

UNIVERSIDAD CENFOTEC
MAESTRÍA EN TECNOLOGÍAS DE BASES DE DATOS
PROYECTO DE INVESTIGACIÓN APLICADA

Diseño de una estrategia base para la elaboración de un sistema de Inteligencia Tributaria para el Ministerio de Hacienda

Ing. José Alberto Leiva Obando, MBA.

Julio, 2013

Dedicatoria

A mi familia por el apoyo y paciencia durante todos estos años y por alentarme a no desistir de alcanzar mis metas y crecer cada día.

Agradecimientos

A todas las personas involucradas desde las entrevistas hasta las revisiones y observaciones durante el proceso para, con su experiencia, enriquecer el proceso de investigación y desarrollo.

Tabla de contenido

Capítulo I	Introducción	1
1.1	Introducción.....	1
1.1.1	Antecedentes.....	1
1.2	Definición y descripción del problema	6
1.3	Justificación.....	8
1.4	Viabilidad.....	11
1.4.1	Punto de vista técnico.....	11
1.4.2	Punto de vista operativo	11
1.4.3	Punto de vista económico	12
1.5	Objetivos	13
1.5.1	Objetivo general.....	13
1.5.2	Objetivos específicos	13
1.6	Marco institucional.....	14
1.6.1	Misión	14
1.6.2	Visión.....	14
1.6.3	Valores institucionales	14
1.7	Alcances y limitaciones	15
1.7.1	Alcances	15
1.7.2	Limitaciones.....	16
Capítulo II	Marco teórico	18
2.1	Estado del arte	19
2.2	Inteligencia tributaria	24

2.3	Bases de datos tabulares	25
2.4	Reportes Ad Hoc	28
Capítulo III	Marco metodológico	30
3.1	Tipo de investigación	31
3.1.1	Población	31
3.1.2	Técnicas	31
3.1.3	Procedimientos	32
3.1.4	Enfoque de la investigación	33
Capítulo IV	Propuesta de solución	34
4.1	Situación actual	35
4.2	Propuesta de solución	35
4.2.1	Repositorio único	36
4.2.2	Niveles de datos	42
4.2.3	Arquitectura del proceso de carga	45
4.3	Prototipo de estrategia	49
4.3.1	Estrategia de recolección de datos de contribuyentes	50
4.3.2	Estandarización de datos entre instituciones	52
4.3.3	Consolidación de datos de las fuentes	55
4.3.4	Utilización de herramientas de análisis	56
4.3.5	Seguimiento de casos	58
4.4	Prototipo de aplicación	59
4.4.1	Base de datos	60
4.4.2	Modelo de datos	60
4.4.3	Diseño de aplicación	70

Capítulo V	Conclusiones y recomendaciones.....	87
5.1	Conclusiones.....	88
5.2	Recomendaciones.....	90
Referencias		92
Apéndices		96
A.1	Entrevista.....	97
A.2	Formulario renta.....	103
A.3	Formulario venta.....	104

Índice de tablas

Tabla 1: Ejemplo tabla conversión identificaciones.....	40
Tabla 2: Diccionario de tablas	62
Tabla 3: Diccionario de columnas.....	62
Tabla 4: Diccionario de operaciones	62
Tabla 5: Personas	63
Tabla 6: Registro de búsqueda	64
Tabla 7: Registro de columnas.....	65
Tabla 8: Registro de procesos.....	66
Tabla 9: Secciones de expedientes.....	66
Tabla 10: Categorías de expedientes.....	67
Tabla 11: Detalles de expediente	67
Tabla 12: Tipos de operación	68
Tabla 13: Tipo de datos.....	68
Tabla 14: Catálogo de operadores	68
Tabla 15: Periodicidades	69
Tabla 16: Filtros de los casos.....	69
Tabla 17: Casos estudio.....	70
Tabla 18 Tabla de filtros de casos.....	72
Tabla 19 Diccionario de Tablas	76
Tabla 20 Diccionario de columnas.....	78
Tabla 21 Resultado de búsqueda.....	84

Índice de figuras

Figura 1: Esquema estrella Tomado de: www.dataprix.com	21
Figura 2: Esquema copo de nieve Tomado de: www.dataprix.com	22
Figura 3: Índices por fila y por columna	27
Figura 4: Esquema conceptual Contribuyente centrista Tomado de: Elaboración propia	40
Figura 5: Estructura conceptual tablas de declaración Tomado de: Elaboración propia	42
Figura 6: Niveles de declaración Tomado de: Elaboración propia	43
Figura 7: Jerarquía de periodicidades Tomado de: Elaboración propia	45
Figura 8: Proceso de carga Tomado de: Elaboración propia	45
Figura 9: Prototipo propuesta de inteligencia tributaria Tomado de: Elaboración propia.....	49
Figura 10: Obtención de datos situación actual Tomado de: Elaboración propia.....	53
Figura 11: Formato estandarizado de datos Tomado de: Elaboración propia.....	54
Figura 12: Datos – Información – Conocimiento Tomado de: Elaboración propia.....	56
Figura 13: Casos de estudio creados Tomado de: Elaboración propia	71
Figura 14: Mantenimiento de casos de estudio Tomado de: Elaboración propia	71
Figura 15: Mantenimiento de filtros de casos Tomado de: Elaboración propia	73
Figura 16: Definición de operaciones	74
Figura 17: Flujo de casos de estudio Tomado de: Elaboración propia.....	75
Figura 18: Tablas creadas Tomado de: Elaboración propia.....	76
Figura 19: Mantenimiento de tablas Tomado de: Elaboración propia	77
Figura 20: Resultado tablas Tomado de: Elaboración propia	77
Figura 21: Columnas creadas Tomado de: Elaboración propia	78
Figura 22: Mantenimiento de columnas Tomado de: Elaboración propia.....	79
Figura 23: Resultado de columnas	80
Figura 24: Flujo de diccionarios Tomado de: Elaboración propia.....	80
Figura 25: Casos de estudio para ejecución Tomado de: Elaboración propia.....	81
Figura 26: Ejecución de operaciones Tomado de: Elaboración propia	82
Figura 27: Búsqueda de personas Tomado de: Elaboración propia	84
Figura 28: Datos del contribuyente Tomado de: Elaboración propia.....	85
Figura 29: Desglose de datos del contribuyente.....	85
Figura 30: Detalle de los datos de la persona Tomado de: Elaboración propia	86

Figura 31: Principal elemento de un sistema de Inteligencia Tomado de: Elaboración propia	97
Figura 32: Principal factor de fracaso Tomado de: Elaboración propia	98
Figura 33: Principal inconveniente en la carga de datos Tomado de: Elaboración propia.....	99
Figura 34: Inconveniente al realizar las cargas Tomado de: Elaboración propia	100
Figura 35: Modelo de desarrollo preferido Tomado de: Elaboración propia.....	101
Figura 36: Herramienta más utilizada por los analistas Tomado de: Elaboración propia	102
Figura 37: Formulario D101 (Renta) Tomado de: http://costarica.eregulations.org	103
Figura 38: Formulario D104 (Ventas) Tomado de: http://dgt.hacienda.go.cr/	104

Resumen

La investigación tiene como objetivo diseñar una estrategia base para la elaboración de un sistema de Inteligencia Tributaria para el Ministerio de Hacienda. Los objetivos específicos de la misma son:

- Definir los modelos de declaración a evaluar como ejemplos para la estrategia de inteligencia.
- Identificar los procesos involucrados en una estrategia de inteligencia tributaria.
- Preparar un prototipo de una aplicación que interactúe con el repositorio unificado para la creación de casos de estudio fiscales.
- Proponer un modelo base para un repositorio unificado de la información fiscal para renta y ventas.
- Proponer un prototipo de la estrategia de inteligencia tributaria para el manejo y carga de las declaraciones de renta y venta.
- Investigar el modelo actual de presentación y almacenamiento de las declaraciones de renta y ventas.

Se analizan la forma en que actualmente son presentadas las declaraciones de impuestos en Costa Rica, como parte de la investigación se establecen los modelos de declaración con los cuales trabajar como referencia para definir un modelo.

Adicionalmente, se identifican los procesos base involucrados en una estrategia de inteligencia tributaria, desde el proceso de presentación hasta los procesos posteriores de seguimiento y control. Esto con el fin de ir delineando los posibles pasos de la estrategia base a proponer.

Con la información recopilada acerca del tipo de herramientas que utilizan los analistas de datos en contextos similares, se define un prototipo de una herramienta para la revisión de expedientes, extracción y filtrado de los datos por parte de los analistas.

Como parte de las tareas se propone un modelo de repositorio unificado, en el cual, dadas las características estructurales de las declaraciones, se utiliza un almacenamiento del formulario en una tabla con todas las casillas. En este almacenamiento se establecen niveles de datos, en los cuales, conforme se va subiendo se agrupan los montos por la granularidad del período, hasta llegar en el nivel superior, a tener un combinado entre varias declaraciones con los campos más relevantes para el analista, de manera que al realizar su labor no requiere ir a combinar varias tablas de consulta.

De acuerdo con los puntos anteriores, se establece un prototipo de una estrategia base, que considere el almacenamiento de los datos internos y externos por medio de las cargas, el modelo del repositorio, y la aplicación de consulta y creación de casos.

Finalmente, de acuerdo con la información pública disponible acerca de la manera en la cual son presentados actualmente las declaraciones de impuestos en Costa Rica. Además, se investigan aspectos como otros proyectos ejecutados y contrataciones abiertas por Hacienda para proyectos de este tipo, a fin de tener una mayor visión de la situación actual en este tema.

Capítulo I. Introducción

1.1 Introducción

1.1.1 Antecedentes

El Ministerio de Hacienda, como ente recaudador, ha sido el encargado, dentro de sus funciones, de la recepción de las declaraciones de impuestos por parte de los contribuyentes. De esta manera, que dadas las distintas características de los impuestos, se generan formularios y reglas particulares para cada tipo de declaración. Estos formularios, son puestos a disposición de las personas físicas y jurídicas para que sean completados con los datos fiscales requeridos para cada período.

En años anteriores los formularios estaban disponibles únicamente en papel, para que fuesen llenados y entregados a los bancos de forma manual. Sin embargo, como en todo proceso manual, los datos registrados en el papel por los contribuyentes o sus contadores no pueden ser validados al momento. Asimismo, una vez entregados a las oficinas o bancos, para verificar la consistencia de los datos se requería de un proceso manual para validar los valores de cada una de las casillas. De este modo, dado que los formularios pueden contener muchas casillas de información y por la cantidad de formularios distintos, era requerido un trabajo adicional, por parte de los funcionarios, para el ingreso de estos valores en los sistemas.

Era necesario considerar problemas comunes en casos como estos, como por ejemplo el tiempo requerido para interpretación de los valores escritos a mano que no son totalmente claros, la detección de inconsistencias, la ilegibilidad de la declaración, el traslado de los documentos a diferentes áreas geográficas, con el peligro de que se dañaran o fuesen extraviados. También, dentro de los inconvenientes de un proceso manual, es que no existen controles totalmente

confiables para evitar alteraciones las declaraciones y mucho menos para rastrear el origen de una alteración.

Dentro de los esfuerzos de tributación fueron introduciéndose herramientas tecnológicas para llenar y validar de forma automática las declaraciones. De esta manera, el Ministerio de Hacienda, pone a disposición de los contribuyentes los sistemas EDDI y Declar@. Asimismo, puso a disposición la plataforma Tribunet, en la cual pueden cargarse los archivos de las declaraciones en línea. En este sentido, Hacienda define en su sitio a Tribunet como “El primer sitio Web que la Dirección General de Tributación puso a disposición de los sujetos pasivos para el cumplimiento de obligaciones tributarias vía Internet.”(Hacienda, 2010)

Tributación ha dado los pasos para ir automatizando este tipo de procesos, entre estos se cuenta la obligatoriedad de la presentación de algunos impuestos de forma digital, en sustitución de los formularios en papel. Por ejemplo, como se indica en la Prensa Libre del 25 de mayo de 2012, las declaraciones de Ventas deben presentarse de forma obligatoria de forma electrónica a partir del 1 de junio. Asimismo, en el mismo artículo el ministro de Hacienda Edgar Áyales indica: “Buscamos facilitar el cumplimiento de las obligaciones tributarias y mejorar el control y la calidad de información registrada en sus bases de datos.” (Prensa Libre, 2012)

En el 2009, en La Gaceta 192 se publica una resolución de Hacienda que indica “Se establece el uso obligatorio de los servicios electrónicos de Tributación Digital para todas las personas físicas o jurídicas, públicas o privadas, en lo que respecta al suministro de información de trascendencia tributaria.” (La Gaceta, 2009)

Desde 2007 la dirección de Tributación menciona proyectos de Data warehouse, como se puede incluir en la presentación de la Dirección General de Informática de ese año. Aún se está en proceso de establecer los mecanismos mediante los cuales se pueda obtener y consolidar la información, tanto interna (como las declaraciones, datos de contribuyentes y otras) como la proveniente de otras instituciones (bancos, aseguradoras, migración y otras). Adicionalmente, en

este momento algunos de los procesos analíticos utilizan datos aislados que se cargan en hojas de cálculo para su análisis y establecer posibles casos sospechosos. No obstante, a nivel tecnológico queda camino por recorrer para el establecimiento de un proceso automatizado que facilite las labores de Inteligencia Tributaria.

