

LA REVOLUCIÓN DE LOS DATOS BIBLIOGRÁFICOS, CIENTÍFICOS Y CULTURALES

ARIEL ALEJANDRO RODRÍGUEZ GARCÍA
Coordinador



Z666.7

R47

La revolución de los datos bibliográficos, científicos y culturales / Coordinador Ariel Alejandro Rodríguez García.
- México : UNAM. Instituto de Investigaciones Bibliotecológicas y de la Información, 2020.

xv, 346 p. – Colección: Metadatos

ISBN: 978-607-30-2996-4

1. Metadatos bibliográficos. 2. Datos vinculados. 3. Big data 4. Repositorios institucionales. 5. BIBFRAME (Modelo conceptual). I. Rodríguez García, Ariel Alejandro, coordinador. II. ser.

Diseño de cubierta: Oscar Daniel López Marín

Primera edición, 2020 D.R. © UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO

Ciudad Universitaria, 04510, Ciudad de México

Impreso y hecho en México

ISBN: 978-607-30-2996-4

Publicación dictaminada

Contenido

INTRODUCCIÓN	ix
1. DATOS BIBLIOGRÁFICOS	1
I. Retro y prospectiva de la investigación bibliotecológica sobre metadatos	3
ARIEL ALEJANDRO RODRÍGUEZ GARCÍA	
II. <i>BIBFRAME</i> : Un modelo de metadatos para la web semántica	17
FILIBERTO FELIPE MARTÍNEZ ARELLANO	
III. Metadatos, datos enlazados e interoperabilidad: reflexiones en el dominio bibliográfico	33
FABIANO FERREIRA DE CASTRO	
IV. De los metadatos para la organización de la Información a la tecnología <i>middleware</i> para los servicios de las bibliotecas: la biblioteca inteligente	63
GERARDO BELMONT LUNA	
V. Enlazando datos: un modelo conceptual orientado a objetos para el diseño de bases bibliográficas	79
LIZBETH BERENICE HERRERA DELGADO	
VI. Informe desde las trincheras: Transformando unidades de catalogación para incluir servicios de repositorio	113
LISA FURUBOTTEN	

2. DATOS CIENTÍFICOS141

I. Factores determinantes para la implementación
del esquema de metadatos para repositorios de datos
de investigación de la Política de Ciencia Abierta en México143

MIGUEL ADOLFO GUAJARDO MENDOZA

II. Habilidades requeridas por el profesional de la Industria 4.0.
Aparición de un nuevo perfil para el área de producción,
en la cuarta Revolución Industrial.161

MARCO BRANDÃO

III. Implementación de un repositorio unificado para
conformar el Sistema de Inteligencia Institucional del INE183

ALEJANDRO VERGARA TORRES

IV. Calidad en los metadatos:
reto para asegurar la participación en repositorios
nacionales manteniendo los estándares institucionales205

ALMA BEATRIZ RIVERA AGUILERA, ELISA CRUZ ROJAS,

BENJAMÍN ALEJANDRO GUERRERO OLIVERA Y ÓSCAR ALFONZO PEREYRA

V. Interoperabilidad en plataformas
de publicación y distribución de libros digitales.....223

JENNY TERESITA GUERRA GONZÁLEZ

3. DATOS CULTURALES	245
I. Metadatos para preservación digital	247
JUAN VOUTSSÁS MÁRQUEZ	
II. El Sistema de Información para el Registro Universitario de Espacios y Activos Culturales (SI-RUEyAC)	265
CATALINA NAUMIS PEÑA, NATALIA VELAZCO PLACENCIA Y ARIEL ALEJANDRO RODRÍGUEZ GARCÍA	
III. La ciudadanía digital y sus mediciones: el caso del estado de ánimo de los tuiteros en México	283
HÉCTOR ALEJANDRO RAMOS CHÁVEZ	
IV. El papel de los metadatos en la construcción de datos enlazados para bibliotecas	301
EDER ÁVILA BARRIENTOS	
V. La integración del usuario final en la creación de metadatos	325
PATRICIA HERNÁNDEZ SALAZAR	

Implementación de un repositorio unificado para conformar el Sistema de Inteligencia Institucional del INE

ALEJANDRO VERGARA TORRES
Instituto Nacional Electoral. México

INTRODUCCIÓN

El Instituto Nacional Electoral (INE) es el organismo autónomo encargado de organizar las elecciones presidenciales, de senadores y diputados federales en México; también es el encargado de administrar el listado nominal y padrón electoral mediante el cual a los ciudadanos se les provee de una credencial para votar que hoy en día funciona como el medio de identificación personal. El INE también se encarga de la capacitación de los ciudadanos que fungen como funcionarios de casilla el día de la Jornada Electoral, fiscaliza los recursos públicos que utilizan los partidos políticos en sus campañas y es el encargado de garantizar el uso de sus prerrogativas en radio y televisión, entre otras funciones que derivan en una gran acumulación de información. El presente capítulo da cuenta de las diferentes acciones realizadas para la implementación de un repositorio unificado de información a partir del cual se conformó el Sistema de Inteligencia Institucional del Instituto Nacional Electoral.

