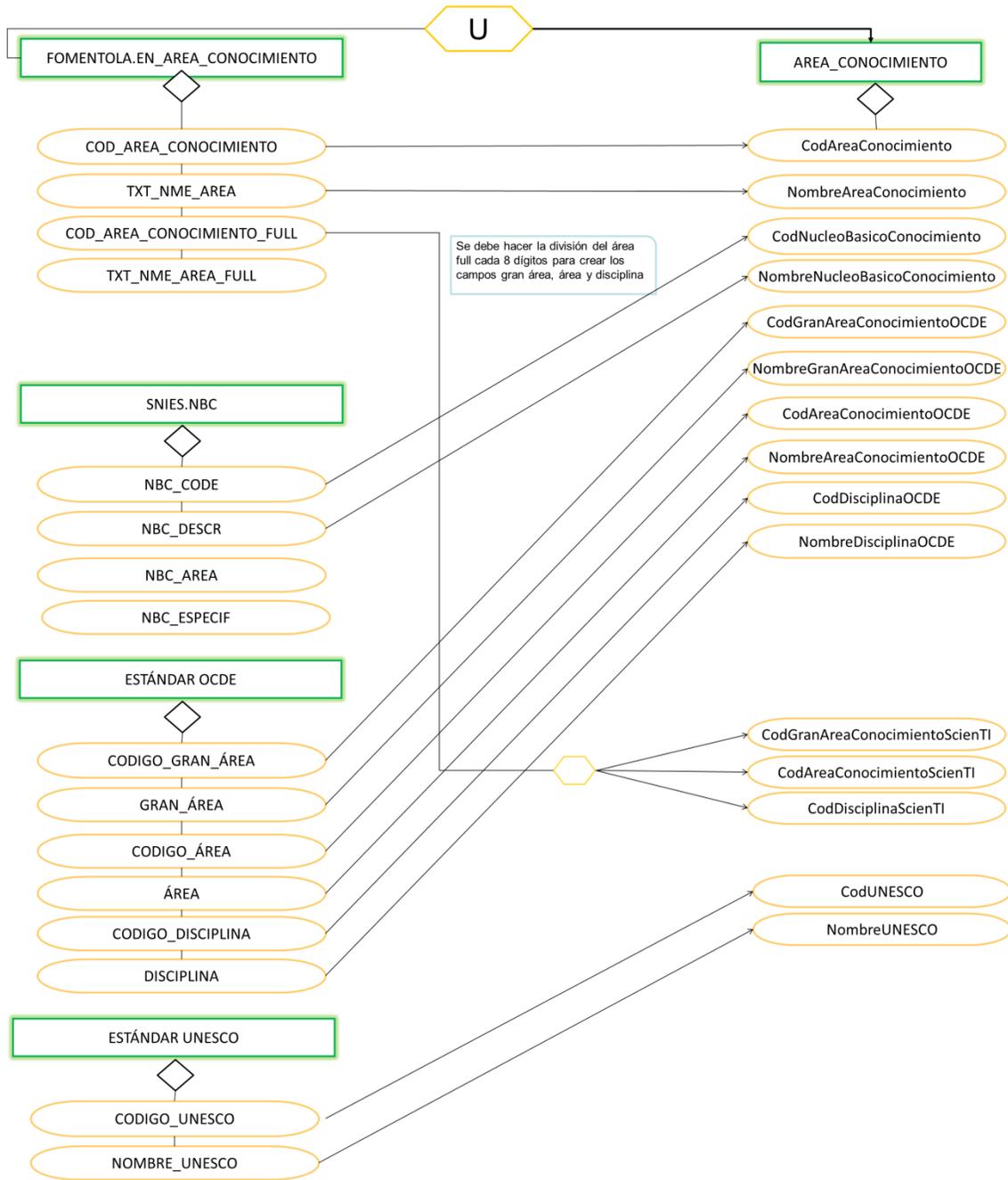


Imagen 45 Dimensión Áreas del Conocimiento

8.3.6 Dimensión ubicación Geográfica

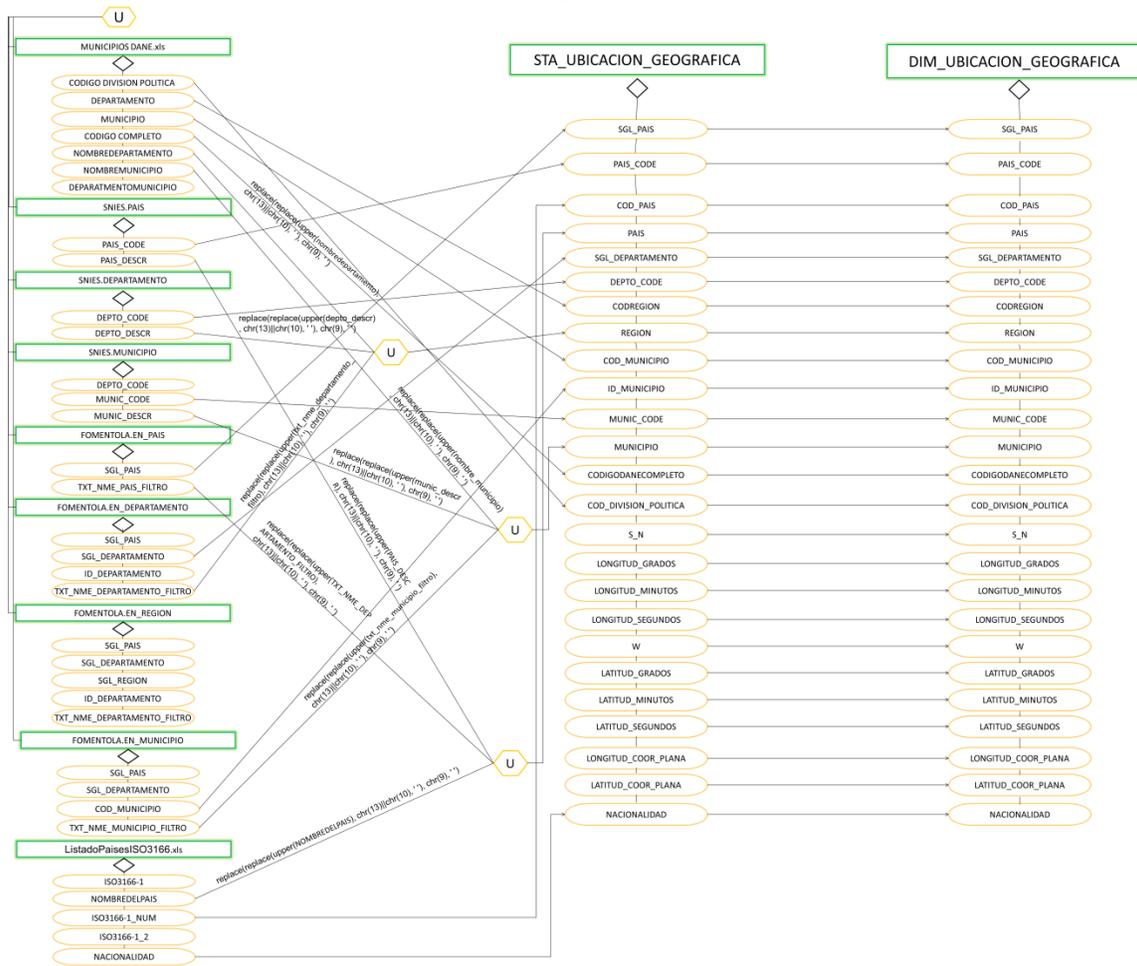


Imagen 46 Dimensión Ubicación Geográfica

8.3.7 Dimensión Tiempo

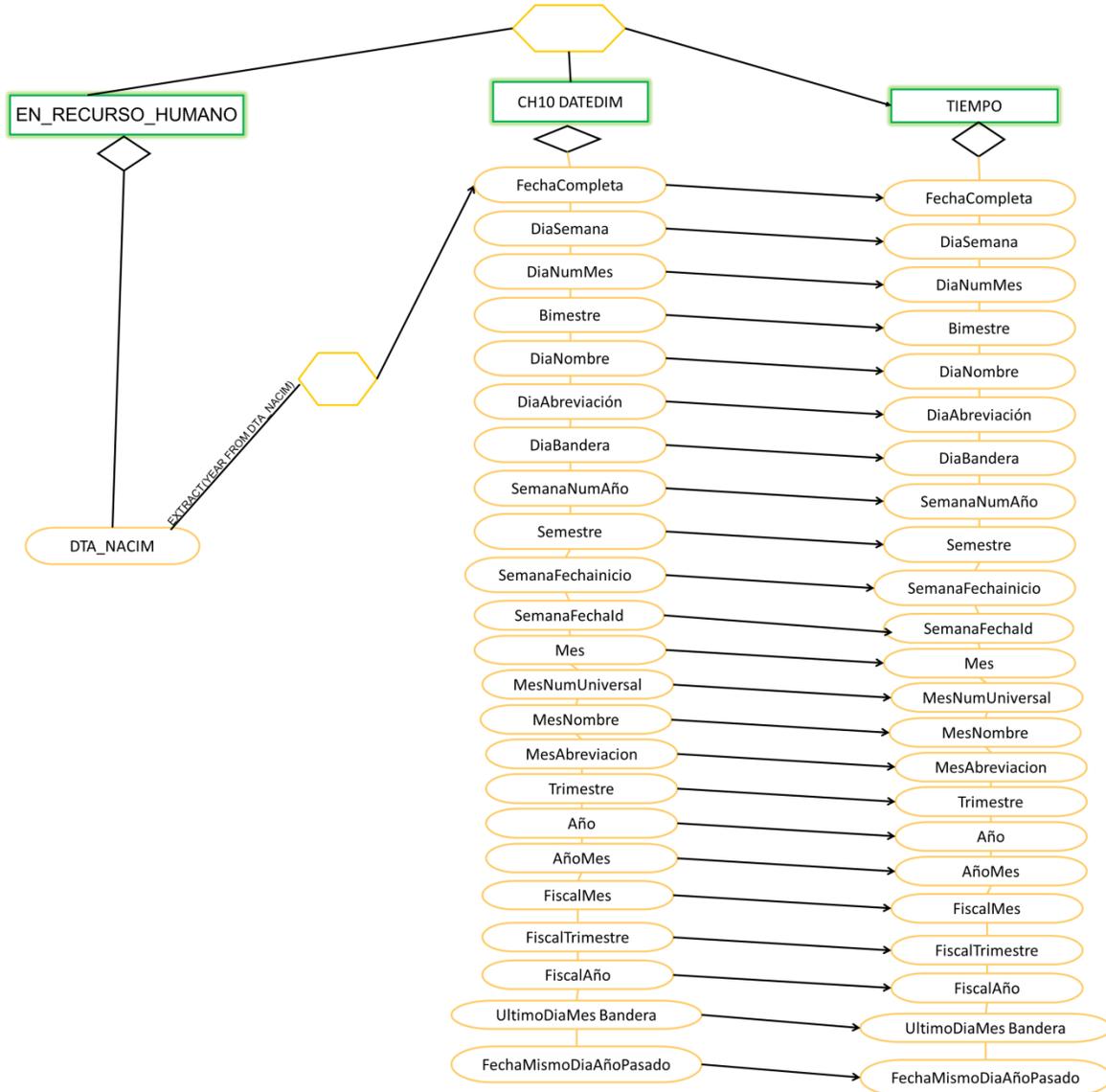


Imagen 47 Dimensión Tiempo

8.3.8 Dimensión Revista

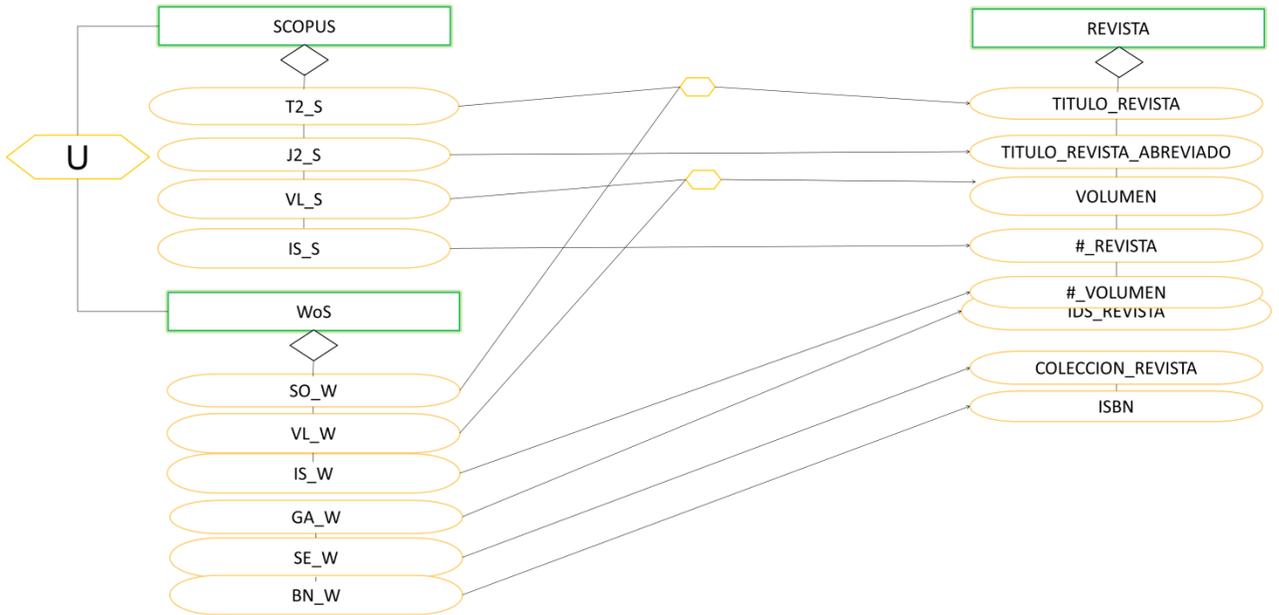


Imagen 48 Dimensión Revista

8.3.9 Dimensión Producto

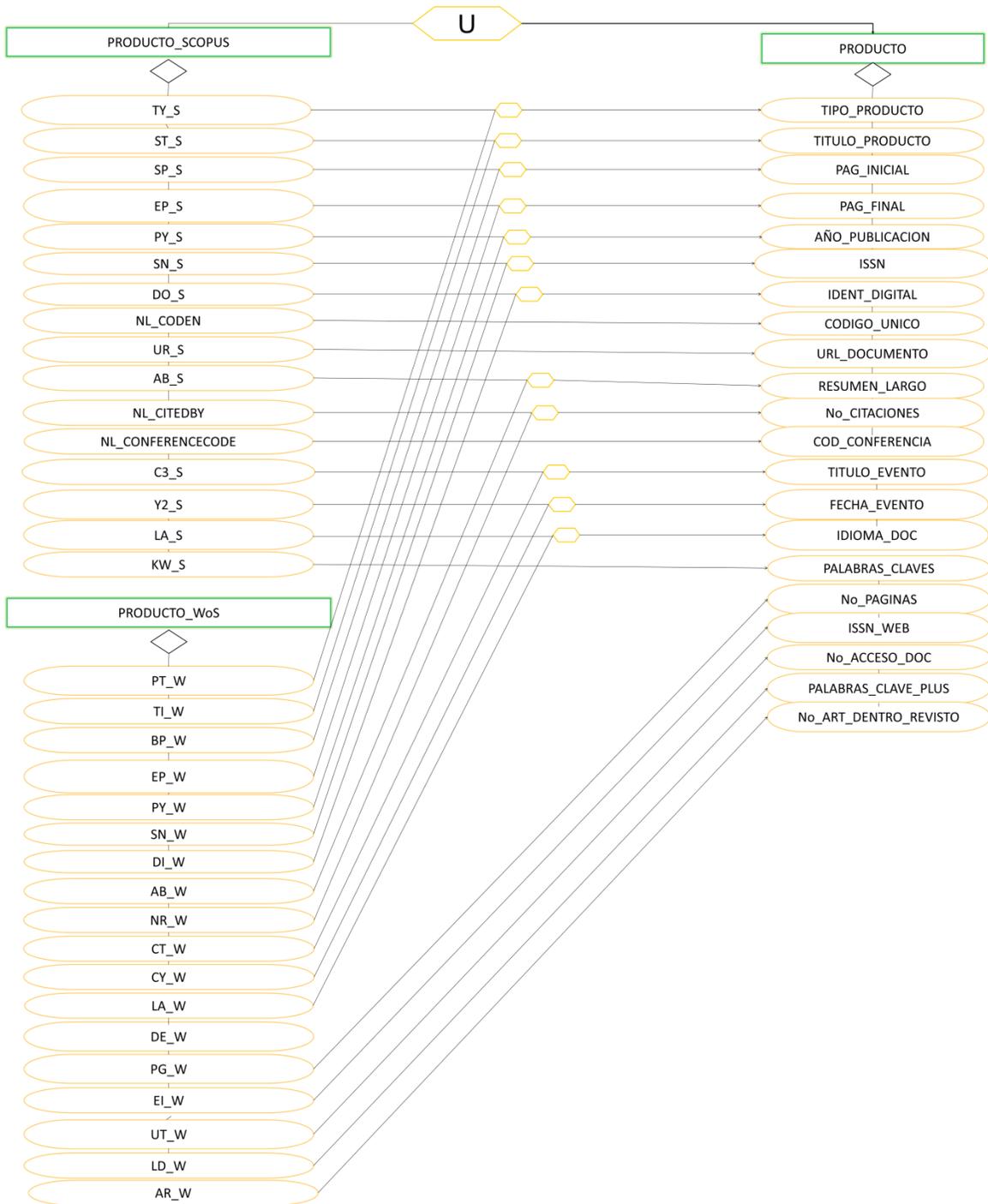


Imagen 49 Dimensión Producto

8.3.10 Dimensión Autor