Dentro de los proyectos también puede mencionarse la factura electrónica, que pretende dar un mayor control sobre los ingresos de los profesionales independientes en cuanto a sus declaraciones de la renta. Lo anterior, por cuanto estos tienen un potencial alto de incumplimiento tributario, por medio de elusión o evasión de impuestos. En cuanto a esto, se indica que “anualmente las personas físicas tienen un 46% de incumplimiento del potencial tributario; es decir, de lo que podría recaudar Hacienda se pierde un 0,9% del PIB, aproximadamente \$200 mil millones.” (Prensa Libre, 2012)

El viceministro de Hacienda indicó en noviembre de 2012 que: “que realizan labores de inteligencia fiscal para cruzar información e implementan el proyecto de factura electrónica para un mayor control. “ (Prensa Libre, 2012)

En contraposición a estos proyectos, otros han pasado por problemas que incluso han causado la suspensión de los mismos luego de la inversión realizada. Uno de estos casos, es el proyecto de Tributación Digital, el cual fue pospuesto en el 2010, al respecto, La Nación indicaba que “el sistema le haría mucho más eficiente en el proceso de análisis de la información, lo cual generaría una mejor fiscalización.” (La Nación, 2010). Este proyecto, buscaba que los contribuyentes hicieran todos sus trámites en línea. El proyecto tenía un costo de \$25 millones. Como menciona en el mismo artículo el especialista en impuestos Eric Thompson “la ampliación de cruces de información desestimula el incumplimiento.”

A partir de 2005 el gobierno de Costa Rica ha realizado esfuerzos para la automatización y el manejo electrónico de los datos de los contribuyentes y sus declaraciones. Lo anterior con el fin de tener datos consolidados con los que

posteriormente sea posible realizar estudios y obtener estadísticas de pago o identificar posibles casos de auditoría fiscal.

Dentro de los esfuerzos realizados se presentó el proyecto Tributación Digital, un proyecto de \$26 millones. Como indicó a El Financiero, Rowland Espinoza, ex Viceministro de Hacienda “El proyecto se componía de 14 módulos, con una inversión total por parte del Estado de \$26 millones” (El Financiero, 2011). No obstante, la empresa encargada del mismo se declaró en quiebra y el proyecto fracasó, con un costo para el estado de \$18 millones de dólares. De este modo, como informa El Financiero (2011):

Una inversión estatal cercana a los \$18 millones para desarrollar el programa Tributación Digital está en arenas movedizas, mientras los tribunales de justicia resuelven el litigio entre el Ministerio de Hacienda y la empresa BearingPoint, contratada en el 2007 para desarrollar la plataforma tributaria.(Párrafo 1)

A partir de esta experiencia, con mayores reservas en cuanto a realizar inversiones tan altas, Hacienda ha impulsado proyectos internos con el apoyo de consultores externos. De esta manera, dentro de estos proyectos de automatización y digitalización de la información se puede mencionar a Po@s y Tributación Direct@, los cuales han sido planteados con el objetivo de consolidar datos y facilitar los procesos de pago a los contribuyentes.

Lo anterior, aunado a nuevos lineamientos legales para paulatinamente establecer la obligatoriedad de la presentación electrónica de las declaraciones va delineando el camino por el cual Tributación está enfocando sus esfuerzos. De esta forma, se facilita el proceso de consolidación de datos de las declaraciones, para análisis en Hacienda, y por otro lado, para el pago de los contribuyentes, reduciendo la manipulación manual de datos y eliminando el elemento humano en varios de los procesos.

A partir de tener mayor volumen para consolidación de los datos de las distintas fuentes, Hacienda pretende mejorar su potencial de análisis. De este modo, se han desarrollado sistemas para mejorar el manejo electrónico de las declaraciones. Como se indica en El Financiero (2011):

Es así como sus informáticos lograron desarrollar plataformas como la Po@s y Tributación Direct@. La primera pretende administrar una base de datos gigante que permita realizar controles cruzados sobre los tributos que pagan personas físicas y jurídicas.

El segundo proyecto sería el sustituto directo de Tributación Digital y tiene como labor fundamental desarrollar la declaración y recaudación de impuestos con base en medios electrónicos. (Sección Plataformas, párrafo 2)

Esto por cuanto la evasión fiscal es una de las causas del déficit fiscal. Adicionalmente, la presentación en papel en ocasiones ha provocado que algunas declaraciones de los contribuyentes contengan datos alterados o incongruentes, de forma que es posible observar casos en que profesionales pagan porcentajes muy bajos de sus ingresos, comparados con otros profesionales en su misma actividad. Por tanto, estos casos pueden ser candidatos a ser auditados pero actualmente no se tienen las estrategias para mejorar este proceso.

1.2 Definición y descripción del problema

En el contexto actual en el cual se discute acerca del déficit fiscal, el Banco Central de Costa Rica estima que para 2013 llegue aproximadamente al “5% del producto interno bruto” (CRHoy, 2012). Asimismo, dentro de las posibles razones del déficit fiscal actual se señalan la evasión y la elusión (ver adelante) de las personas físicas y jurídicas para realizar un pago menor de impuestos.

En los últimos años se ha buscado por parte de las entidades de gobierno realizar los controles respectivos para obtener una recaudación mayor de impuestos a partir del justo pago por parte de los profesionales y las empresas. Dado lo anterior se hace cada vez más necesario la implementación de sistemas automatizados para la detección de posibles casos de estudio en las auditorías fiscales, de forma que se promueva una mayor transparencia con el fisco por parte de los contribuyentes. A pesar de lo anterior, muchos de los procesos aún siguen siendo seleccionados al azar, por cuanto no se ha avanzado a un proceso de unificación de los datos y detección automatizada de posibles casos de estudio.

La elusión consiste en buscar, por medio de caminos provistos por la legislación o vacíos en la misma, reducir la cantidad de impuestos a pagar. Por otra parte, la evasión consiste en el no pago de los impuestos que corresponden. Por lo tanto, para el proyecto de investigación el enfoque se dará en la evasión.

El problema a investigar es determinar una posible estrategia base de inteligencia tributaria a partir del cual, extendiéndose en posteriores etapas, sea posible mejorar las labores de análisis y seguimiento por parte de los encargados. Para lo anterior, se requiere verificar los datos de los contribuyentes y de sus declaraciones para establecer los casos candidatos a revisión. No obstante, este proceso debe abarcar la consolidación de los datos, el análisis de los mismos y mecanismos de seguimiento de los casos auditados. Sin embargo, ese es un boceto

de todas las labores que serían requeridas incrementalmente para tener una estrategia tributaria integral para el país.

Para efectos de la investigación se plantea utilizar un conjunto reducido y simplificado de entidades y reglas que permitan representar de forma conceptual las ideas planteadas en el proceso de la investigación. A pesar de la gran cantidad de posibilidades de declaraciones existentes, el proyecto busca enfocarse en la determinación de un modelo base para el establecimiento de un repositorio unificado con fines analíticos en caso de realizar procesos de Inteligencia Tributaria.

Adicionalmente, dada la confidencialidad de los datos, estos no serán incluidos dentro del texto. De esta forma, las observaciones, análisis y revisiones fueron realizados con información disponible al público, como por ejemplo, la estructura de las declaraciones.

1.3 Justificación

De acuerdo al estudio sobre evasión fiscal del Ministerio de Hacienda, se informa que para el año 2010 el monto de la evasión “pretende reducir las erogaciones del Estado en un 1% del producto interno bruto (PIB) y aumentar los ingresos en un 2%”(CRHoy, 2012). Adicionalmente, el viceministro de Hacienda Álvaro Ramos, señala que “Una parte del incumplimiento responde a la figura de elusión, es decir, formas a las que el contribuyente recurre para evadir el pago de impuestos aprovechando ambigüedades y vacíos en las leyes” (CRHoy, 2012)

Se han planteado diversos escenarios para conseguir la disminución del déficit, entre los cuales se encuentra la definición de nuevos mecanismos de recaudación y nuevos impuestos. De acuerdo a lo anterior, se plantea el desarrollo de un nuevo plan fiscal, como lo informa La Nación (2013) “...El Gobierno trabaja en una nueva propuesta fiscal que llevará a la Asamblea Legislativa entre mayo y junio de este año, y que incluirá medidas para mejorar los ingresos, el gasto y el financiamiento de la deuda...” Asimismo han surgido iniciativas para tener mayor automatización en el manejo de los datos fiscales y en el control de los contribuyentes. Por ejemplo, se tiene que en 2011 se licita un sistema de inteligencia tributaria para el Ministerio de Hacienda, como lo reseña CentralAmericaData (2011) “El Ministerio de Hacienda licita la contratación de consultoría en servicios informáticos para el desarrollo y mantenimiento de un sistema de inteligencia tributaria.”.

Se busca la digitalización de declaración y la presentación y pago por medios electrónicos. De esta forma, en los últimos años se han hecho esfuerzos con el fin de que los contribuyentes puedan presentar sus declaraciones por Internet. Como informa La Nación (2012) al respecto:

“Las declaraciones de retenciones del impuesto sobre la renta serán las terceras que se podrán presentar sólo por Internet.

Primero fueron, en diciembre del 2011, las declaraciones del impuesto sobre la renta y luego, desde junio pasado, las declaraciones del impuesto sobre las ventas.”

El gobierno ha invertido en proyectos millonarios y ha obtenido pocas ganancias en la lucha por mejorar la recaudación. Por lo tanto, se ha buscado asesoría con entidades extranjeras, para determinar modelos a seguir para los nuevos sistemas. En este sentido, Rowland Espinoza indica que se tienen convenios con la Agencia Estatal de Administración Tributaria, y señala que “Se escogió a España para firmar el convenio, pues Hacienda considera que ese país goza de buenos resultados en el nivel de recaudación fiscal.”(El Financiero, 2011).

A nivel mundial los últimos años han sido de crisis financieras en distintas economías. En casos como el de la crisis griega, a partir de gran cantidad de beneficios a los burócratas y un déficit fiscal elevado se ha presentado una crisis que en este momento tienen sometida a la economía griega en una situación de amplias restricciones impuestas por otros países europeos que adicionalmente sirven de apoyo financiero para evitar que la crisis se contagie al resto de la comunidad europea.

Una vez visto el ejemplo de Grecia y a la luz del acelerado crecimiento del déficit fiscal en Costa Rica, se hace necesario plantearse de forma responsable los ajustes que van a ser necesarios para que la economía no se derrumbe y entre en una crisis que pueda llegar a tener un alto precio social. A diferencia de Grecia, Costa Rica no tiene el respaldo de economías fuertes, por lo cual, aun cuando cuantitativamente las dos economías no son comparables, los síntomas pueden ser comparados y es oportuno determinar a tiempo las medidas preventivas para poder enrumbar la economía.

En la sociedad actual, muchos datos de las transacciones cotidianas se almacenan de forma electrónica. No obstante, en gran parte de las ocasiones, estos datos están en diferentes formatos y distribuidos en distintos repositorios. Lo anterior,

dificulta el análisis de la información, dado que al no tener una centralización y estandarización de la información fiscal, se complica realizar estudios más fiables de la situación actual del país en este aspecto.

Se busca, de acuerdo a la información que ya se captura de los contribuyentes y sus declaraciones, definir una estrategia base que permita tener un modelo unificado para realizar labores de Inteligencia Tributaria.

En resumen, el proyecto pretende, a partir de una herramienta tecnológica, crear una estrategia base para un modelo de Inteligencia Tributaria, que pueda ser extendido para distintas declaraciones e información fiscal. De esta manera se podrá tener un origen unificado para el estudio y detección de blancos fiscales, es decir, de contribuyentes que de acuerdo a los datos de sus declaraciones y los personales, puedan ser sujetos de una auditoría fiscal. Finalmente, para esto, es necesario integrar los datos a sistemas unificados y automatizados que permitan realizar estudios aún más sofisticados y exhaustivos. Finalmente, el proyecto se enfoca en un ámbito reducido pero que puede servir de referencia base para que estudios futuros implementen nuevos métodos. Lo anterior, con el fin de aportar al país soluciones que permitan aumentar la recaudación con el objetivo de ir reduciendo el déficit fiscal, para la mejora continua de los indicadores económicos y sociales, si se hace el uso adecuado de los mismos.

1.4 Viabilidad

1.4.1 Punto de vista técnico

El Ministerio de Hacienda licitó un sistema de inteligencia por 9600 horas anuales, lo que corresponde a un grupo de 10 analistas externos interactuando con los internos que conocen las estructuras y procesos relacionados con los contribuyentes y las declaraciones de Renta y Ventas, así como las experiencias anteriores. Este grupo, puede servir como base para la discusión y la generación de los elementos necesarios para el establecimiento de la estrategia requerida.

Adicionalmente, están a disposición del equipo los recursos tecnológicos para la elaboración del proyecto, así como acceso a los analistas y las estructuras y datos disponibles en el proyecto de Inteligencia Tributaria.

En cuanto a Hardware y Software, Hacienda realiza esfuerzos por fortalecerse tecnológicamente. Un ejemplo de esto lo reseña Oscar Rodríguez para el caso de recaudación de impuestos en que indica que “Tributación hizo un análisis de su plataforma y determinó la necesidad de robustecer su sistema de trámite de los formularios digitales en línea” (Rodríguez, 2013).

1.4.2 Punto de vista operativo

Para el proyecto, estaría a disposición de Hacienda recursos externos e internos, de acuerdo a las contrataciones publicadas y adjudicadas, con el cual es posible, si es requerido, realizar los análisis e implementaciones.

De este modo, siempre dentro del contexto de los acuerdos de confidencialidad, se posee acceso a datos y estructuras para la investigación de forma controlada y sin poder extraer los mismos desde los servidores. Adicionalmente, se utiliza el conocimiento obtenido en el trabajo de campo, por parte

de los analistas, en cuanto a los elementos involucrados en los procesos tributarios de captura de declaraciones y análisis de datos. Por lo tanto, para los estudios se dispone del tiempo y recursos para realizar laboratorios con estructuras semejantes a las reales y establecer las propuestas.