Diversos factores, entre ellos la acumulación de grandes volúmenes de datos, han hecho del análisis de datos una tecnología esencial y cada vez más necesaria. Las empresas y organizaciones que procesan su información mediante herramientas analíticas se vuelven más productivas y rentables que otras al explotar de mejor manera la información que producen, les rodea y poseen, y utilizarla a su favor para la toma de decisiones. Por ende, la adopción de herramientas analíticas se está acelerando y cobrando mayor relevancia en muchas organizaciones, y más aún, tomando en consideración los debates del gobierno abierto y la ciudadanía digital.

Sin embargo, uno de los retos mayores, que conlleva un gran esfuerzo, es el de recabar y concentrar la información histórica y presente en un solo repositorio; asegurar su homogeneidad, calidad y consistencia es tan solo el primer preparativo para poder explotarla. Por ello, un primer esfuerzo importante para analizar múltiples fuentes de información es contar con un repositorio unificado en donde se concentren y convivan diversas fuentes de datos con una estructura homogénea y correctamente relacionada para poder realizar análisis de esas grandes cantidades de información ya integradas.

El Sistema de Inteligencia Institucional (SII) a grandes rasgos se compone de: 1) un repositorio unificado de información o *dataware house*; 2) un componente analítico que permite el modelado y análisis de la información que alberga, y 3) un componente de visualización con el cual pueden producirse diversos gráficos y análisis visuales. Este sistema es una herramienta informática que favorece la cultura analítica dentro del Instituto, cuyo objetivo principal es integrar en un solo repositorio la información que se ubica en bases de datos históricas y presentes de los diversos sistemas operados por el Instituto Nacional Electoral, así como diversas

fuentes de datos externas (Instituto Nacional de Estadística y Geografía o bien información producida por el Instituto Federal de Telecomunicaciones), previamente depuradas y homogeneizadas para facilitar su consulta y análisis, lo que permite la explotación de información a partir de la construcción de diversos productos de información (tableros de control, gráficas, consultas dinámicas, mapas y herramientas visuales, entre otros).

El Sistema de Inteligencia Institucional le permite al Instituto:

- Contar con una base de conocimiento de la información que se produce en los sistemas de información que permite conocer y comparar información histórica o que provenga de otros procesos electorales.
- Lograr la homogeneidad de la información al cruzar datos cartográficos, relacionados con procesos electorales (locales o federales) y con base en un calendario proveniente de diversos sistemas de información.
- Obtener datos de diversos sistemas de información y de áreas operativas distintas.
- Contar con un modelo integral que permita acoplar los datos contenidos en los sistemas de información con los que cuenta el Instituto.
- Garantizar la permanencia del conocimiento concentrada en un repositorio institucional único.
- Contar con una herramienta que realiza diversos análisis de información que pueden ser de índole cuantitativo, predictivo, de comportamiento o ponderativos.

- Consolidar la información institucional a pesar de que los sistemas informáticos desarrollados para cada área del Instituto no guarden relación directa.
- Cotar a las áreas del Instituto de un herramienta con capacidades tecnológicas sin precedente que permita hacer eficiente y fortalecer la toma de decisiones.

El objetivo de este capítulo se centra en compartir las diferentes etapas realizadas para la implementación de este proyecto, las grandes dificultades que conlleva un proyecto de esta magnitud, su gestión y relevancia para que proyectos como éstos logren construir acervos de información de calidad explotables y permitan salvaguardar el conocimiento acumulado de la mejor manera para generar un valor agregado para las organizaciones y para la ciudadanía mediante una mejor toma de decisiones.

SOBRE EL ANÁLISIS DE DATOS, LA CIENCIA DE DATOS, LA INTELIGENCIA DE NEGOCIOS Y EL *BIG DATA*

Comprender el uso de herramientas analíticas, en particular aquellas relacionadas con la inteligencia de negocios, requiere conocer un poco del contexto en donde este tipo de soluciones se ubican dentro del amplio campo del análisis de datos, por ello a continuación se describirán los diferentes tópicos relacionados con el análisis de datos.

El primer término que es necesario comprender es el de la ciencia de datos, el cual relaciona y da origen a la gran mayoría de las herramientas analíticas hoy conocidas. David Donoho (2017), en su artículo “50 years of Data Science” da

cuenta del origen de la necesidad de analizar los datos; como el autor lo cuenta, hace más de cincuenta años, John Tukey propuso reformar la estadística académica y en su publicación “The Future of Data Analysis” puntualiza sobre la existencia de una aún no reconocida ciencia cuyo objetivo era “aprender de los datos” o bien del “análisis de los datos”. Diez o veinte años después, John Chambers, Bill Cleveland y Leo Breiman promueven dentro de círculo de la academia de la estadística expandir sus límites más allá del dominio clásico de la teoría estadística.