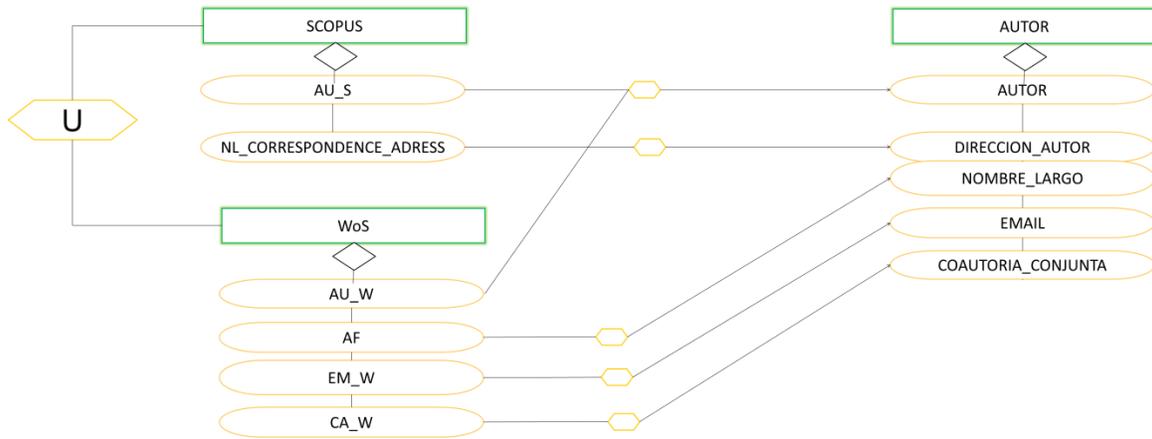


Imagen 50 Dimensión Autor

8.3.11 Dimensión Objetivo Socio Económico



Imagen 51 Dimensión Objetivo Socioeconómico

8.4 Tablas de Hechos

8.4.1 Tabla de Hechos de Inversión

8.4.1.1 Indicadores

Tabla 9 Gasto

Nombre	Gasto en CTI
Variables	<p>7. Valor Total del Gasto interno o intramuros</p> <p>a. Gastos Corrientes</p> <ul style="list-style-type: none"> • Total anual de salarios del personal de CTI • Total anual de remuneraciones del personal de CTI • Total anual de primas, vacaciones, contribuciones a fondos de pensiones, pagos de seguridad social, impuestos salariales, etc. • Total anual de compra de materiales, suministros y equipos de apoyo a las actividades ACTI que no hacen parte de los gastos de capital

	b. Gastos de Capital <ul style="list-style-type: none"> Total del capital fijo utilizado en programas de ACTI 8. Valor Total Gasto Extramuros 9. Tipo de Recursos 10. Tipo de Institución 11. Tipo de Actividad 12. Objetivo Socioeconómico	
Fuente de Información	Barrus	
Cálculo	Suma de total de gasto en Intramuros y Extramuros Por año	
Frecuencia	Anual	
Desagregaciones	Temática	<ul style="list-style-type: none"> Tipo de Recursos Tipo de Institución Tipo de Actividad Objetivo Socioeconómico
	Geográfica	<ul style="list-style-type: none"> Departamental
	Específica	<ul style="list-style-type: none"> Institucional
Tratamiento de las variables	Se debe calcular en pesos constantes	
Limitaciones del indicador	Se debe contar con la información real de la(s) institución(es) la información no es publica	