Asimismo, se tiene acceso a licencias de Software y de infraestructura para realizar pruebas y laboratorios de las propuestas.

Por lo tanto, dada la naturaleza del proyecto propuesto, el patrocinador provee los recursos humanos y tecnológicos para su realización. De esta manera, una vez finalizado el proyecto, el patrocinador podrá evaluar posteriormente la propuesta para presentarla como una opción de desarrollo en la Administración Tributaria.

1.4.3 Punto de vista económico

Como se indica en CentralAmericaData (2011), el Ministerio de Hacienda licitó y presupuestó un contrato para atención de proyectos de Inteligencia Tributaria. En la licitación se indican 9600 horas anuales durante tres años. Dentro de este contexto, con el presupuesto aprobado para tal fin, de acuerdo a las prioridades del Ministerio es factible destinar los recursos necesarios para la evaluación y aprobación de modelos para tener una estrategia de Inteligencia Tributaria adecuada.

1.5 Objetivos

La definición de los objetivos de la investigación se ha planteado de acuerdo a la taxonomía original de Bloom. De esta forma, el proceso se ha realizado desde los diferentes niveles, representados en los objetivos específicos. Es así, como conforme avanza el proyecto se van configurando los elementos necesarios para alcanzar el objetivo general de la investigación.

1.5.1 Objetivo general

Diseñar una estrategia base para la elaboración de un sistema de Inteligencia Tributaria para el Ministerio de Hacienda.

1.5.2 Objetivos específicos

- Definir los modelos de declaración a evaluar como ejemplos para la estrategia de inteligencia.
- Identificar los procesos involucrados en una estrategia de inteligencia tributaria.
- Preparar un prototipo de una aplicación que interactúe con el repositorio unificado para la creación de casos de estudio fiscales.
- Proponer un modelo base para un repositorio unificado de la información fiscal para renta y ventas.
- Proponer un prototipo de la estrategia de inteligencia tributaria para el manejo y carga de las declaraciones de renta y venta.
- Investigar el modelo actual de presentación y almacenamiento de las declaraciones de renta y ventas.

1.6 Marco institucional

En el contexto de los planes estratégicos de Hacienda, se ha buscado mejorar los procesos de recaudación presentando opciones para facilitar los procesos al contribuyente y dando los pasos tendientes a mejorar los procesos de control de forma que se consiga una recaudación adecuada y justa, de acuerdo a las necesidades del país.

1.6.1 Misión

“Garantizar a la sociedad costarricense la efectiva y justa recaudación de los impuestos, uso adecuado del financiamiento del Sector Público y coadyuvar a la asignación de los recursos, con eficiencia, eficacia y transparencia, mediante una política hacendaria sostenible y la activa rectoría del Sector Financiero, para contribuir al desarrollo económico y social del país.” (Ministerio de Hacienda, 2010)

1.6.2 Visión

“Ser un actor estratégico en el desarrollo socio-económico, con una política hacendaria que impulse el desarrollo económico y social sostenido y la competitividad nacional, en un marco de estabilidad macroeconómica, brindando servicios de calidad, con tecnologías de información y un desempeño transparente y eficiente.” (Ministerio de Hacienda, 2010)

1.6.3 Valores institucionales

“Excelencia, compromiso, honestidad, integridad, trabajo en equipo.” (Ministerio de Hacienda, 2010)

1.7 Alcances y limitaciones

1.7.1 Alcances

En el contexto tributario existen distintos tipos de declaraciones que van en función de las actividades económicas o las características de cada contribuyente sea físico o jurídico. Por tanto, para efectos del proyecto se trabajará con un solo tipo de declaración.

En el proyecto se pretende determinar una estrategia básica para consolidar de forma unificada la información fiscal, tomando como ejemplo las declaraciones de renta y venta. Adicionalmente, generar un modelo por niveles para la consulta de analistas y los especialistas fiscales, aprovechando características como la representación tabular de los datos, semejante a la estructura del formulario, y extendiéndose a posibles usos de índices columnares, como alternativa a los modelos dimensionales tradicionales. De este modo, se pretende, a partir de las declaraciones de la renta, diseñar las bases que permitan en el futuro extender a otras declaraciones, de acuerdo al diseño por niveles que se establezca con declaraciones de renta. De este modo el proyecto se enfoca únicamente en declaraciones de renta para personas físicas, excluyendo otros tipos de personas.

La estrategia no busca definir una tecnología específica para su implementación. Por el contrario, se pretende generar modelos que puedan adaptarse a distintos motores de forma semejante y sin requerir licenciamientos específicos.

1.7.2 Limitaciones

Dentro de las limitaciones del proyecto se encuentra la confidencialidad de los datos fiscales. Por lo tanto, los datos fuente del estudio serán simulados. Dado lo anterior, el proyecto considera los resultados obtenidos a partir de los datos pero los datos de declaraciones y contribuyentes no serán divulgados ni incluidos de forma total o parcial en la documentación del proyecto dado que, por su nivel de confidencialidad y la legislación costarricense, estos datos no son públicos.

El proyecto utilizará datos e información pública disponible en la red, como las estructuras de los formularios de renta (Apéndice A.2) y ventas (Apéndice A.3), artículos de prensa y otros que puedan ser utilizados para definir los componentes requeridos para una estrategia de inteligencia.

Asimismo, como parte del proyecto no se incluye la realización de pruebas de rendimiento o comparativas entre distintos modelos. Lo anterior, por cuanto para este tipo de pruebas se requiere de mayor volumen de datos que los disponibles en este momento para ser analizados.

Es por tanto que se pueden resumir las restricciones de la siguiente forma:

- Contratos de confidencialidad, datos confidenciales y legislación. Por lo tanto no se puede disponer de datos reales.
- Herramientas disponibles para el desarrollo de los experimentos.
- Utilización de datos simulados relativos únicamente a personas físicas y declaraciones de renta.
- No divulgación de la información o datos de estudio.
- Poca o ninguna documentación acerca de procesos de monitoreo automatizado para estas labores.
- Una muestra de datos simulada de un período de tiempo determinado que permita evaluar los resultados.

- Documentación acerca de procesos de Inteligencia Tributaria en otras regiones.
- No se pretende realizar pruebas de rendimiento ni de algoritmos, dado que la cantidad de datos disponibles no es suficientemente significativa para tal fin.

Sin embargo, a pesar de las limitaciones listadas, para efectos del proyecto de investigación es posible continuar en la elaboración de la estrategia, dado que las mismas no son un impedimento para el cumplimiento de los objetivos planteados.

Capítulo II. Marco teórico

2.1 Estado del arte

En el 2006 el Centro Interamericano de Administraciones Tributarias (CIAT), a partir de un grupo de trabajo formado por expertos de diferentes países de América y Europa, generó un manual de referencia para lo relativo a la inteligencia fiscal. Respecto a esto CIAT indica:

Con el propósito de proporcionar solidez y eficacia a los trabajos de inteligencia fiscal, este manual identifica áreas de especial interés tanto para la implementación de esas actividades, como de su perfeccionamiento. Cada una de ellas fue examinada y detallada, de forma que puedan servir como referencia a las administraciones tributarias de los países miembros del CIAT. (CIAT, 2006)

De este modo, el manual generado por este grupo de trabajo, además de ser un grupo de recomendaciones, es una recopilación de las prácticas realizadas por los países participantes dentro de sus actividades de Inteligencia Tributaria, por lo cual, no se trata de un manual meramente teórico sino de experiencia y prácticas.

Asimismo, a nivel de Latinoamérica, incluyendo a Costa Rica, se han realizado estudios y esfuerzos en cuanto a sistemas de inteligencia tributaria. De esta manera, en setiembre del 2009, el Ministerio de Hacienda presentó el documento titulado “Plan Nacional de prevención y lucha contra el fraude fiscal”. De este modo, el Ministerio de Hacienda menciona la importancia de las tecnologías de información como “la gestión de las organizaciones, mediante el manejo de grandes volúmenes de datos necesarios para la toma de decisiones” (Ministerio de Hacienda, 2009)

De la misma forma, el Sistema Integrado del Servicio Nacional Integrado de Administración Aduanera y Tributaria de la República Bolivariana de Venezuela, ha coincidido, al igual que en Costa Rica, y en el CIAT, en que una de las dificultades

existentes en los procesos de Inteligencia Tributaria es la unificación y estandarización de datos de las diferentes fuentes.

De esta forma, las diversas fuentes coinciden en la necesidad de consolidar información de distintas fuentes de información para completar lo relativo al expediente tributario de un contribuyente. De este modo, los datos pueden ser utilizados como un medio para la obtención de información que permita detectar posibles fraudes fiscales. Por lo tanto, uno de los puntos medulares es la consolidación de esta información.

Es así, como por ejemplo el Ministerio de Hacienda, en su plan señala como uno de los problemas, respecto a la interacción con las diversas instituciones involucradas, "... plataformas informáticas con una capacidad limitada de almacenamiento y procesamiento, equipos obsoletos e incompatibles, dificultades en la estandarización de los datos y organización muy disímil, así como falta de herramientas tecnológicas para transferir y acceder los datos" (Ministerio de Hacienda, 2009)

Lo anterior por cuanto, de acuerdo al estudio de Hacienda, en algunas instituciones los datos no están estructurados o estandarizados en cuanto a formatos. Asimismo, indica que "...muchas veces no se cuenta con una estrategia definida y en otros casos, aún cuando se afirma que existe una estrategia formal, no se elaboran y ejecutan los procedimientos..." (Ministerio de Hacienda, 2009).

Además de la integración de esta diversidad de fuentes, uno de los puntos que tienen mayor relevancia corresponde al repositorio en el que se van a almacenar. De esta manera, los diversos organismos mencionan la necesidad de implementar Data warehouse para facilitar las tareas de Inteligencia. Es así como, en el grupo de trabajo del CIAT, Chile menciona la necesidad de un Data warehouse "...que permitirá realizar Inteligencia de negocios con toda la información que se encuentra disponible en las bases de datos..." (CIAT, 2006).

De este modo, se tiene de forma centralizada información de diversas fuentes, tanto instituciones como contribuyentes, que puede ser utilizada para realizar cruces, detectar diferencias y determinar posibles blancos de investigación para la autoridad tributaria.

Ahora bien, los trabajos publicados acerca de inteligencia tributaria coinciden en la importancia de dos puntos de partida:

- Obtención e integración de los datos de diversas fuentes.
- Creación de un Data warehouse para ser utilizado para los análisis.

En cuanto al tema de implementación, dentro de los diseños más utilizados están los de estrella (como puede verse en la figura 1) o copo de nieve (snowflake, como se muestra en la figura 2). Como señala Joseph Curto, “...En un Data warehouse relacional encontramos esquemas en estrella y en copo de nieve que premian por encima de todo las consultas...” (Curto, 2007). De este modo, el esquema estrella representa de forma desnormalizada dimensiones y hechos, a partir de los cuales pueden realizarse los análisis. De igual forma, el esquema copo de nieve, tiene una representación un poco más normalizada.

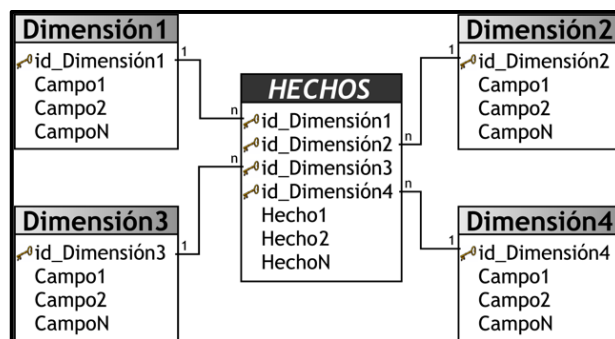


Figura 1: Esquema estrella
Tomado de: www.dataprix.com

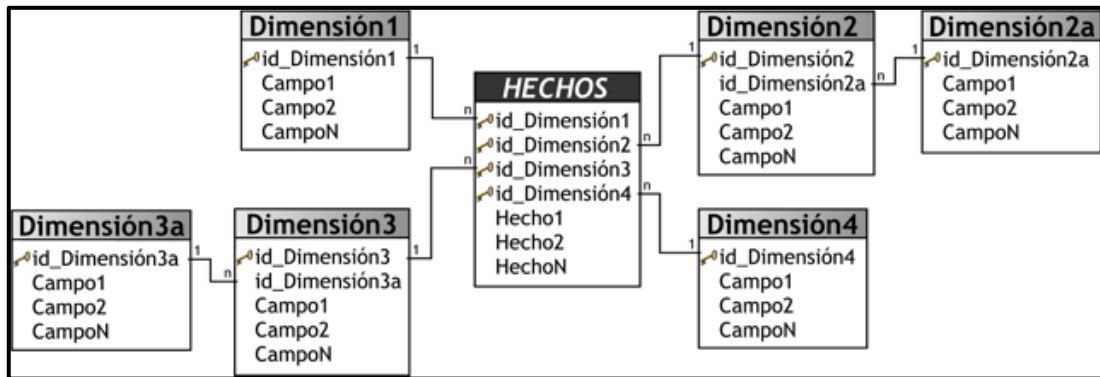


Figura 2: Esquema copo de nieve
Tomado de: www.dataprix.com

Estos esquemas, a diferencia de las tablas de un sistema transaccional (las cuales se encuentran normalizadas), tienen como objetivo tener bajos tiempos de respuesta ante las consultas de los usuarios. De este modo, el almacenamiento, a diferencia de los sistemas relacionales, no es un factor crítico.

Los Data warehouse permiten la generación de Cubos, que pueden ser utilizados por los usuarios expertos en diversas herramientas para cruzar información y generar los reportes. Estos son los esquemas usados por lo general al modelar sistemas para el análisis y toma de decisiones. De esta forma, los datos obtenidos de la fuente se distribuyen en las tablas de dimensiones y de hechos para generar el modelo.