Chambers llamó la atención de la academia para enfatizar la preparación de los datos más que el modelado estadístico, y Beiman promovió un mayor énfasis en la predicción que en la inferencia. Cleveland incluso sugirió el pegajoso nombre Data Science para ese campo.

En este sentido y como John D. Kelleher y Brendan Tierney (2018) refieren en su libro *Data Science*, nunca ha sido más fácil que hoy en día para las organizaciones recopilar, almacenar y procesar datos. El uso de la ciencia de datos está impulsado por el aumento de los grandes datos y las redes sociales, el desarrollo de la informática de alto rendimiento y la aparición de métodos tan potentes para el análisis y modelado de datos como aprendizaje profundo. La ciencia de datos abarca un conjunto de principios, definiciones de problemas, algoritmos y procesos para extraer patrones no obvios y útiles de grandes conjuntos de datos y está estrechamente relacionado con los campos de la minería de datos y el aprendizaje automático, con un alcance aún más amplio.

Como Keheller y Tierney (2018) refieren en la introducción de su libro, el registro de información se distingue en dos grandes grupos, 1) datos transaccionales, los cuales incluyen información de cada evento, la venta un artículo, la entrega

de algún bien, etcétera, y 2) datos no transaccionales, como los datos demográficos.

El desarrollo de las computadoras ha contribuido a un masivo incremento de información que es recolectada y almacenada en bases de datos con tablas de un campo por instancia y una columna por atributo; en este sentido, las bases de datos son la tecnología natural para obtener transacciones estructuradas o datos operacionales. Sin embargo, hoy en día, en razón de la gran variedad de datos que se generan, empezaron a surgir dificultades para analizar esas grandes cantidades de información. Uno de los principales problemas era que los datos se encuentran almacenados en múltiples bases de datos independientes entre sí. Esto provocó el desarrollo de una tecnología llamada *data warehouses* o repositorios de datos en donde la información es obtenida a lo largo de la organización e integrada para que pueda ser analizada de una mejor manera. Éste es el origen de las herramientas analíticas de inteligencia de negocios, las cuales se dedican a procesar múltiples fuentes de datos estructurados provenientes de bases de datos de una organización para su cruce, integración y posterior análisis. Este tipo de tecnología es la que se empleó para la integración del Sistema de Inteligencia Institucional y que más adelante se detallará ampliamente.

Adicionalmente, en las últimas dos décadas el uso de dispositivos móviles conectados a Internet, el crecimiento y la masificación de las redes sociales combinado con diversas plataformas en línea han tenido un impacto dramático en la cantidad de datos recopilados. No solo ha aumentado la cantidad de datos, sino también su diversidad, fotografías, tuits, correos electrónicos, mensajes, videos, imágenes, etcétera, todos ellos relacionados con los metadatos (datos que describen la estructura y las propiedades de los datos en bruto) de cada uno de estos eventos, a partir de lo cual podemos empezar a comprender

el significado del término *big data*, que a menudo se define en términos de las tres “V”: el Volumen extremo de datos, la Variedad de los tipos de datos y la Velocidad a la que deben procesarse los datos. Esto ha derivado en el desarrollo de nuevas tecnologías de consulta y búsqueda de información que no se encuentre estructurada como sí lo están las bases de datos relacionadas, es decir, las nuevas tecnologías de *big data* se enfocan en procesar grandes bases de datos no relacionadas. Es decir, que la existencia de *big data* ha promovido el desarrollo de nuevos esquemas de procesamiento de datos ya que al manejarse grandes cantidades de información a velocidades muy grandes puede ser muy útil desde una perspectiva computación y de velocidad distribuir los datos en múltiples servidores.

El *big data* es sólo un gran volumen de datos concentrado, pero el valor agregado es el conocimiento que obtenemos a partir de estos, lo cual sólo se logra a partir de estructurarlos, modelarlos y explotarlos para influir valor en la toma de decisiones.

En resumen, podríamos decir que la ciencia de datos ha permitido el desarrollo de diversas tecnologías, algoritmos, modelos de procesamiento de información para trabajar tanto con información estructurada como no estructurada, y en ese sentido podemos decir que la ciencia de datos engloba múltiples disciplinas relacionadas con el estudio de los datos, y *big data* es una de ellas, pues se concentra en optimizar la explotación de fuentes de datos no relacionadas y relacionadas haciendo uso de algoritmos precisos de análisis y explotación de información y usando múltiples computadoras para su procesamiento mientras que las herramientas de inteligencia de negocios están enfocadas a explotar y consultar múltiples bases de datos de una organización, para lo cual hace uso de tecnologías como *data warehouses* para concentrar la

información en un repositorio único que resuelva la problemática de múltiples fuentes de datos que no guardan una relación estrecha.

Es así que el Sistema de Inteligencia Institucional pretende integrar en un solo repositorio las bases de datos relacionadas y diversas fuentes adicionales de información estructurada para que puedan ser consultadas, procesadas y analizadas de manera eficiente, bajo un modelo de datos multidimensional unificado; es decir, gobernado por un modelo de datos global que permita relacionar todas las fuentes que se desee analizar con la garantía de que sean homogéneas, que puedan intersectarse y que se eliminen las inconsistencias que surgen de contar con múltiples bases de datos dispersas.