Tabla 10 Financiamiento

Nombre	Financiamiento en CTI	
Variables	<p>8. Total de recursos propios</p> <p>9. Total de recursos públicos</p> <ul style="list-style-type: none"> • Total de recursos de la nación • Total de recursos territoriales • Total de recursos por regalías <p>10. Recursos Privados</p> <ul style="list-style-type: none"> • Total de recursos por asociaciones • Total de recursos empresariales • Total de recursos internacionales • Total de recursos de otras entidades del conocimiento <p>11. Tipo de Recursos</p> <p>12. Tipo de Institución</p> <p>13. Tipo de Actividad</p> <p>14. Objetivo Socioeconómico</p>	
Fuente de Información	Barrus	
Cálculo	Suma de total de la financiación por año	
Frecuencia	Anual	
Desagregaciones	Temática	<ul style="list-style-type: none"> • Tipo de Recursos • Tipo de Institución • Tipo de Actividad • Objetivo Socioeconómico
	Geográfica	<ul style="list-style-type: none"> • Departamental
	Específica	<ul style="list-style-type: none"> • Institucional
Tratamiento de las variables	Se debe calcular en pesos constantes	
Limitaciones del indicador	Se debe contar con la información real de la(s) institución(es) la información no es publica	

8.4.2 Tabla de Hechos de Producción

8.4.2.1 Indicadores

Tabla 11 Producción

Nombre	Financiamiento en CTI	
Variables	9. Total de producción bibliográfica de autores vinculados a instituciones colombianas Web of Science 10. Total de producción bibliográfica de autores vinculados a instituciones colombianas Scopus 11. Total de producción bibliográfica de autores vinculados a instituciones colombianas SciELO 12. Tipo de Documento 13. Índice Documental 14. Área de Conocimiento OCDE 15. Disciplina 16. Coautoría	
Fuente de Información	Scopus, Web of Science, SciELO	
Cálculo	Conteo de productos por año	
Frecuencia	Anual	
Desagregaciones	Temática	<ul style="list-style-type: none"> • Tipo de Documento • Índice Documental • Área de Conocimiento OCDE • Disciplina • Coautoría
	Geográfica	<ul style="list-style-type: none"> • Departamental • Países de Latinoamérica
	Específica	
Tratamiento de las variables	<ul style="list-style-type: none"> • Se debe tener en cuenta la vinculación del autor a la institución. • Se debe tener en cuenta que el documento puede estar escrito en idiomas diferentes y ser el mismo estos documentos deben contar una sola vez 	
Limitaciones del indicador		

8.4.3 Tabla de Hechos de Investigadores

8.4.3.1 Indicadores

Tabla 12 Investigadores

Nombre	Investigadores	
Variables	9. Investigadores 10. Sexo 11. Fecha de Nacimiento 12. Nivel de Formación 13. Área de conocimiento OCDE 14. Institución 15. Tipo Institución 16. Clasificación OCyT (Activo vinculado a grupo, Activo no vinculado). 17. Entidad Territorial	
Fuente de Información	ScienTI, Snies	
Cálculo	Conteo de investigadores por año de acuerdo al criterio de actividad	
Frecuencia	Anual	
Desagregaciones	Temática	<ul style="list-style-type: none"> Área de conocimiento OCDE Tipo de Institución Máximo Nivel de Formación
	Geográfica	<ul style="list-style-type: none"> Departamental
	Específica	<ul style="list-style-type: none"> Sexo Rango de Edad
Tratamiento de las variables	<ul style="list-style-type: none"> Fecha de Nacimiento: Se extrae el año de nacimiento del investigador y se calcula la edad del mismo de acuerdo al año de consulta, luego se agrupan en los siguientes rangos: <ul style="list-style-type: none"> 0 a 20 21 a 30 31 a 40 41 a 50 51 a 60 > 60 Máximo Nivel de Formación: Se consulta el historial de formación del investigador y se selecciona el máximo nivel de formación. Tipo Institución: Se establece de la institución que avala al grupo de investigación al que pertenece el investigador. 	

Limitaciones del indicador	<ul style="list-style-type: none"> Ubicación geográfica: Se establece del departamento en donde se ubica la institución a la que pertenece el investigador.
----------------------------	--

9 DISEÑO DE VISUALIZACIÓN

CIENCIA, TECNOLOGÍA E INNOVACIÓN EN COLOMBIA

- INVESTIGADORES
- INVERSIÓN
- PRODUCCIÓN
- COMPARE INDICADORES

El presente documento hace parte del proyecto de grado titulado *DISEÑO DE DATAMART EN LAS TEMÁTICAS DE PRODUCCIÓN E INVERSIÓN DE CIENCIA TECNOLÓGICA E INNOVACIÓN EN COLOMBIA*, este es un diseño de lo que podría llegar a tener un sistema basado en la información integrada de las áreas de Producción, Inversión e Investigadores.

La información que verán a continuación fue tomada del Informe anual de indicadores de Ciencia y Tecnología 2015, que pueden encontrar en <http://ocyt.org.co/es-es/InformeAnualIndicadores/ArtMID/542/ArticleID/273/Indicadores-de-Ciencia-y-Tecnolog237a-Colombia-2015>.

Este diseño fue creado con el objetivo de dar una herramienta para poder iniciar la explotación analítica del diseño conceptual expuesto en el proyecto.