Con el desarrollo de las nuevas tecnologías han surgido nuevas opciones que pueden ser consideradas para el desarrollo de esquemas analíticos para Inteligencia, y en este caso, para Inteligencia Tributaria. Es así como, algunas bases de datos, permiten la generación de índices columnares, como por ejemplo es posible en Microsoft SQL Server 2012 (Olmedo, 2012), y dadas las características de una declaración de renta o ventas, permiten almacenar en una sola estructura los datos requeridos, sin necesidad de establecer gran cantidad de tablas para cruzar, ni conocer cómo se han dividido las tablas en el repositorio unificado.

Parte del estudio a realizar, consiste en mostrar la integración de los datos de declaraciones de Renta y Ventas, no en los esquemas tradicionales de estrella y

copo de nieve, sino en forma tabular, aprovechando elementos como los índices columnares.

En el mismo sentido, utilizando un sistema de niveles, se busca representar como los datos asociados a las declaraciones pueden ser agrupados por distintos criterios, la mayor parte de ellos de periodicidad. De este modo, por medio de diccionarios de datos y herramientas personalizadas, los usuarios analistas tendrán a disposición una herramienta que permita consultar la información de las declaraciones de forma precalculada, y sin requerir grandes esfuerzos para realizar cruces de información, dado que las declaraciones se modelan de acuerdo a las estructuras de las mismas en los formularios que presentan los contribuyentes. Lo anterior permite una mayor analogía entre lo real y lo estudiado, con el fin de que el usuario pueda comprender de forma más natural los datos que está analizando y que no requiera tanto apoyo técnico para realizar los cruces que realiza actualmente de forma manual.

Por lo tanto, el proceso de la investigación se enfoca en obtener y generar información a partir de una secuencia definida de pasos descritos a continuación:

- Trabajo de campo analizando la situación actual en el sitio y con los analistas.
- Revisión del trabajo realizados por otros equipos de inteligencia y recomendaciones de entes especializados, por ejemplo la CIAT.
- Establecimiento de la base para la estrategia que se propone en el proyecto.
- Ejemplificar mediante un prototipo el funcionamiento de la parte tecnológica involucrada en el proyecto.
- Generación de conclusiones y recomendaciones generales para el establecimiento de una estrategia de inteligencia tributaria.

2.2 Inteligencia tributaria

La inteligencia tributaria o fiscal, es un proceso de apoyo para las entidades encargadas de la fiscalización y monitoreo de la política tributaria. Mediante diferentes técnicas y herramientas, la inteligencia tributaria pretende facilitar la detección de posibles blancos de auditoría sospechosos de evadir o eludir sus obligaciones tributarias.

El Centro Interamericano de Administraciones Tributarias (CIAT) en su manual sobre inteligencia fiscal la define como: *“la actividad de recopilar información con la finalidad de asesorar a la administración tributaria en la planificación y ejecución de sus atribuciones legales, particularmente en el combate a la elusión fiscal, a la evasión fiscal y al contrabando”* (CIAT, 2006)

La inteligencia tributaria busca, por medio de cruces, estadísticas, patrones de comportamiento y elementos semejantes, determinar comportamientos anómalos ante posible realización de actividades ilícitas, como el blanqueo de capitales. Pueden hacerse análisis históricos de los datos fiscales de las personas físicas y jurídicas. De esta manera, como lo define CIAT, la inteligencia fiscal puede tener características reactivas o proactivas.

Es así, como de acuerdo a CIAT, la inteligencia tributaria es proactiva cuando *“relacionada con la actividad de obtención y análisis de las informaciones, haciendo posible la elaboración de pronósticos en el área tributaria”* (CIAT, 2006) y es reactiva por cuanto *“puede asumir la forma de investigación, ayudando de forma decisiva a la obtención de pruebas”* (CIAT, 2006)

Para realizar esta labor, las entidades de inteligencia tributaria requieren tener a disposición los datos de diferentes entidades como bancos, aseguradoras, servicios y otros. A partir de esto, los datos pueden ser procesados y analizados desde las distintas perspectivas que permitan los análisis reactivos y proactivos para la detección de blancos.

Como lo señala Carlos Camacho la importancia de la inteligencia tributaria consiste en “detectar personas, ya sean físicas o jurídicas, que puedan tener signos aparentes de riqueza y no paguen sus tributos” (Camacho, 2011). Respecto a este tema, el mismo autor indica:

Esos signos se detectan por medio de sus consumos, información de sus proveedores, bienes inscritos a su nombre y otros mecanismos crecientes de información masiva, al tiempo que no aparezcan de forma coherente cumpliendo con sus deberes, en primera instancia formales y más aún los deberes materiales, es decir, el pago de impuestos. (Párrafo 8)

2.3 Bases de datos tabulares

Dentro de los motores de bases de datos relacionales transaccionales se ha utilizado como regla de diseño la normalización de las tablas, al menos en tercera forma normal, lo cual permite reducir la redundancia de datos y el tamaño de la base de datos.

En contraposición, los modelos multidimensionales, por lo usual, utilizan esquemas de estrella o copo de nieve, los cuales representan las estructuras de forma no normalizada para efectos analíticos, dado que en este contexto tiene preferencia el tiempo de respuesta que el tamaño de la base de datos.

Las bases de datos tabulares presentan la posibilidad de crear un modelo muy parecido a lo que utilizan los analistas con sus herramientas, por ejemplo las hojas de cálculo. Como lo indica Microsoft (Vitt & Cameron, 2012), este modelo da familiaridad a los usuarios dado que es más parecido a las herramientas utilizadas en análisis (en contraposición a un modelo dimensional). Estos modelos dan mayor flexibilidad dado que las tablas no se trabajan en los términos de las dimensiones y los hechos. Con esto, se facilitan las labores de diseño e implementación del componente, dado que no tendrían la complejidad de agregar los datos de acuerdo a las estructuras dimensionales.

Dentro del concepto de bases de datos columnares se han agregado características que permitan mejor rendimiento en ejecución. Por lo tanto, para las bases de datos columnares se pueden utilizar no sólo los índices tradicionales sobre fila, sino que adicionalmente pueden ser incluidos índices de tipo columnar, como son descritos a continuación.

2.3.1.1 Índices columnares

A nivel de bases de datos, la definición y creación de índices es uno de los procesos importantes dado que estos son los que permiten obtener un mejor tiempo de respuesta en las consultas a las tablas de la base de datos. Adicionalmente, la creación de índices tiene implicaciones en cuanto al tamaño que van a ocupar y al ingresar datos en las tablas, lo cual debe considerarse para mantener buenos tiempos de respuesta.

Usualmente los índices son realizados para columnas de una fila y otorgando características de agrupamiento, unicidad u otras, de acuerdo a la función de las columnas involucradas dentro de la fila. No obstante, existe un tipo de índices que trabaja sobre columnas, y el cual puede adecuarse de mejor manera a las características de las bases de datos tabulares, este tipo se llama índice tabular.

Los índices columnares son una herramienta dentro del contexto OLAP, como lo indica se indica en (Olmedo, 2012):

El uso de índices columnares está orientado a los Data warehouse, porque son bases de datos orientadas al análisis, con muchas operaciones de lectura, carga de datos incrementales y manejan histórico de datos con grandes cantidades de datos. (Párrafo 13)

Adicionalmente, Olmedo agrega que, el índice columnar es “un tipo de índice del motor relacional con la particularidad que se almacenan en columnas y nos aportan una mayor capacidad para resolver consultas complejas de una manera

mucho más rápida” (Olmedo, 2012). De esta forma, en un contexto tabular, los índices dan la posibilidad de consultar grandes volúmenes de datos de forma más rápida. En la Figura 3 se muestra la distribución de los índices por filas y columnas en el almacenamiento de la base de datos.

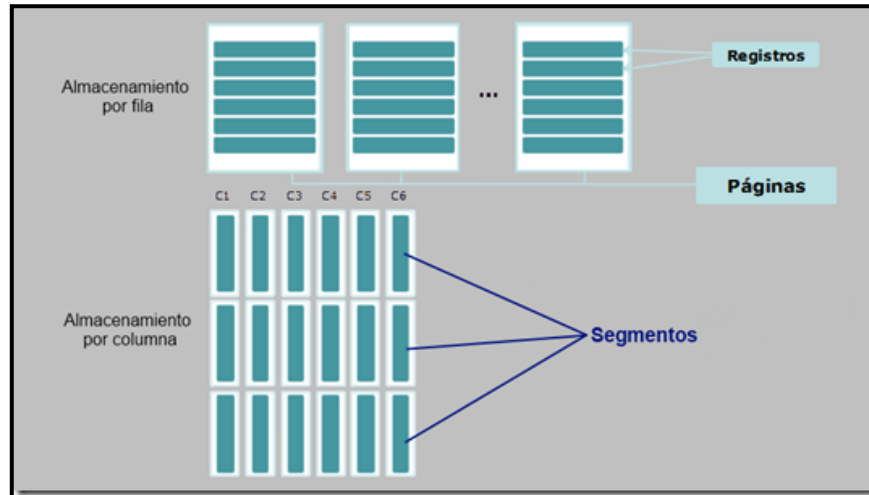


Figura 3: Índices por fila y por columna

Fuente: 1 Olmedo, 2012

Los índices columnares tiene la característica de inhabilitar a la tabla para que le puedan ser ingresados registros. Por lo tanto, cuando se utilicen, al momento de realizar las cargas será necesario deshabilitar temporalmente el índice para luego ponerlo nuevamente a disposición.

2.4 Reportes Ad Hoc

Dentro de los análisis de datos relacionados a la inteligencia de negocios, los reportes constituyen un punto de encuentro entre los datos almacenados y los usuarios. Es así, como a partir de estos reportes los usuarios pueden tener una visión de lo que está ocurriendo con los datos sin necesidad de tener conocimiento técnico de cómo está construido internamente el repositorio.

Asimismo, resulta común que los reportes predefinidos no necesariamente abarquen todas las posibles combinaciones y consultas que los usuarios requieran para su trabajo. Wayne Eckerson (Eckerson, 2012) clasifica a los usuarios en dos grupos:

- Usuarios casuales: son usuarios pasivos y que utilizan la información para su trabajo y que el 80% del tiempo utilizan reportes predefinidos.
- Usuarios expertos: Para estos la información es su trabajo. Son los que pasan generando información y acceden a la información mediante herramientas avanzadas generando sus propios reportes de acuerdo a las necesidades.

Los reportes Ad Hoc dan la posibilidad a los usuarios, y específicamente a los usuarios expertos, de ampliar el ámbito de sus análisis sin depender de los departamentos de TI. Como Eckerson indica, uno de los objetivos de los reportes Ad Hoc es potencializar a los usuarios para crear sus propios reportes sin necesidad de tener que esperar semanas o incluso meses para que el departamento de TI pueda atender su solicitud.

Como lo menciona Katelyn Bornak (Bornak, 2012), con los reportes Ad Hoc los usuarios, aún con poco entrenamiento pueden modificar o crear los reportes de acuerdo a los requerimientos individuales de cada uno. Por lo tanto, permiten una mejor utilización de recursos a nivel de la empresa dado que, como se mencionó anteriormente, los usuarios al estar en la capacidad de generar sus

reportes y análisis permiten que los departamentos de tecnología puedan enfocarse en otros proyectos que le agreguen valor a la organización.

Capítulo III. Marco metodológico

3.1 Tipo de investigación

3.1.1 Población

La población de estudio corresponde a una representación de las declaraciones de renta y ventas de una muestra simulada de los contribuyentes. Esto por cuanto los datos reales de declaraciones y contribuyentes no pueden ser utilizados dado que están protegidos por la legislación costarricense. Se hace necesario representar de forma genérica los posibles datos que se van a registrar, considerando propuestas de estructuras de almacenamiento basados en las estructuras cercanas a los ambientes reales. Es así, que para la generación de las estructuras se utilizan como base la distribución de casillas de los formularios disponibles en Internet, y la definición de los mismos disponibles en las guías como los instructivos para el llenado del impuesto de la renta y venta disponibles en Internet, en los que se describen los campos y lo que representa. Un ejemplo de este tipo es el “Instructivo para el llenado del formulario D101” (DGT, 2010)

Para las pruebas del prototipo se utiliza una muestra de declaraciones de renta, ventas e informativas. Lo anterior considerando que, dentro de las características del modelo, renta es anual, informativa es anual y ventas es mensual. Dado que el proyecto no abarca evaluaciones de rendimiento y dado que no se dispone de datos reales para pruebas, las muestras simuladas se utilizan para las pruebas del modelo de datos propuesto.

3.1.2 Técnicas

Se realizaron entrevistas a los analistas involucrados en el trabajo de campo con sistemas y usuarios de inteligencia (Ver apéndice A.1). De este modo, se

recopilará información acerca de las características de los orígenes, las declaraciones y los datos del contribuyente.

Una vez realizadas las entrevistas, se procederá a verificar las estructuras y los datos de cada origen. De esta manera, se analizará de acuerdo a las recomendaciones acerca del manejo de información fiscal para el análisis para realizar el prototipo de un modelo de base de datos por niveles para las declaraciones de Renta y Venta.

Asimismo, de acuerdo al modelo y las recomendaciones se establecerá un prototipo de una estrategia para el manejo de los datos fiscales para efectos de realizar procesos de inteligencia tributaria.

Finalmente, de acuerdo a cada una de las etapas anteriores, se elaborará un prototipo de un sistema que utilice algunas características del modelo propuesto por el proyecto.