IMPLEMENTACIÓN DE UN REPOSITORIO UNIFICADO

La inteligencia de negocios es una práctica que busca fortalecer la toma de decisiones a partir de la utilización de información relevante de alta calidad, y maximizar su valor a través del análisis, la definitividad y la interpretación.

En el contexto de los sistemas de información, la Inteligencia de negocios es una tecnología que facilita la recopilación y el procesamiento de datos para identificar correlaciones y patrones significativos para apoyar las decisiones dentro de una organización.

El Dataware House Institute describe que la inteligencia de negocios une datos, tecnología, análisis y conocimiento humano para optimizar las decisiones organizacionales y, en última instancia, impulsar el éxito y logro de metas en una organización. Los programas de inteligencia de negocios suelen combinar un almacén de datos de empresa y una plataforma de inteligencia de negocios o conjunto de

herramientas para transformar datos en información organizacional útil.

En este sentido, el Sistema de Inteligencia Institucional fue concebido como un proyecto que relaciona información entre el pasado, presente y futuro para la toma de decisiones estratégicas. Una institución de gobierno tan grande y relevante como el Instituto Nacional Electoral cuenta con áreas ejecutivas de administración (ejecutivos), funcionarios (personal operativo y de servicio), clientes o receptores de los servicios (ciudadanos, partidos políticos y sujetos regulados) y productos (servicios proporcionados; por ejemplo, la credencial para votar).

En esa labor diaria, el INE produce grandes cantidades de información que, como en la gran mayoría de instituciones y empresas, se encuentra dispersa en múltiples bases de datos independientes entre sí, por lo cual no es posible comparar el pasado con el presente fácilmente y en ese sentido la cualidad de los acervos no permite obtener información con valor analítico.

Por ello, el potencial del proyecto reside en relacionar la diversa información que el INE posee y ha generado en el pasado y sigue produciéndose en el presente, conciliarla y concentrarla en un único repositorio, relacionarla y permitir su procesamiento y análisis para la toma de decisiones estratégicas. Es decir, concebir un proyecto que permitiera entretener la diversa información con que cuenta el Instituto en sus bases de datos históricas y presentes para que puedan ser consultadas a partir de herramientas analíticas que doten a la institución de capacidades analíticas que más adelante permitan a la ciudadanía y otros sujetos la explotación de esas nuevas capacidades de análisis. Porque una forma de lograr la efectividad en la administración pública es utilizar tecnología que soporte y coadyuve en una mejor

toma de decisiones, y qué mejor que las decisiones se basen en información, hechos y datos existentes.

En resumen, las herramientas analíticas fortalecen y mejoran la toma de decisiones porque en la evolución de la sistematización de procesos el primer paso es el sistematizar procesos, y al hacerlo logramos sistematizar la operación diaria de una labor en particular, conseguimos que cada transacción quede registrada y con ello somos capaces de guardar la historia en la operación de un proceso operativo; con ello dotamos a las instituciones de sistemas operacionales que nos permiten conocer el ejercicio diario de sus acciones.

Pero eso no es suficiente para dotar a una institución de capacidades analíticas, la sistematización de información no es una labor suficiente para mejorar la toma de decisiones, la clave está en dar el siguiente paso, que es analizar la información; para ello es necesario contar con mecanismos y herramientas que nos permitan procesar múltiples fuentes de datos, conciliar la información histórica y presente de una organización, y asegurar que la obtención de respuestas a preguntas claves puedan ser viables. Para ello, la primera decisión era crear un repositorio único donde pudieran integrarse las múltiples bases de datos que posee el INE y generar un meta modelo, un modelo multidimensional, que permitiera cruzar y llevar a cabo diversas consultas de información para generar productos analíticos.

Al igual que en una gran biblioteca, esa labor fue necesaria para poder concentrar en una sola herramienta el acervo de información generado por las múltiples bases de datos que genera la institución en un primer ejercicio que le permita al INE fortalecer sus capacidades analíticas de la información que posee para que no sólo pueda saber “qué pasó” y “qué está pasando”, lo cual se logra en primer nivel implementando sistemas operacionales, sino poder

anticiparse al “qué puede pasar” y “qué debe planearse” con un enfoque estratégico y no solo operativo, con base en indicadores vinculados al negocio identificado en cada proceso sistematizado y vinculado.