Imagen 52 Diseño de Inicio

CIENCIA, TECNOLOGÍA E INNOVACIÓN EN COLOMBIA

- INVESTIGADORES
- INVERSIÓN
- PRODUCCIÓN
- COMPARE INDICADORES

GENERALES

POR DEPARTAMENTO

INVESTIGADORES

Mujeres 7.294 63%

Hombres 4.320 37%

RANGO DE EDAD

ÁREAS OCDE

No DE INVESTIGADORES **11.566**

2014

■ Otros ■ Pregrado Universitario ■ Maestría ■ Doctorado

Imagen 53 Diseño de Tablero de Investigadores

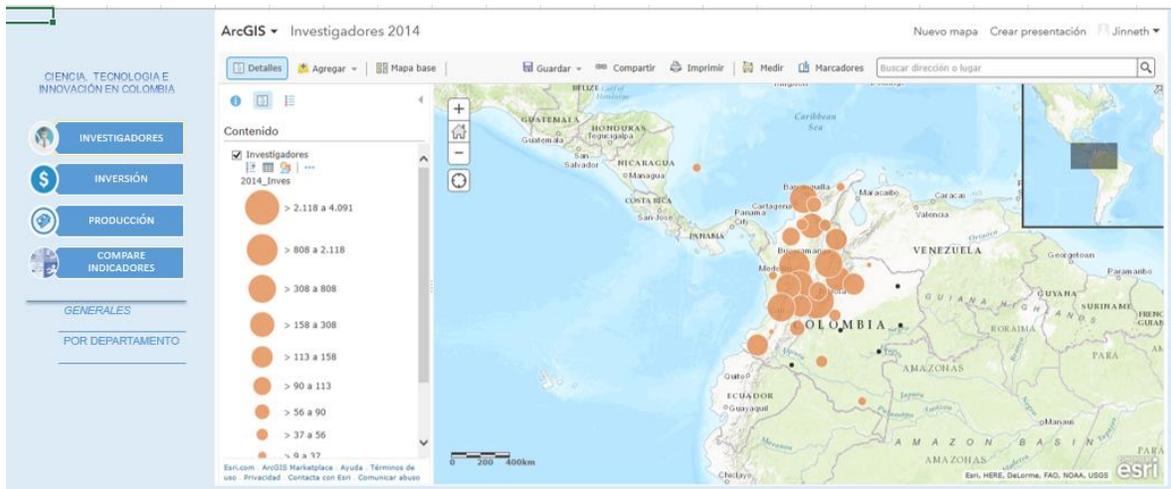


Imagen 54 Diseño de Tablero de Investigadores por Departamento

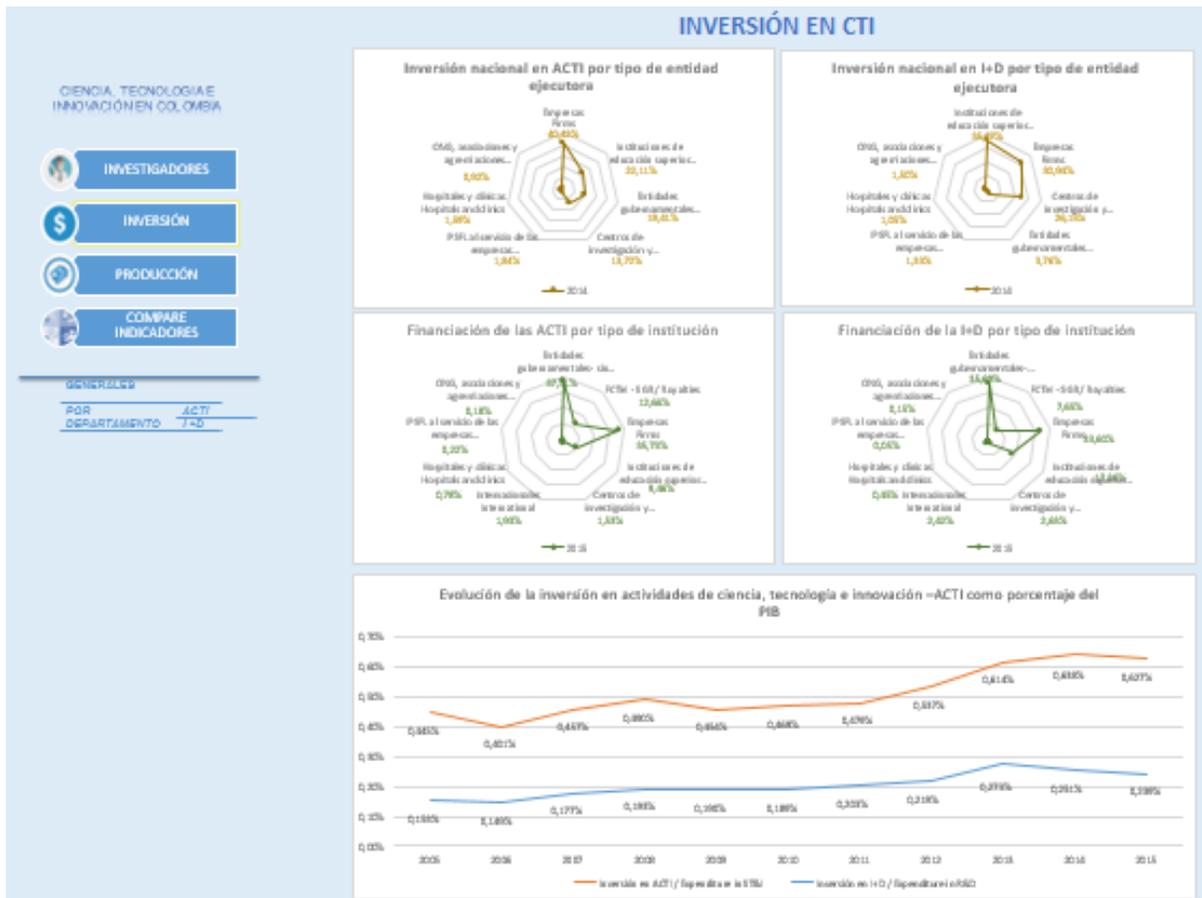


Imagen 55 Diseño de Tablero de Inversión

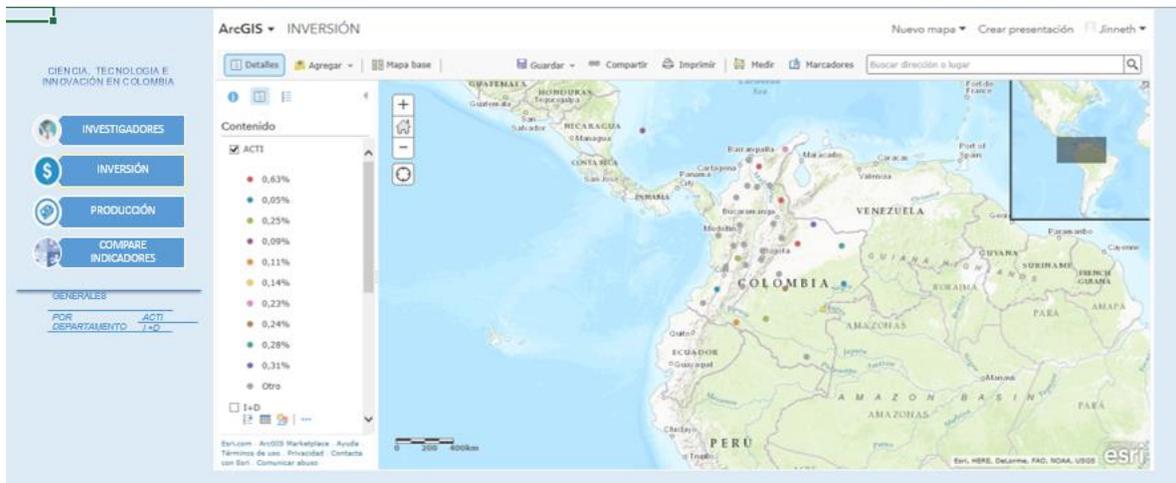


Imagen 56 Diseño de Tablero de Inversión por Departamento

PRODUCCIÓN EN CTI

CENSA, TECNOLOGÍA E INNOVACIÓN EN COLOMBIA

INVESTIGACIÓN

PRODUCCIÓN

CON NUEVOS INDICADORES

Gráficos POR DEPARTAMENTO



Diferencias en tipos de usuarios en la producción bibliográfica de autores vinculados a instituciones colombianas de acuerdo a cómo donde se encuentran los registros

Indicador	2006	2007	2008	2009	2010	2011	2012	2013	2014	Tendencia	
Metodología de Investigación	116	87	124	244	337	374	388	433	384	[Line Chart]	
	224	268	882	864	1,848	1,188	1,238	1,283	1,284	[Line Chart]	