3.1.3 Procedimientos

El primer paso es revisar los medios actuales en que son administradas las declaraciones de renta y venta. Adicionalmente, deben verificarse orígenes y formatos, requeridos para establecer las reglas de la consolidación de datos en los repositorios base. Asimismo, es necesario estandarizar los datos de manera tal que se tenga forma de relacionar a los contribuyentes con las declaraciones de forma consistente y eficiente.

Posteriormente, se debe definir un modelo para el repositorio unificado, de acuerdo a las características de la información y que permita realizar labores de análisis sobre datos fiscales.

3.1.4 Enfoque de la investigación

La investigación será de tipo cualitativa por cuanto la información que se obtiene a nivel de la institución e instituciones externas no es evaluable de forma cuantitativa dado que corresponde más a elementos del conocimiento y experiencia de los involucrados, que de valores medibles y cuantificables.

Se realizará un diseño evaluativo. De este modo, se ha recopilado de forma paulatina información de las fuentes internas así como de experiencias externas. Es así como la información obtenida se evalúa con el fin de identificar los elementos involucrados en el proceso de Inteligencia Tributaria, y de este modo tener las bases para plantear un camino base que permita establecer la propuesta de la estrategia. De esta manera, el trabajo va dirigido hacia una evaluación de campo e interacción con los involucrados para obtener el conocimiento de la situación interna, para posteriormente, a la luz de las experiencias en entidades externas, buscar puntos de encuentro para la solución.

De acuerdo con las entrevistas realizadas con el fin de recopilar información acerca de los procesos involucrados en la generación se buscó clarificar los elementos involucrados en distintos procesos de inteligencia en diferentes entornos.

De este modo se busca, a partir de la experiencia actual, comprender los procesos para guiar hacia una estrategia para la administración de los datos relacionados con las declaraciones de Renta y Venta. De esta forma, se hará la recolección y análisis de los datos para la formulación de la propuesta y tratar de establecer un camino adecuado para los requerimientos de inteligencia tributaria.

Capítulo IV. Propuesta de solución

4.1 Situación actual

4.2 Propuesta de solución

Por medio de la observación y consulta con algunos de los analistas, se han identificado puntos importantes que pueden ser considerados dentro de una propuesta de solución para tener un sistema de Inteligencia Tributaria. De este modo, es de relevancia considerar las reservas, que por lo general, tienen, tanto a nivel técnico como de usuario, en la utilización de Data warehouse y Cubos para análisis.

Si los datos no están consolidados en un repositorio unificado que permita tener a disposición de los analistas un punto de entrada único para sus revisiones, la consulta de los sistemas de forma separada puede aumentar los tiempos de búsqueda de datos y disminuir la eficiencia en contraposición a un sistema centralizado y automatizado.

Deben considerarse mecanismos que faciliten la adaptación de los analistas de acuerdo a lo que ya conocen. De esta manera, inicialmente los puntos a ser considerados son:

- Establecer las estructuras para el repositorio único
- Diseñar los procesos de carga
- Probar, por medio de un prototipo, la forma de consultar contra el repositorio único.

4.2.1 Repositorio único

El primer paso para establecer un proceso de inteligencia tributaria funcional y eficiente es el establecimiento de un lugar centralizado que sea el origen a partir del cual las herramientas puedan conectarse para las consultas. Adicionalmente, este repositorio consolidaría la información de las distintas fuentes tanto internas como externas, de forma que se tenga en un punto único la información estandarizada y homologada.

Es así como, en términos generales este repositorio debe contener información relacionada al registro de las personas jurídicas o físicas, sean o no contribuyentes registrados, declaraciones, datos de aseguradoras, bancos y otras instituciones. Lo anterior, con el fin de proveer a los equipos de inteligencia tributaria de un medio por el cual les sea posible obtener y cruzar información que pueda ser utilizada para determinar comportamientos anómalos que sean candidatos a ser auditados para establecer casos de elusión y evasión.

Este componente sería alimentado de fuentes diversas en formato y estructura. Por lo tanto, para los procesos de integración deben establecerse mecanismos de homologación de la información externa respecto a lo interno, manejo de datos faltantes, calidad de los datos. Lo anterior, dado que se requiere que los datos registrados identifiquen de forma correcta a los contribuyentes, dado que errores al asociar los datos de diversas fuentes que corresponden a un mismo individuo pueden generar información incorrecta al momento de definir los candidatos a ser estudiados.

Sumado las consideraciones de los orígenes y formatos de los mismos al momento de realizar los procesos de extracción, transformación y carga, es necesario definir la estructura que tendrá el repositorio analítico. Lo anterior, por cuanto esta estructura es determinante para el diseño de los procesos de carga, las reglas de negocio de las mismas, así como la interacción del usuario con las mismas

a partir de las herramientas, de forma que la experiencia de usuario sea natural y de fácil asimilación.

Tradicionalmente, los sistemas analíticos son vistos desde la perspectiva dimensional. Es decir, una base de datos desarrollada a partir de tablas de dimensión y hechos. A partir de lo anterior, las tablas de dimensión contienen los atributos cualitativos, y los hechos las medidas relacionadas.

En este esquema, las tablas usualmente no están normalizadas, se requiere velocidad en el tiempo de respuesta en contraposición a eficiencia en el espacio utilizado. Por lo tanto, estos repositorios pueden ser de gran tamaño y deben establecerse estrategias para la carga de los datos en las dimensiones y los hechos. A partir de estas dimensiones y hechos es posible definir cubos de consulta para que los analistas crucen la información y generen las estadísticas.

No obstante, dentro de las observaciones realizadas en entrevistas en distintos lugares, puede corroborarse de manera empírica y por conversaciones con los analistas, la resistencia a la utilización de cubo, dado que en ocasiones para los usuarios resultaba complicado o los tiempos de capacitación y herramientas disponibles no cumplían con las expectativas. Adicionalmente, dados los fracasos en proyectos anteriores existe a nivel técnico y de usuario una reacción adversa a términos como Data warehouse, cubo u otros relacionados a inteligencia de negocios.

Dado que los datos están descentralizados, ha sido labor de los usuarios encargados interactuar directamente con los datos crudos. Por lo anterior, los analistas han desarrollado ciertas destrezas en la utilización de herramientas como hojas de cálculo, para realizar cálculos y cruces de los valores para efectos analíticos. Por lo tanto, para estos usuarios lo natural ha sido ver los datos de forma tabular y aplicar determinadas fórmulas. Sin embargo, estos procesos no están estandarizados, centralizados ni automatizados.

Es importante resaltar el hecho de que los datos de las declaraciones, en su mayoría, corresponden a valores numéricos. También, las declaraciones representan, para el caso de renta, una fila con las columnas de montos para cada contribuyente, y la de ventas, un conjunto de filas y columnas con los datos. Entonces, las declaraciones pueden ser vistas como una tabla con varias filas, en las que las columnas representan los datos de los contribuyentes y los montos asociados. Es así, que con esta información, los encargados cargan la hoja de cálculo y pueden analizar y buscar posibles blancos.

Dado lo anterior, en un escenario como este podría ser adecuado la utilización de un modelo tabular para representar los datos fiscales. De este modo, para una primera etapa se establecería un almacenamiento tabular por niveles. Este esquema es independiente del motor, dado que utiliza las mismas reglas del modelo relacional. Para efectos de los laboratorios, el modelo fue probado y ejecutado tanto en Oracle como en SQL, manteniendo la misma funcionalidad y estructura física de las tablas.

En términos más específicos de la tecnología, actualmente herramientas como MSSQL Server 2012 permiten la creación y utilización de índices tabulares para mejorar el rendimiento. Además, permite tener modelos OLAP de forma multidimensional, pero adicionalmente introducen el almacenamiento por medio de bases de datos tabulares, con índices columnares.

De acuerdo a las características de las declaraciones se propone el desarrollo de una base de datos tabular por niveles, en la cual las estructuras reflejen los elementos involucrados en los elementos, como las declaraciones, de forma que pueda ser consumida más naturalmente por medio de otras herramientas.

Se busca tener un repositorio en el cual se almacenen tanto los valores de los orígenes así como columnas calculadas, de acuerdo a las necesidades de los clientes, de forma que la cantidad de cálculos al momento de consulta se reduzca en lo posible y que los usuarios puedan hacer sus reportes de la forma como ya están habituados a hacerlo.

Dentro del modelo se deberían definir niveles, con el fin de poder comparar elementos que tienen diferente periodicidad (por ejemplo las declaraciones de venta que son mensuales, contra renta que son anuales). Adicionalmente, pueden tenerse niveles compuestos en los que se consoliden elementos de diferentes declaraciones.

Cada uno de los niveles tiene un conjunto de detalle y de columnas diferentes. El nivel más bajo es el que contiene la totalidad del detalle. También, de acuerdo a las entrevistas y necesidades de los usuarios, los siguientes niveles tendrían menos columnas o nuevos elementos calculados. De esta forma, puede irse navegando entre los niveles desde el superior a los inferiores, de forma que pueda obtenerse el rastro completo de la información fiscal.

Es importante resaltar, que el modelo tabular es una base a partir de la que posteriormente puedan extenderse nuevas opciones como crear cubos sobre la base ya centralizada. De igual forma, en futuros trabajos podría ser relevante estudiar en varias tecnologías, la implementación por niveles utilizando bases de datos Relacionales – Objeto para establecer herencia entre las tablas de los distintos niveles.

4.2.1.1 Modelo centrado en el contribuyente

Los análisis de Inteligencia tributaria, en su mayoría buscan determinar posibles casos de estudio y auditoría. De esta manera, estas verificaciones son realizadas sobre los datos del contribuyente. Por lo tanto, se recaba información para los contribuyentes tanto a lo interno como de entidades externas.

Tomando en consideración la importancia del contribuyente, el modelo propuesto tiene que ser de cierta manera centrado en el contribuyente. Lo anterior, por cuanto las distintas entidades que se representan en la estructura del repositorio único están relacionadas a los datos del contribuyente, como se representa en la Figura 4.

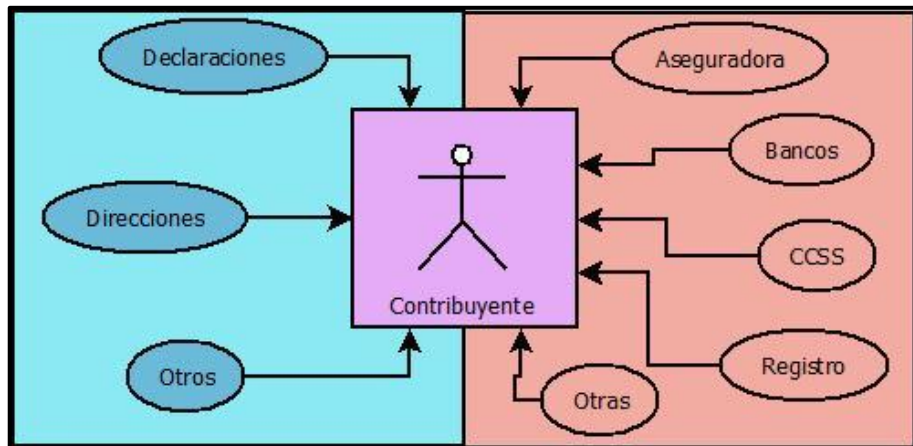


Figura 4: Esquema conceptual Contribuyente centrista
Tomado de: Elaboración propia

Es así como los datos generales de las personas (sean físicas, jurídicas o extranjeras) son almacenados en la tabla PERSONAS. La tabla de personas es el centro del modelo, dado que las otras tablas lo referencian. De igual forma, al ser una tabla de referencia y de poco cambio no se utiliza versionamiento como el resto de colecciones. Para las relaciones con las otras colecciones se utiliza una llave sustituta, a la cual se referirá como número de identificación fiscal (NIF). A partir de este esquema es posible realizar análisis con los datos de los contribuyentes y adicionalmente crear expedientes para revisión de los datos que están registrados, con el fin de tener un panorama más amplio de los datos registrados.

Para efectos de los procesos de carga, deberá disponerse de una tabla de conversión de las identificaciones por origen. Lo anterior, por cuanto las identificaciones en cada uno de los orígenes pueden tener representaciones diferentes. El objetivo de esta tabla es que cada contribuyente pueda ser asociado de forma única con el NIF que le corresponda, con el fin de evitar tener los datos de un mismo contribuyente registrado como si fuese diferentes personas, lo cual sesgaría los análisis.

NIF	Identificación	Origen
4565	10123045640	Hacienda
4565	0101230456	Seguro
4565	1-0123-0456	Banco
...	...	...

Tabla 1: Ejemplo tabla conversión identificaciones

Como puede verse en el ejemplo anterior, una misma identificación puede ser representada de diversas maneras de acuerdo a la institución que la provee. Así por ejemplo, para Hacienda, una cédula física nacional, además de los dígitos usuales agrega un verificador de dos dígitos, de la misma forma puede ocurrir que las identificaciones tengan separadores u otros dígitos adicionales, siempre para representar a la misma persona.

4.2.1.2 Datos generales

Dentro de las estructuras propuestas se establece una sección general con datos que se comparten para cada uno de los elementos involucrados. Los datos de esta sección se dividen en datos del contribuyente y de fecha.

Para lo correspondiente a los datos del contribuyente debe ser definida una llave sustituta que permita identificarlo en cada una de las estructuras. De esta manera, se define una tabla llamada Persona, la cual tiene los datos generales del contribuyente y la identificación. Dado que las identificaciones varían de acuerdo al tipo de persona (física, jurídica o extranjero), no resulta práctico utilizarlo como llave a través de las tablas de declaración. Por lo tanto, como llave sustituta se define un número de identificación tributaria (NIT).