IMPLEMENTACIÓN DEL PROYECTO

Un sistema de inteligencia de negocios está integrado de tres grandes componentes:

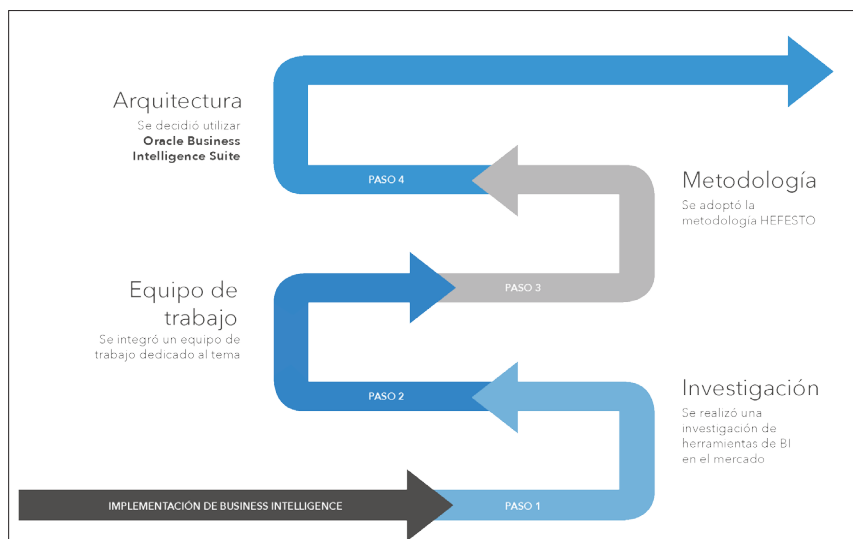
1. Un repositorio único o central (*data warehouse*), el cual sirve como una gran biblioteca centralizada de información donde se integran las múltiples fuentes de datos estructuradas (es decir bases de datos, información obtenida de archivos Excel, CSV u otras fuentes que tengan un estructura de orden previo). En este repositorio se diseñará un modelo multidimensional, el cual formará parte del gran metadato que describa las relaciones que vayan descubriendo al vincular cada una de las bases de datos que integren y cada uno de los datos de estas múltiples bases de datos.
2. Un componente analítico. Herramientas analíticas que llevan a cabo la extracción y carga de información al repositorio único que genera el modelo multidimensional y adicionalmente un modelo basado en cubos de información e indicadores a partir de Cubos OLAP (On Line Analytical Process) mediante el cual se pueden agilizar las consultas de grandes cantidades de información a partir de la estructura multidimensional y no sobre la tradicional estructura rela-

- cional de las bases de datos.
3. Un reporteador o visualizador. Herramienta orientada a usuarios de negocio para que exploren y analicen la información que necesitan de manera rápida y sencilla. Permite crear visualizaciones basadas en distintos conjuntos de datos y mostrar el resultado en la misma página, como si fuera una combinación entre respuestas a preguntas y tableros de control.

Para implementar un proyecto de inteligencia de negocios, fue necesario ejecutar los siguientes pasos:

A partir de la fase inicial del proyecto, se distinguieron las

Gráfico 1. Pasos previos
a la implementación de un proyecto de inteligencia de negocios



Fuente: Elaboración propia.

necesidades tecnológicas y del grupo de trabajo que llevaría a cabo la gestión del proyecto, y se resaltó la importancia de la ejecución de cada paso:

Paso 1. Es necesario realizar una investigación previa para conocer la situación actual en el mercado tecnológico sobre herramientas de inteligencia de negocio que puedan implementarse.

Paso 2. Integrar un equipo de trabajo es esencial, ya que un proyecto de estas características demanda amplios conocimientos en materia de análisis de datos, así como de administración de la infraestructura donde se ejecutarán estas herramientas.

Paso 3. Identificar metodologías y un marco de referencia permitirá tener una idea clara de ejercicios previos; para este proyecto se implementaron dos metodologías. La primera es HEFESTO, metodología creada por Bernabeu Ricardo Darío disponible bajo licencia GNU FDL, la cual se fundamenta en una amplia investigación, comparación de metodologías existentes y experiencias propias en procesos de confección de almacenes de datos. Consta de cuatro fases: análisis de requerimientos, análisis de las bases de datos orientadas al procesamiento de transacciones OLTP (On-Line Transactional Processing), modelo lógico del almacén de datos y proceso ETL (Extract-Transform-Load), que corresponde a las actividades de extraer los datos de las fuentes OLTP. Asimismo, se tomó como referencia a The Dataware Housing Institute (TDWI), organización que por más de veinte años se ha dedicado a la investigación y asesoría sobre inteligencia de

negocios (*business intelligence*), *data warehousing* y analíticos alrededor del mundo. El objetivo de investigación de este organismo se centra exclusivamente en diversas situaciones a las que se enfrentan líderes y ejecutantes para entregar un profundo y amplio entendimiento del negocio y los retos técnicos involucrados en la implementación y el uso de inteligencia de negocios, *data warehousing* y soluciones analíticas. El TDWI creó el Modelo de Madurez Analítico (Analitycs Maturity Model) con el objetivo de ayudar a la necesidad que tienen las organizaciones de conocer de calidad de las implementaciones de soluciones analíticas en comparación con otras para identificar las mejores prácticas y basarse en ellas. Por lo tanto, este modelo sirve para determinar en qué nivel se encuentra una institución o empresa con respecto al grado de madurez de adopción de un modelo analítico al interior de la organización.