	218	824	1,882	1,428	1,428	1,728	1,418	2,314	2,844	2,468	[Line Chart]
Metodología de Investigación en Salud	272	828	624	768	837	983	1,182	1,126	1,892	983	[Line Chart]
	818	811	998	1,278	1,882	1,882	1,461	2,888	2,188	2,124	[Line Chart]
	174	214	298	296	484	497	626	626	646	646	[Line Chart]
Bases de Datos	128	221	284	264	464	884	621	744	822	824	[Line Chart]
	882	724	982	1,828	1,722	2,126	2,294	2,788	2,872	2,214	[Line Chart]
	986	1,188	1,262	1,782	2,182	2,242	2,898	2,144	2,282	2,881	[Line Chart]

* Usuarios de Bases de Datos: Metodología de Investigación, Metodología de Investigación en Salud y Bases de Datos.
 Usuarios de Bases de Datos: Metodología de Investigación, Metodología de Investigación en Salud y Bases de Datos.
 * Metodología de Investigación: Metodología de Investigación y Metodología de Investigación en Salud.
 Metodología de Investigación: Metodología de Investigación y Metodología de Investigación en Salud.

Imagen 57 Diseño de Tablero Producción

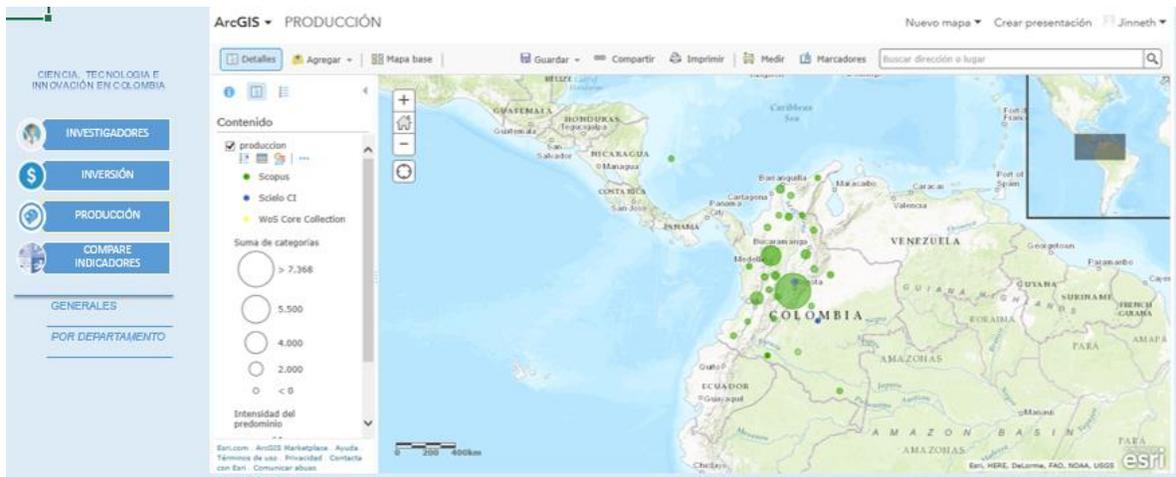


Imagen 58 Diseño de Tablero Producción por Departamento

10 CONCLUSIONES

El SNCTI cuenta con información valiosa y relevante que debe ser integrada y estandarizada para apoyar la toma de decisiones.