De esta forma todas las fuentes de datos del Ministerio de Hacienda tendrían un dato en común que las relaciona. Mediante el NIT se puede obtener de los diferentes sistemas de Hacienda los datos del contribuyente, sus declaraciones, pagos, etc. Adicionalmente el NIT sirve para segmentar a los contribuyentes por actividad económica.

Adicionalmente, dado que los datos de los contribuyentes y las declaraciones pueden variar, para cada registro en las tablas se agrega un indicador para establecer si ese es la última versión o no. De esta forma, siguiendo un comportamiento de dimensión lentamente cambiante de tipo 2, se mantiene tanto la versión actualizada de los datos como los históricos que existan del mismo.

La tercera sección del área común corresponde a la segregación de fechas. En esta sección se segregan las fechas que el analista indique. De esta manera, se almacena de forma segregada para una fecha determinada el día, mes, año, trimestre, semestre y año fiscal correspondiente. Está segregación tiene el fin de facilitar búsquedas y agrupaciones por períodos para todos los elementos.

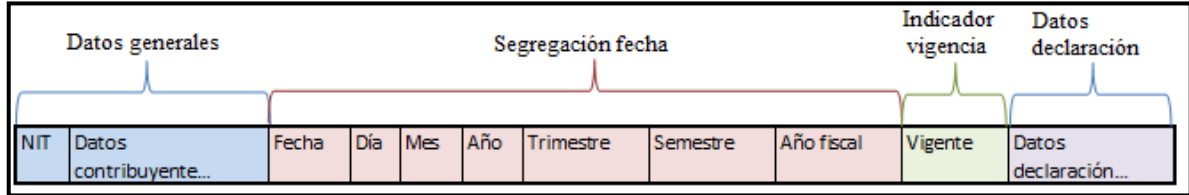


Figura 5: Estructura conceptual tablas de declaración
Tomado de: Elaboración propia

4.2.2 Niveles de datos

Dadas las características de las declaraciones algunas representan datos correspondientes a distintos periodos, sean estos diarios, mensuales, bimensuales, anuales u otros. De esta forma, si se quisiera comparar un modelo anual contra uno mensual, primero sería necesario realizar la sumatoria de las doce declaraciones mensuales de ese periodo para poder representarlos en los mismos términos. Lo anterior, implica consumir recursos de la base de datos en la ejecución de las operaciones y posiblemente incrementar los tiempos de espera.

En pruebas preliminares el modelo mantenía solo el almacenamiento en el nivel más bajo. Por lo tanto, si se intentaba comparar una declaración mensual con una anual, en tiempo real se realizaba la agrupación de la declaración mensual por contribuyente para usar el total en la comparación. No obstante, este proceso consumía muchos recursos y el tiempo de respuesta se alargaba más de lo que el analista estaba dispuesto a esperar. A raíz de lo anterior se replanteo la forma de representar los datos hacia una representación por niveles.

En el contexto del modelo propuesto se incluye el concepto de niveles. Los niveles corresponden al equivalente a una pirámide en la que cada nivel representa un grado de detalle y de periodicidad distintos. De esta manera, sería posible tener un nivel de ventas que incluya todo el desglose de las ventas para cada mes, un nivel superior que corresponda al resumen total de las ventas del mes y un nivel superior con el total de todas las ventas durante el periodo fiscal.

Si se requiere comparar una declaración anual, como la de renta, contra una mensual, como ventas, la operación sería realizada ya no sobre el nivel mensual, sino sobre el nivel de ventas mensual. Con esto, la operación se reduce a una consulta simple sobre la tabla.

En el esquema de niveles es posible tener súper niveles, es decir, mezclar niveles del mismo tipo de periodicidad para crear un nivel combinado. De esta manera, el nivel combinado tendría un subconjunto de columnas de todos los niveles que lo componen, y permite poner a disposición del analista los datos desde una sola tabla, con lo cual no se requiere consultar varias tablas para obtener los datos de un mismo contribuyente.

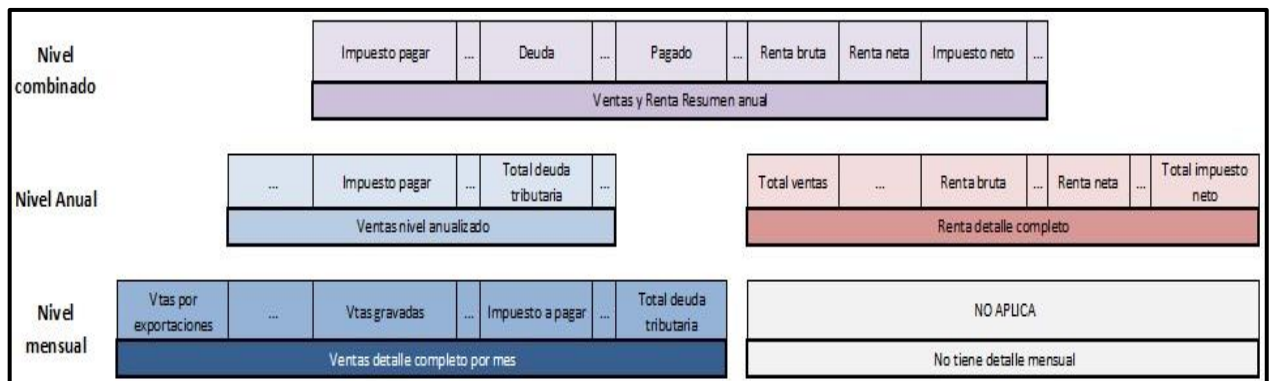


Figura 6: Niveles de declaración
Tomado de: Elaboración propia

Para la creación de los niveles y las columnas calculadas se requiere recabar con los usuarios cuales son los elementos que consideran importantes para sus procesos de revisión, cuales son las fórmulas más comunes que utilizan y cuáles son las combinaciones más usuales que deben realizar. Todo esto, con el fin de plasmar

a nivel de las estructuras tabulares estos elementos y dejarlos disponibles desde base de datos sin necesidad de consumir recursos del motor realizando cálculos que se sabe son frecuentes por parte de los usuarios.

De acuerdo al esquema de niveles, las tablas que se ubican en el mismo nivel pueden compararse de forma equivalente entre sí sin necesidad de realizar cálculos o agrupaciones en ejecución como si ocurriría al tratar de comparar una declaración de tipo mensual con una anual. No obstante, el mismo modelo da la posibilidad de realizar operaciones de este tipo, aunque requieren de más procesamiento. Por lo tanto, de preferencia, los campos que se requieren operar o combinar de esta manera de forma frecuente se recomienda tenerlos cargados en los niveles correspondientes.

Si se comparan dos colecciones, una de periodicidad mensual y la otra anual, para efectos del estudio sería necesario realizar la totalización de las declaraciones mensuales para el año seleccionado, de forma que se tenga el total anual para ambos casos. En contraposición, si por ejemplo se selecciona una periodicidad mensual, las colecciones anuales, trimestrales, semestrales, o de cualquier periodicidad mayor a la mensual, quedan excluidas del estudio por cuanto no habría una forma fiable de poner ambas declaraciones en los mismos términos (dado que por ejemplo dividir una declaración anual entre doce no necesariamente daría un valor totalmente certero dado que ciertas declaraciones, como la de ventas, pueden verse afectadas con altas y bajas de acuerdo a las temporadas).

Las colecciones que pueden actuar por sí mismas o proporcionadas, son las que tengan periodicidades iguales o menores a la seleccionada. En la figura puede observarse una jerarquía de periodicidades, donde una anual contiene al resto, y una semestral está contenida dentro de las anuales e incluye a las mensuales y trimestrales.

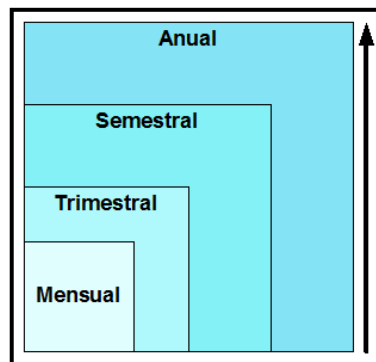


Figura 7: Jerarquía de periodicidades
Tomado de: Elaboración propia

4.2.3 Arquitectura del proceso de carga

Para la carga de los datos en el repositorio unificado, es necesario considerar las características de integración involucradas desde los diferentes orígenes hasta la base de datos final. Lo anterior, dado que los datos de los diferentes orígenes tiene que ser puestos en los mismos términos al momento de ser cargados en el repositorio final.

Para el proceso de carga se utiliza un modelo de carga hacia un repositorio intermedio (o Staging Area), conocido como tanque. Una vez ahí se pasa por un proceso de estandarización de los datos para realizar un segundo proceso de extracción, transformación y carga, ahora del tanque a la base de datos unificada. A continuación puede observarse el gráfico con la arquitectura base del sistema.

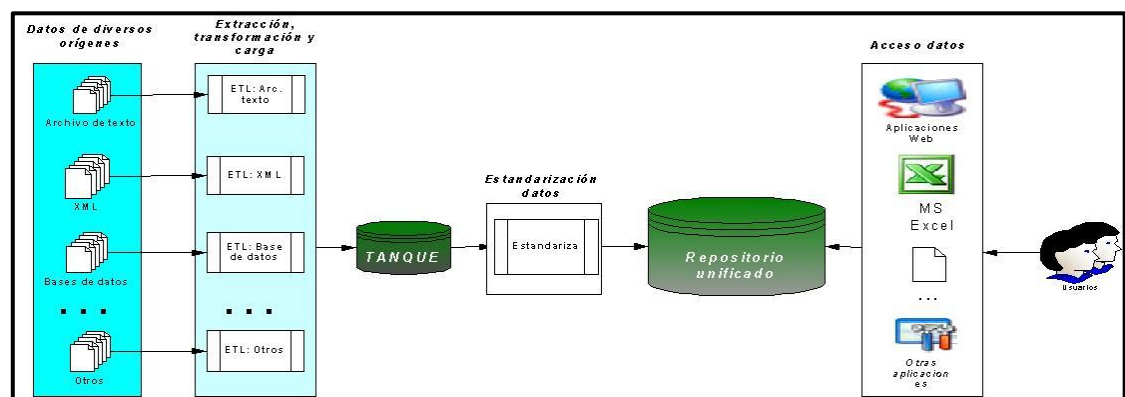


Figura 8: Proceso de carga
Tomado de: Elaboración propia

Los elementos que intervienen son los siguientes:

- *Orígenes de datos:* Los documentos recibidos van a provenir de distintos entes, tales como bancos, aseguradoras, Hacienda y otros. Asimismo, los datos son contenidos en diferentes formatos, desde archivos de texto, XML, bases de datos y otros tipos.
- *Proceso de extracción transformación y carga hacia el tanque:* debido a las diferencias en los orígenes de datos y sus formatos la extracción, transformación y carga de los objetos no es un proceso genérico. En contraposición se hace necesario crear cada uno de los procesos de acuerdo a los diferentes formatos y estructuras de los archivos.
- *Tanque:* En el tanque se realiza el traslado puro de los datos desde el origen a las tablas de mapeo.
- *Estandarización:* El proceso de estandarización consiste en obtener los datos que han sido cargados en el tanque y realizar las cargas de los campos estandarizados de acuerdo al formato que utilice el repositorio único para enlazar los datos entre sí. De esta forma las cédulas se traducen como NIT y se realizan las relaciones entre las personas y los datos que tienen asociados.
- *Repositorio único:* Es el repositorio unificado tabular con los datos estandarizados y que pueden ser explotados mediante técnicas de inteligencia. Es una estructura contribuyente – centrista. El objetivo es que los analistas puedan realizar sus análisis con las estructuras cargadas, los campos pre-calculados y nuevos campos que requiera generar.
- *Acceso a datos:* El acceso a los datos al repositorio único puede hacerse de diversas formas. Una vez consolidados los datos en un solo repositorio permite que los usuarios tengan un único punto de consulta

en un formato estandarizado en lugar de tener la información dispersa en las distintas fuentes, como ocurre actualmente. Además, es posible acceder, con los permisos correspondientes, al repositorio utilizando diferentes herramientas, sean estas construidas a la medida o creadas por terceros (Ej. MS Excel), de forma que los analistas puedan interactuar con los datos de la forma que requieran.

4.2.3.1 Actualización

Como se muestra en el diagrama de arquitectura, los datos que van a alimentar el tanque pueden provenir de un sinnúmero de instituciones y sistemas diferentes. De la misma forma, cada origen podría tener reglas de formato distintas, así como sus formatos van a variar de acuerdo a las características propias de lo que se va a trasladar. Por tanto, en el momento que se considere necesario realizar programas automatizados para la carga de los datos debe considerarse esto como un proceso no genérico, de acuerdo a las características propias de los sistemas.

Por lo tanto, para cada origen es necesario crear un paquete de actualización que permita inicialmente cargar desde la fuente hasta el tanque, haciendo los procesos de extracción, transformación y carga propios de la misma y de acuerdo a las reglas y el formato de la fuente cargada. Una vez cargada la fuente en el tanque el proceso continuaría con la estandarización de los datos y su carga en el repositorio final, realizando el mapeo a la colección que corresponda.

Como se ha mencionado, el proceso de actualización conlleva la extracción de los datos y las transformaciones de los mismos, de forma que puedan ser mapeados en el tanque y posteriormente en el modelo. De esta forma, al llegar el momento de la estandarización, un contribuyente que ingrese por medio de su identificación debe ser mapeado en la colección Personas, para extraer el NIF que lo va a representar en las relaciones.

Algunos programas van a tener una implementación sencilla, que consiste únicamente de ir a un directorio específico para extraer los datos de un archivo de texto que deja automáticamente el sistema con los datos de origen. Otros programas van a tener que implementar consultas a bases de datos para extraer la información.

Es claro que agregar una nueva colección siempre va a implicar la creación de un paquete de actualización.