Paso 4. Definir la arquitectura tecnológica que se empleará para la explotación de información, el proceso completo de carga y la publicación de tableros y diversos análisis que puedan ser consultados, la cual involucre ambientes específicos de desarrollo, pruebas y producción.

A partir de este marco de referencia y tomando en consideración los pasos definidos por la metodología HEFESTO, se pueden iniciar los trabajos para integrar las múltiples fuentes de información y lograr la construcción de análisis de información identificados, para lo cual se ejecutarán las siguientes fases:

La primera es analizar e identificar los requerimientos analíticos y descubrir las preguntas clave a responder para así ubicar los indicadores y las perspectivas de análisis que será necesario mapear hacia las fuentes de datos origen y con ello lograr un modelo conceptual.

La segunda fase comprende el análisis de las fuentes de datos a cargarse en el repositorio único. Para determinar los indicadores a implementarse, se establecen las correspondencias de estos a cada uno de los campos de información que provienen de las fuentes de datos origen y se crean los niveles de granularidad; es decir, se ubican los campos que contendrá cada perspectiva, ya que a partir de ellos se examinarán y filtrarán los indicadores, lo que logrará un modelo ampliado.

La tercera fase corresponde a la generación de un modelo lógico del *dataware house*, en donde se diseñan las tablas de dimensiones, las tablas de hechos y se detallan las uniones; es decir, se genera el mapa lógico final que será creado en el repositorio único donde se construirá el modelo que albergue los datos provenientes de las diversas fuentes.

La cuarta fase corresponde a la creación de los procesos de carga y extracción de los datos origen para poblar el repositorio único tomando como base el modelo lógico recién creado en el repositorio.

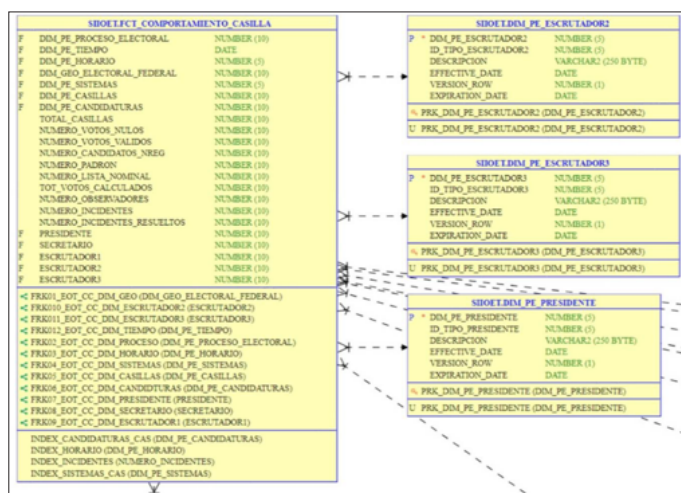
Este camino de cuatro fases deberá ejecutarse para cada una de las fuentes de información que pretenden colocarse en el repositorio único. Una labor esencial es realizar una correcta gestión del modelo multidimensional diseñado. Algo muy importante es ubicar indicadores que se identifiquen como datos maestros que son empleados por múltiples fuentes de datos y que gobiernan la relación entre ellos; por ejemplo, algunas dimensiones maestras que son comunes entre múltiples fuentes de datos son el tiempo, elementos

geográficos y catálogos genéricos que eran empleados por el modelo lógico para permitir las uniones entre múltiples fuentes de datos de manera homologada y bien definida.

Este trabajo permite contar con un modelo multidimensional capaz de relacionar múltiples fuentes de datos que ahora pueden ser consultadas a partir de una perspectiva de cubos de información en indicadores; es decir, se logró abstraer la complejidad de los modelos de datos operacionales y se implementó un modelo más afín al negocio, lo que resolvió así la complejidad de consultas y la dispersión de las fuentes de información.

A continuación, se presenta como ejemplo un fragmento del modelo multidimensional, en particular refiriendo a información obtenida de una fuente de datos histórica donde se registraron las cifras referentes al conteo de actas de escrutinio y cómputo de una elección:

Gráfico 2. Fragmento del modelo multidimensional correspondiente al modelo de datos de información correspondiente al conteo de actas



Fuente: Elaboración propia.

Asimismo, se presenta una vista de una estrella contenida dentro de la constelación total del modelo multidimensional:

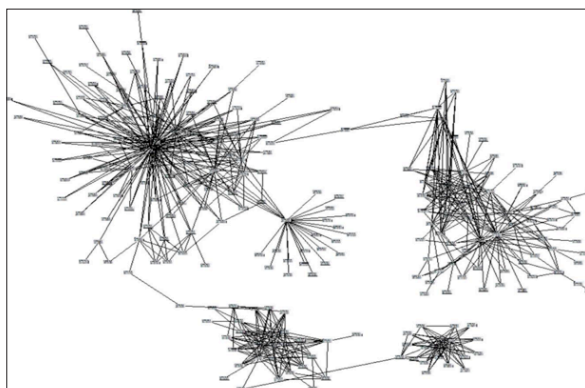
Gráfico 3. Fragmento del modelo multidimensional correspondiente al modelo de datos de información correspondiente al conteo de actas



Fuente: Elaboración propia.