Si la información estuviera disponible desde las entidades territoriales las decisiones se tomarían informadas y con datos reales.

La información no solo debe ser integrada, también debe ser íntegra y de calidad. Para ello es necesario ajustar, depurar y verificar los sistemas que capturan la información.

No toda la información del SNCTI cuenta con la calidad apropiada, por lo que se recomienda realizar un análisis de calidad por fuente de información con la idea de fortalecer y enriquecer la información con la que cuenta SNCTI.

La relación entre investigadores y autores no se identifica fácilmente puesto que los investigadores no firman sus documentos de la misma forma siempre, por lo que se debe capacitar a los investigadores para que la firma sea identificada y fácilmente identificada la relación de su producción. Así mismo, se debe realizar la depuración de las diferentes firmas de los investigadores en las fuentes de información.

BIBLIOGRAFIA

- BAGLIETO A., BALLESTEROS A., BARCELÓ M., CORREAS J., FERNANDEZ P., GOMEZ S., *et al.* Hacia una economía del conocimiento. Price water house Cooper. 2001.
- CINDI HOWSON. Successful Business Intelligence: Secrets to Making BI a Killer App. Hardcover. 2008
- ELZBIETA MALINOWSKI · ESTEBAN ZIM'ANYI. Advanced Data Warehouse Design. Springer. 2008.
- INMON W. Building the Data Warehouse. Wiley. 2002.
- INSTITUTO COLOMBIANO DE NORMAS TECNICAS. Normas Colombianas para la presentación de trabajos de investigación. Sexta actualización. Bogotá D.D., ICONTEC, 2008 NTC 1486.
- KIMBALL R., ROSS M. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. Wiley. 2002
- KIMBALL R. The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses. John Wiley&Sons. SecondEdition. 2002.
- KIMBALL R., REEVES L., ROSS M., THORNTHWAITE W. The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses. Wiley. 2002
- LARISSA T. Moss. Business Intelligence Roadmap. Addison-Wesley. 2003.
- MONROY S. REV. INNOVAR. Vol. 16, No. 28, JULIO A DICIEMBRE DE 2006. Nuevas políticas y estrategias de articulación del sistema de ciencia, tecnología e innovación colombiano.
- RAFANELLI M. Multidimensional Databases: Problems and Solutions. Idea Group Publishing. 2003. ISBN: 1-59140-086-4.

- RITACCO M., CARVER A. The business value of business intelligence: a framework to measuring the benefits of business intelligence. Business Objects. 2007.
- SID A, LARISSA T MUSGO. Data Warehouse Project Management. Addison-Wesley. 2000.
- MÓNICA SALAZAR ACOSTA, JORGE LUCIO, ÁNGELA RIVERA, CAMILO DÍAZ. Metodología para el cálculo del gasto en ciencia, tecnología e innovación en Colombia. Estudio realizado para la Dirección de Desarrollo Empresarial del Departamento Nacional de Planeación, y el Ministerio de Hacienda. 2006.
- JORGE LUCIO, SANDRA CAROLINA RIVERA TORRES, JINNETH TIQUE ORTIZ, DIANA LUCIO-ARIAS, HENRY MORA HOLGUÍN, CLARA INÉS PARDO MARTÍNEZ, ANDREA GUEVARA, GLORIA INÉS PEREA, DANIEL RICARDO TORRALBA BARRETO, MARIA PAULA GARAVITO MUÑOZ, FABIÁN HERNÁN MELO MARTÍNEZ, NATALIA MÁRQUEZ-BUSTOS, SANDRA MILENA ZÁRATE RINCÓN, NAYIBE CASTRO NOVOA. Indicadores de ciencia y tecnología Colombia 2015.2016.

REFERENCIAS ELECTRONICAS

- GARTNER GROUP. Howard Dresner Biography, disponible en línea en [Http://www.gartner.com/research/fellows/asset_79427_1175.jsp](http://www.gartner.com/research/fellows/asset_79427_1175.jsp)
- TORLONE R. Conceptual Multidimensional Models. Università Roma Tre. Disponible en línea en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.6570&rep=rep1&type=pdf>.
- LEONARDO TORRES OCHOA. Resumen de Procesos de PMBOK, disponible en línea [Http://www.scribd.com/doc/16671508/Resumen-de-Procesos-de-PMBOK](http://www.scribd.com/doc/16671508/Resumen-de-Procesos-de-PMBOK).
- http://docs.politicascsti.net/reportes/CO_SI.pdf
- <http://ocyt.org.co/es-es/informacion-institucional>
- <http://www.colciencias.gov.co/scienti>
- <http://www.ricyt.org/indicadores>
- <http://indicadores.micit.go.cr/dashboards/7953/inversion-en-act-e-id-2011/>
- <http://bi.mineduacion.gov.co:8380/eportal/web/snies1/departamento-de-domicilio-de-la-ies>
- <http://bi.mineduacion.gov.co:8380/eportal/web/men-observatorio-laboral/ubicacion-geografica>