4.2.3.2 Frecuencia de actualización

Por su naturaleza, cada colección de datos va a tener frecuencias de actualización diferentes. Esos procesos de actualización normalmente implican el procesamiento de grandes volúmenes de datos, por lo que lo más conveniente sería ejecutar estos procesos durante la noche y madrugada.

La frecuencia de actualización varía de acuerdo a las características de cada colección y las periodicidades asociadas. Así habrá algunos procesos que se ejecuten diariamente, mientras que otros se pueden ejecutar semanal, mensual, semestral o anualmente.

4.3 Prototipo de estrategia

La inteligencia tributaria requiere de datos de diferentes fuentes tanto internas como externas para el establecimiento de los posibles casos sospechosos de evasión o elusión fiscal. De esta forma, una estrategia preliminar de inteligencia fiscal, entre otras cosas debería considerar:

- Estrategia de recolección de datos de contribuyentes
- Estandarización datos entre instituciones
- Consolidación de datos de las fuentes
- Utilización de herramientas de análisis
- Seguimiento de casos

Como puede observarse en la siguiente figura, el proceso abarca los datos internos y externos que deben ser consolidados en el repositorio unificado. Una vez llegado a este punto, se requieren herramientas que permitan explotar los datos para generar casos de estudio. Finalmente, los casos generados deben ser administrados por medio de un sistema que dé seguimiento en cada uno de los pasos del proceso de auditoría.

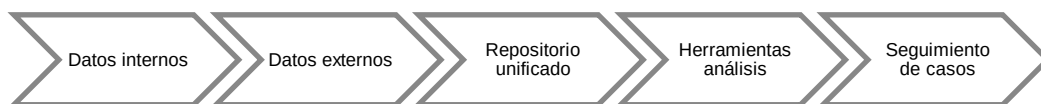


Figura 9: Prototipo propuesta de inteligencia tributaria
Tomado de: Elaboración propia

A continuación se describen de forma general cada uno de los elementos anteriormente mencionados.

4.3.1 Estrategia de recolección de datos de contribuyentes

Como se mencionó en capítulos anteriores, el Ministerio de Hacienda ha dado pasos en la definición del establecimiento de obligatoriedad de la entrega electrónica de datos de los contribuyentes y las declaraciones de los distintos impuestos.

De esta manera, se puede reducir la cantidad de errores de digitación y cálculos de los montos a pagar dado que estos se realizan de forma automatizada por los sistemas. Al mejorar los mecanismos de recaudación y dar mayores opciones para recaudar los datos, se puede llevar a incentivar la presentación puntual de las declaraciones de los contribuyentes, dado que los procesos pueden hacerse de forma virtual y por Internet.

Las políticas de este tipo deben venir acompañadas de una infraestructura robusta y una aplicación confiable, que permitan a los contribuyentes presentar y pagar sus obligaciones en línea.

La automatización de los procesos de generación y presentación de las declaraciones de impuestos, permiten que los datos puedan ser validados y registrados en línea y en el momento, de forma tal que la calidad de los datos sea mejor. Anteriormente, los formularios registrados a mano podían incluir errores de digitación que no eran detectados hasta el momento de la digitación manual en los sistemas.

Con la presentación electrónica de las declaraciones y el registro inmediato en los sistemas centrales de Hacienda, Tributación ha ido, paulatinamente, poniendo a disposición de los contribuyentes el pago en línea de sus obligaciones. De esta manera, dado que los contribuyentes ya no requieren hacer filas en los bancos y tienen a disposición los medios electrónicos para el pago las 24 horas del día durante los siete días de la semana, el proceso de pago se facilita e incentiva. Lo anterior por cuanto, el pago por Internet puede hacerse desde cualquier lugar y el proceso requiere pocos minutos.

Los registros automatizados pondrían a disposición los datos fiscales en un tiempo menor del que requeriría digitar una a una todas las declaraciones. Con esto se pretende reducir errores de digitación y el tiempo de trabajo requerido por los funcionarios de Hacienda en el proceso manual y aprovechar los recursos para tener los datos para análisis a la mayor brevedad posible.

El Ministerio de Hacienda ya ha dado los primeros pasos en la definición legal de la obligatoriedad de la presentación por Internet de algunos impuestos (como el de ventas) y avanza con otros, como el de la renta (que actualmente se presenta en papel y por Internet). Este proceso va a permitir incrementar la calidad de los datos capturados, así como la validación y verificación de algunos datos de la declaración al momento de ser presentada, para reducir inconsistencias que puedan ser detectadas de forma automática en ese momento y realizar cálculos al momento de intereses y total a pagar.

En el mismo sentido, se da la posibilidad, por medio del sistema de Conectividad, de realizar los pagos por medio de las plataformas de los bancos, tanto en línea como físicas, aumentando de esta manera la cantidad de canales de pago disponible y facilitando el proceso de pago, de forma que sea rápido y conveniente.

Dentro de los puntos a mejorar se encuentra la infraestructura, para aumentar la robustez de la plataforma y permitir a todos los contribuyentes presentar a tiempo sus declaraciones. Para esto se requiere que la plataforma tecnológica sea diseñada de forma tal que puedan ser balanceadas las cargas de datos y permitan múltiples conexiones en los días pico de presentación, para evitar contratiempos por parte de los que presentan así como para que Tributación no requiera solicitar prorrogas en las fechas por problemas en sus sistemas.

4.3.2 Estandarización de datos entre instituciones

Adicionalmente a los datos internos de Tributación, como las declaraciones y la información de los contribuyentes, un elemento de relevancia que debe ser considerado dentro de los análisis fiscales corresponde a la información de fuentes externas como bancos, aseguradoras y otras.

Los sistemas de inteligencia tributaria tienen como fuente de información lo que dicen los contribuyentes. Esta información debe ser contrastada con otras fuentes para determinar posibles inconsistencias que deban ser auditadas. De esta manera, los analistas pueden verificar elementos como signos de riqueza o inconsistencias entre lo que declaran y lo que registran las instituciones.

Un ejemplo de esto es el impuesto solidario. La obligación del pago de este impuesto está definida por el valor de las casas de habitación. De este modo, el monto reportado por las personas físicas a Hacienda puede ser contrastado contra los valores registrados en una aseguradora para la casa. Es así como, Inteligencia Tributaria podría detectar casos de personas que en Hacienda reporten menos de lo que vale su casa de acuerdo a las aseguradoras. No obstante, si la información no está disponible e integrada, estos análisis no serían posibles o requerirían de trabajo adicional para realizarlos.

4.3.2.1 Situación actual

Como parte de una estrategia de Inteligencia Tributaria, uno de los puntos que debe establecerse con claridad y de manera conjunta son los formatos y mecanismos para compartir los datos. Actualmente, de acuerdo a entrevistas y la experiencia empírica de proyectos anteriores, se ha podido observar que las instituciones ponen a disposición de Hacienda la información en distintos formatos de archivo, como archivos de texto, hojas de cálculo, bases de datos Access. De la misma forma, los datos de estos repositorios presentan diferentes formatos de los

datos, como identificaciones o códigos. Entonces, en Hacienda la cédula se almacena con un formato que agrega dos dígitos verificadores, en los bancos las cédulas pueden o no venir con guiones. En el caso de las identificaciones de extranjeros es aún más complejo, dado que Migración puede establecer cambios en la forma de generar las identificaciones.

La no estandarización de los datos implica realizar procesos de carga personalizados para cada institución y, en caso de que las instituciones realicen cambios, deben adaptar cada uno de estos paquetes. Además, cuando se necesite información de una nueva institución, deberá hacerse un nuevo desarrollo que se adapte a la misma. Este proceso puede consumir mucho tiempo dado que prácticamente se requiere un paquete de carga por institución y puede requerir gran cantidad de validaciones de acuerdo a cada formato recibido. Asimismo, de acuerdo a lo indicado por los analistas, en caso de que la institución genere un archivo incorrecto, la información del mismo podría tener que ser descartada en su totalidad, con lo cual se pierde potencial información que puede ser útil a los analistas.

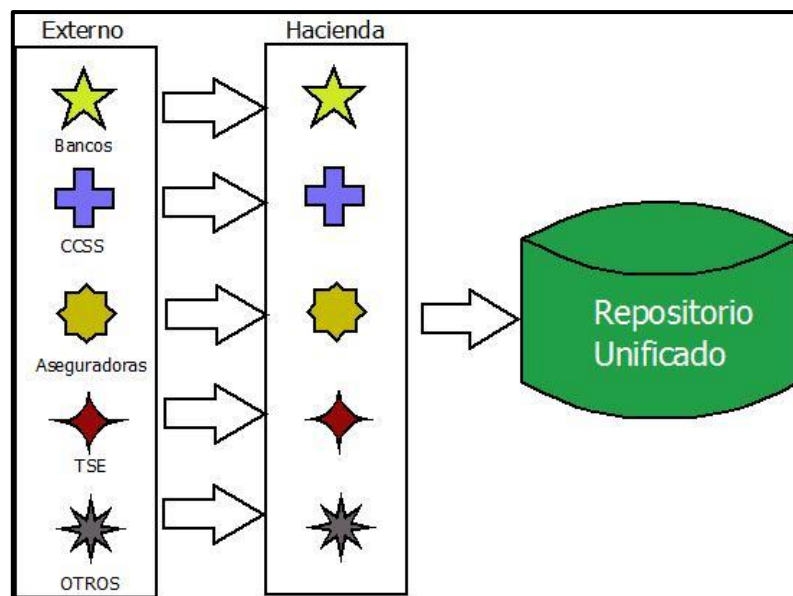


Figura 10: Obtención de datos situación actual
Tomado de: Elaboración propia

Otro punto adicional tiene que ver con la completitud y la calidad de los datos. En algunas instituciones, datos que pueden ser útiles para posibles análisis no están

completos. Por ejemplo, algunos estudios de inteligencia podrían requerir tener disponible para un contribuyente quiénes son sus padres o sus hijos y así realizar un análisis amplio de ingresos en el entorno de la persona. No obstante, el Tribunal Supremo de Elecciones no tiene disponible esta información para algunas personas con más de 20 años. Lo anterior, por cuanto anteriormente los procesos de registro no eran tan estrictos como en la actualidad y no siempre eran registradas las identificaciones de los padres. Por lo tanto, en este momento, para esas personas, no es posible establecer este tipo de relaciones.

4.3.2.2 Escenario propuesto

Dado lo anterior, uno de los puntos que deben ser considerados para que la estrategia de Inteligencia Tributaria sea más eficiente, es la definición, por parte del gobierno, utilizando medios legales, de un estándar entre instituciones para compartir información entre ellas. De esta manera, cada institución tendría que crear los mecanismos para generar archivos bajo el formato estándar de estructura, códigos e identificaciones, así como los canales para compartir los datos, como por ejemplo Servicios Web con seguridad, Servidores de archivos u otros.

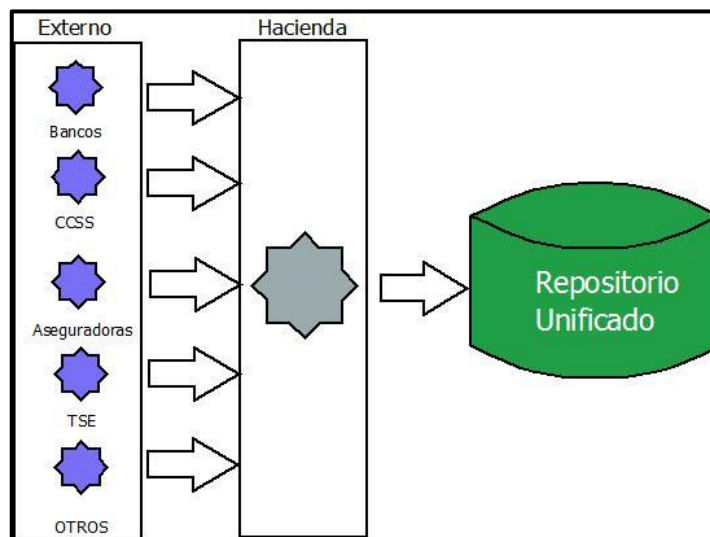


Figura 11: Formato estandarizado de datos
Tomado de: Elaboración propia

Para compartir los archivos pueden definirse, ya sea archivos de texto plano posicional o en formato XML. Cada institución debe indicar claramente la estructura y los campos de cada uno de los campos y los tipos correspondientes. Asimismo, se requiere una definición, a nivel general del país e instituciones, de la forma en que se van a representar las identificaciones para las personas físicas, jurídicas y extranjeras, de manera que sin importar el origen de los datos, pueda ser establecido claramente a quien pertenecen y evitar la fragmentación de datos o la incorrecta asignación de datos a las personas.

En resumen, la definición del proceso de estandarización debería contemplar, al menos:

- Mecanismos de conexión y seguridad entre las fuentes y el recaudador.
- Estandarización de estructuras y formato de archivos.
- Homologación de identificaciones y códigos.
- Definición de protocolos de comunicación y aplicación de cambios.

4.3.3 Consolidación de datos de las fuentes

En secciones anteriores se ha descrito con más detalle lo concerniente a la necesidad de tener un repositorio unificado, las características de su arquitectura y las cargas. Por lo tanto, se reseña de forma resumida lo relativo al mismo en esta sección. Actualmente los analistas necesitan consultar los datos de distintas fuentes o archivos al momento de realizar los estudios. Lo anterior, por cuanto los datos no están centralizados en un repositorio único y estandarizado para tal fin.

Como en otros procesos, los datos son la base a partir de la que puede generarse un conjunto de información de relevancia que permitan concretar cada uno de los valores en conocimiento, como por ejemplo para la detección de anomalías en declaraciones.