Y finalmente, una visión del modelo físico completo solo como referencia de lo que se integra en un modelo multidimensional de un repositorio unificado.

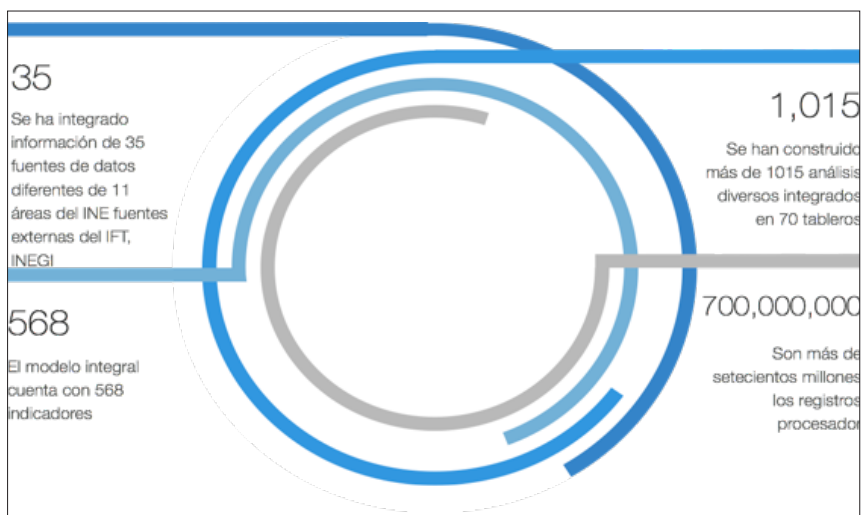
Gráfico 4. Fragmento del modelo físico completo del SII



Fuente: Elaboración propia.

Lo cual, en conjunto, engloba lo siguiente:

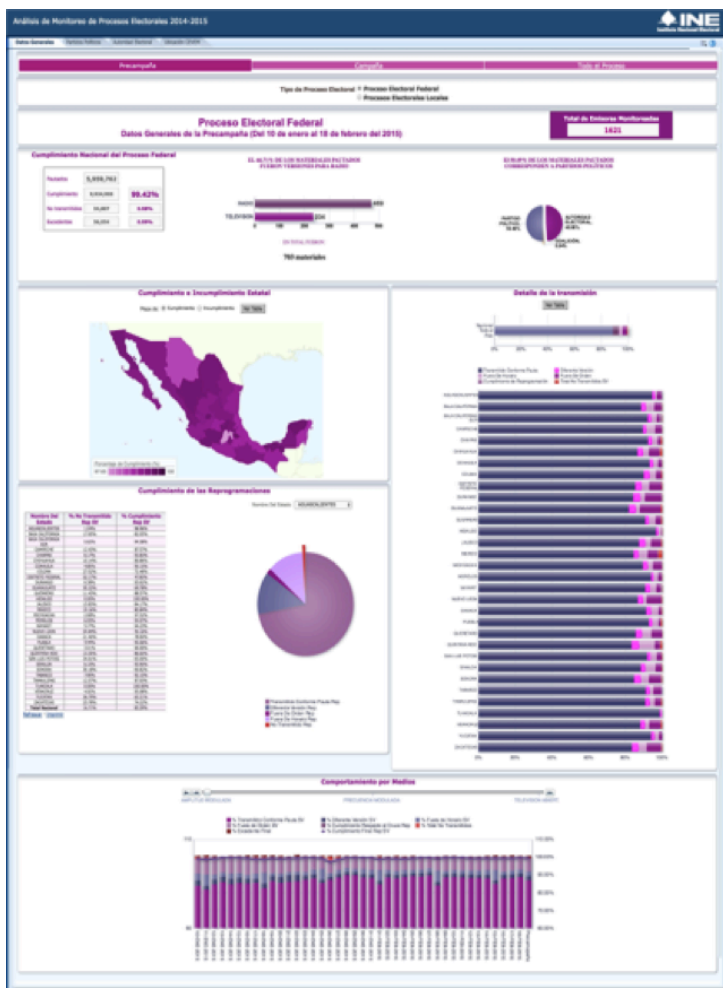
Gráfico 5. Información cargada al repositorio único del SII



Fuente: Elaboración propia.

Ejemplo de un análisis que puede ser generado desde el Sistema de Inteligencia Institucional:

Gráfico 6. Ejemplo de análisis sobre la transmisión de promocionales de radio y televisión para los partidos políticos en México



Fuente: Elaboración propia.

CONCLUSIONES:

El uso de herramientas analíticas cobra mayor relevancia en las organizaciones de gobierno y empresas privadas; estar informados no sólo implica acumular información diariamente, sino tener capacidades de explotar esa información de la mejor manera para la toma de decisiones.

Elegir una herramienta de análisis de información depende del tipo de aplicaciones y usos que se hará de la información; en particular para la inteligencia de negocios un pilar fundamental es contar con un repositorio unificado de información, que más adelante puede complementarse con herramientas de *big data* que permitan el análisis de información no estructurada. Sin embargo, en una organización de gobierno o empresa, mucha de la información que se encuentra en el inventario de los sistemas empleados corresponde a información estructurada, por ello implementar el uso de herramientas analíticas que exploten este tipo de acervo es sustancial para potenciar en análisis al interior de este tipo de organizaciones.

Es importante diferenciar entre una herramienta de inteligencia de negocios y herramientas de *big data*. La distinción se vuelve importante en el tipo de fuentes que buscan analizarse, ya sean estructuradas o no estructuradas, ya que cada una conlleva esfuerzos y objetivos distintos. El valor que hemos descubierto de trabajar en un repositorio unificado es que al integrar las fuentes estructuradas en un modelo multidimensional, se logra la madurez del modelo a largo plazo; es decir, todo el conocimiento que se ocupó para diseñar cada una de las bases de datos independientes y que se emplearon en procesos operativos reales dentro de la organización se han recuperado y fusionado en un modelo que permite hoy asegurar que ante cualquier

cambio futuro en las fuentes el modelo multidimensional será persistente y permitirá acumular conocimiento de diversos años de años previos, presentes y futuros permitiendo la convivencia de múltiples modelos de bases de datos en un modelo ya unificado y funcional, y con plena certeza en que los indicadores están plenamente validados y la información que producen es cierta y de calidad. *Big data* al contrario podría analizar y cruzar fuentes diversas, sean estructuradas y no estructuradas, y proponer hallazgos, identificar patrones que permitan concebir una fotografía mucho más clara de lo que esos datos representan; sin embargo, en una organización donde los procesos de negocio son modelados concebir un gran modelo multidimensional, será sin duda el mayor beneficio al reflejar la operación del negocio con un modelo confiable y bien estructurado.

Noam Chomsky decía en el prefacio de su libro *El conocimiento del lenguaje* que le intrigaban dos problemas referentes al conocimiento humano. El primero era explicar cómo conocemos tanto a partir de una experiencia tan limitada, cómo conocemos tanto teniendo en cuenta que los datos de los que disponemos son tan escasos (problema de Platón), y el segundo es el problema de explicar cómo conocemos tan poco considerando que disponemos de una experiencia tan amplia, cómo conocemos y comprendemos tan poco, a pesar de que disponemos de unos datos tan ricos (problema Orwell). Considero que hoy más que nunca, este segundo problema cobra aún mayor relevancia: el problema no es acumular información. sino lograr discernir dentro de ese gran cúmulo de datos los datos que nos otorguen valor, y es ahí cuando usar herramientas analíticas que nos permitan posicionar por encima de la capa de los datos podrán dotarnos de una inteligencia adelantada, afinada y certeza sobre lo que se debe discernir dentro del conjunto de datos no explorados.

BIBLIOGRAFÍA

- Bernabeu, Ricardo Darío y Mariano García Mattío. 2017. "Hefesto: Data Warehousing: Guía completa de aplicación teórico práctica; metodología Data Warehouse v 3.0". Disponible en https://sourceforge.net/projects/bi-hefesto/?source=typ_redirect.
- Breiman, Leo. 2001. "*Statistical Modeling: The two cultures*".
- Chambers, John M. 1993. "Greater or lesser statistics: a choice for future research. *Statistics and Computing*", vol. 3, núm. 4:182-184.
- Cleveland, William S. 2001. "Data Science: an action plan for expanding the technical areas of the field of statistics", *International statistical review*, vol. 69, núm. 1:21-26.
- Chomsky, Noam. 1985. *El conocimiento del lenguaje*. Madrid: Atalaya.
- Donoho, David. 2017. "50 years of Data Science", *Journal of Computational and Graphical Statistics*. <https://doi.org/10.1080/10618600.2017.1384734>
- Halper, Fern y David Stodder. 2014. "TDWI Analytics Maturity Model". The Data Warehouse Institute. Disponible en <https://tdwi.org/whitepapers/2014/10/tdwi-analytics-maturity-model-guide.aspx>.
- Kelleher y Tierney. 2018. *Data Science*. The MIT Press Essential Knowledge Series: MIT.

La Revolución de los Datos Bibliográficos, Científicos y Culturales. La edición consta de 100 ejemplares. Coordinación editorial, Israel Chávez Reséndiz; revisión especializada, Valeria Guzmán González, revisión de pruebas, Carlos Ceballos Sosa, formación editorial, Oscar Daniel López Marín. Instituto de Investigaciones Bibliotecológicas y de la Información /UNAM. Fue impreso en papel cultural de 90 gr. en los talleres de Grupo Fogra. Año de Juárez 223. Col. Granjas San Antonio. Alcaldía Iztapalapa. Ciudad de México. Se terminó de imprimir en julio de 2020.