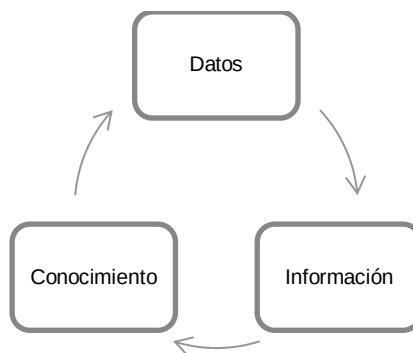


Figura 12: Datos – Información – Conocimiento
Tomado de: Elaboración propia

De esta manera, esta sección de la estrategia corresponde a consolidar los datos provenientes de cada una de las fuentes internas, así como los datos capturados de las instituciones, dentro de un repositorio único y estandarizado, que permita tener un punto común de acceso a los analistas y que adicionalmente provea de datos precalculados.

El repositorio unificado, es el núcleo del sistema de inteligencia, es hacia donde llegan los datos de las diversos orígenes, se almacenan de forma estandarizada y homologada, y adicionalmente, es el lugar de origen para las consultas de los analistas acerca de los datos fiscales.

4.3.4 Utilización de herramientas de análisis

El equipo de analistas y tomadores de decisión en temas de inteligencia tributaria no necesariamente son funcionarios con conocimiento técnico. Por lo tanto, se requiere establecer mecanismos por medio de los cuales puedan acceder a los datos de forma efectiva para realizar sus estudios. De esta manera dos posibles caminos pueden ser seleccionados.

El primero de estos es utilización de herramientas de terceros, como por ejemplo hojas de cálculo, que permitan conectarse al repositorio unificado e interactuar y combinar los datos de la forma que se crea conveniente.

Con ese camino las principales consideraciones tienen que ver con licenciamiento, capacitación y cambios en el futuro, dado que el control de la aplicación está en terceros y se depende de la compatibilidad y cambios entre versiones.

El segundo camino es la construcción de herramientas a la medida. De acuerdo a los requerimientos de los usuarios, estas herramientas pueden ser creadas de forma más cercana a las necesidades del negocio y las características de las estructuras existentes.

De esta forma, por ejemplo, sería posible crear una aplicación que obtenga los datos, permita consultar el expediente de las personas y marcar en el repositorio posibles casos que deban ser estudiados.

En contraposición, una herramienta de este tipo puede llegar a ser complicada y requiere de una amplia colaboración entre usuarios y desarrolladores para recopilar, analizar, diseñar y construir estas aplicaciones y que sean útiles en su función,

Para esto, es importante establecer esta relación usuario – desarrollador, dado que el usuario es el principal beneficiado y conocedor de las reglas, por tanto, puede convertirse en un socio estratégico o en un enemigo fuerte del sistema. Por lo tanto, gran parte del éxito o fracaso de esta etapa se centra en este elemento.

El establecimiento del repositorio único permite la flexibilidad de utilizar tanto herramientas de terceros como a la medida. Anteriormente, con los escenarios desconcentrados el acceso a los datos era complicado y disperso. Con esta propuesta, la herramienta de análisis sería la cara para los usuarios que requieren la información al momento y sin tener que depender de los departamentos de tecnología.

4.3.5 Seguimiento de casos

Parte de los objetivos de la inteligencia tributaria consisten en la detección de casos anómalos que puedan ser considerados como candidatos a auditorías. De esta forma, a partir de los datos almacenados en el repositorio unificado, los cuales son extraídos por medio de las herramientas de análisis, los analistas podrían establecer cuales personas físicas o jurídicas deberán ser auditadas.

Ahora bien, en este escenario, el grupo de Inteligencia requiere de mecanismos que les faciliten el seguimiento de cada uno de los casos. Es así como, una herramienta de seguimiento de casos puede resultar de utilidad para registrar cada uno de los pasos a seguir en el proceso de investigación, desde la apertura hasta los distintos cierres.

El sistema debe ser capaz de obtener del repositorio unificado los datos que se han sido marcados para estudio. A partir de estos, el sistema debe ser capaz de realizar la asignación de cada uno a los grupos de auditores. Adicionalmente, para cada paso en el flujo de la investigación, deben quedar registrados las descripciones y documentos recabados.

Un sistema de inteligencia sin un sistema de seguimiento de casos puede implicar retrabajo o problemas en el control de los estudios. De este modo, la no automatización podría implicar mayor carga de trabajo o duplicación de labores por parte de los analistas, así como la manipulación inadecuada de datos que impida detectar de forma adecuada los casos.

Al automatizar el proceso, cada persona es asignada a un caso específico por el sistema, y por medio de los expedientes generados a partir de la información externa e interna, así como las auditorías de campo, pueden realizar su labor de forma más eficiente y con esto lograr un mayor control de los contribuyentes y sus obligaciones.

4.4 Prototipo de aplicación

La aplicación representada por el prototipo parte de dos elementos importantes para los analistas:

- Las hojas de cálculo
- El expediente del contribuyente.

En el primer caso, la aplicación busca representar la generación de cálculos por medio de hojas de cálculo. De esta forma, se utiliza el concepto de columnas. El analista va creando de forma interactiva una hoja de cálculo con operaciones y filtros. De esta forma, se generan subconjuntos de datos de contribuyentes que puedan ser utilizados para analizarlos.

La segunda parte la aplicación permite revisar el expediente del contribuyente. En el mismo se incluye información tributaria como la obtenida de los entes externos. De esta manera, sea por las hojas de cálculo o las búsquedas específicas, puede revisar el histórico de cada uno de los elementos como declaraciones, teléfonos, direcciones y otros. A continuación, se presenta un prototipo con algunas de las características de una aplicación básica para consulta de datos.

4.4.1 Base de datos

El motor de base de datos no es limitado por el modelo propuesto. Para efectos de los laboratorios se realizaron simulaciones tanto en Oracle como en SQL Server 2012. Lo anterior, por cuanto estos son los motores utilizados en distintas empresas. Las pruebas finales fueron realizadas sobre SQL 2012 por las características que la misma ofrece para la administración de los datos del sistema, como índices columnares y otros que permiten mejorar el rendimiento de las consultas en repositorios con gran cantidad de datos involucrados.

En cuanto al almacenamiento de datos, el proceso consiste en la carga de datos desde los orígenes hacia el tanque para almacenar finalmente en el repositorio final. Asimismo, para el caso de la aplicación de casos se han creado las tablas de diccionario, así como las correspondientes a cada uno de los mantenimientos asociados. Por otra parte, están las tablas con los datos cargados, las cuales son las involucradas en la obtención de la información necesaria para hacer un estudio.

4.4.2 Modelo de datos

A partir de los datos que deben ser almacenados y los posibles procesamiento a realizar, a continuación se describen algunas de las colecciones importantes que podrían ser utilizadas dentro del modelo. Estas colecciones son:

- Diccionarios
- Personas
- Registros
- Expedientes

4.4.2.1 Diccionarios

El diccionario de datos de la aplicación de consulta va a saber cuáles son los campos que componen una colección. Del mismo modo, registra la forma correcta de consultar y mostrar los datos. De esta forma, se tiene una relación entre los campos físicos, como están definidos en las tablas, y los nombres lógicos, como se podrían mostrar en pantalla, en el expediente y en los controles de usuario que correspondan.

De acuerdo con lo anterior se tendría tres diccionarios principales, los cuales serían:

- Diccionario de tablas: Tablas de repositorio único
- Diccionario de columnas: Diccionario de las columnas asociadas a una tabla del repositorio único.
- Diccionario de operaciones: Diccionario de las operaciones para poder aplicar filtros a los casos.

Diccionario de tablas

N°	Identificación campo	Tipo	Descripción
1	IDTABLA	INT	Identificación de la tabla
2	NOMBREFISICO	VARCHAR(30)	Nombre físico de la tabla
3	NOMBRELOGICO	VARCHAR(50)	Nombre lógico de la tabla.
4	ACTIVA	BIT	Indicador de si la tabla permanece activa
5	COD_PERIODICIDAD	VARCHAR2(1)	Código de periodicidad de la tabla.
6	MOSTRAREXPEDIENTE	VARCHAR2(2)	Indica si debe mostrarse en el expediente

Tabla 2: Diccionario de tablas

Diccionario de columnas

N°	Identificación campo	Tipo	Descripción
1	IDCOLUMNA	INT	Identificación de las columnas
2	IDTABLA	INT	Llave foránea de la tabla a la que pertenece
3	NOMBREFISICO	VARCHAR(30)	Nombre físico de la columna
4	NOMBRELOGICO	VARCHA250)	Nombre lógico de la columna
5	TIPODATO	VARCHAR2(30)	Tipo de dato de la columna
6	AGRUPA	BIT	Indica si se puede agrupar por esta columna
7	FILTRA	BIT	Indicador de si es columna filtro.
8	SUMARIZA	NUMBER(1,0)	Indicador de si se puede realizar operaciones matemáticas sobre el campo
9	VISIBLE	CHAR(1)	Indica si es visible al crear el expediente
10	ORDENDESPLIEGUE	NUMBER	Orden de despliegue al momento de cargar el expediente
11	ES_CATALOGO	BIT	Indica si la columna es catálogo
12	PA_CATALOGO	VARCHAR(30)	Se muestra el nombre del catálogo
13	ACTIVA	BIT	Indica si la columna está activa o no
14	COD_TIPODATO	INT	Código de la tabla tipo de dato

Tabla 3: Diccionario de columnas

Diccionario de operaciones

N°	Identificación campo	Tipo	Descripción
1	ID_OPERACION	NUMERIC(18,0)	Identificación de la identificación
2	DESCRIPCION	VARCHAR(255)	Descripción de las operaciones
3	OPERADOR	VARCHAR(30)	Operador lógico o de comparación según sea el caso
4	ID_TIPO	NUMERIC(18,0)	Identificación del tipo de operación
5	ACTIVA	CHAR(1)	Indica si la operación está activa o no
6	VALORDEFECTO	NUMERIC(18,0)	Valor que tiene por defecto
7	SUSTITUIRCERO	VARCHAR(1)	Sustituir el cero por el valor indicado

Tabla 4: Diccionario de operaciones

4.4.2.2 Personas

Los datos generales de las personas (sean físicas, jurídicas o extranjeras) son almacenados en la tabla PERSONAS. La tabla de personas es el centro del modelo, dado que las otras tablas lo referencian. De igual forma, al ser una tabla de referencia y de poco cambio no se utiliza versionamiento como en el resto de colecciones. Asimismo, para las relaciones con las otras colecciones se utiliza una llave sustituta, a la cual se referirá como número de identificación tributaria (NIT).

Además, para mantener la consistencia se recomienda que en las entidades que referencien al nombre e identificación, en lugar de ambos atributos referencien a la tabla de personas con el NIT.

Se recomienda adicionalmente tener una tabla de conversión para las identificaciones, con la cual pueda realizarse el mapeo entre distintas identificaciones a un mismo NIT.

Al momento de la carga es importante considerar estas situaciones para implementar las transformaciones necesarias que permitan relacionar los datos de manera adecuada.

Nº	Identificación campo	Tipo	Descripción
1	NIT	INT	Número de identificación fiscal. Se utiliza como llave sustituta para enlazar las tablas relacionadas.
2	IDU	INT	Número interno que identifica, unívocamente en los sistemas internos de Hacienda, al contribuyente (Obligatorio)
3	TIPOIDENTIFICACION	INT	Tipo de identificación de la persona
4	IDENTIFICACIÓN	VARCHAR(30)	Identificación de la persona
5	NOMBRERAZONSOCIAL	VARCHAR(125)	Apellidos y nombre para personas físicas o razón social para entidades jurídicas. (Obligatorio)
6	PRIMERPELLIDO	VARCHAR(25)	Primer apellido para la persona física
7	SEGUNDOPELLIDO	VARCHAR(25)	Segundo apellido para la persona física
8	NOMBRE	VARCHAR(30)	Nombre de la persona física
9	ANAGRAMADEBUSQUEDA	VARCHAR(10)	Anagrama para procesos de búsqueda.
10	SEXO	INT	
11	FECHANACIMIENTOCONSTITUCION	INT	Fecha de nacimiento para persona física y fecha de constitución para entidades jurídicas, en formato Date: (DDMMAAAA).

Tabla 5: Personas

4.4.2.3 Registros de búsqueda

El contenedor de búsquedas almacena los listados que se van generando para cada uno de los filtros que va realizando el usuario. De esta forma, para efectos del prototipo almacena el identificador del usuario y hasta doce campos numéricos de monto para ingresar datos totalizados o cálculos de relaciones entre los mismos. De esta forma, los cálculos de determinado monto para un período se va almacenando en una consulta para que posteriormente esta tabla sea el conjunto reducido de búsqueda sobre el que se va trabajar.

N°	Identificación campo	Tipo	Descripción
1	NIT	INT	Número de identificación fiscal. Se utiliza como llave sustituta para enlazar las tablas relacionadas.
2	IDPROCESO	INT	Identifica a cual caso de estudio pertenece el registro.
3	USUARIO_BUSCA	VARCHAR(20)	Usuario que realiza la búsqueda
4	ACTIVO	CHAR(1)	Indicador de si la columna permanece activa durante las búsquedas y cálculos o debe ser ignorada.
5	MONTO_0	NUMERIC(17,2)	Columna monto 0
6	MONTO_1	NUMERIC(17,2)	Columna monto 1
7	MONTO_2	NUMERIC(17,2)	Columna monto 2
...	...	...	...
153	MONTO_148	NUMERIC(17,2)	Columna monto 148
154	MONTO_149	NUMERIC(17,2)	Columna monto 149

Tabla 6: Registro de búsqueda

4.4.2.4 Registro de columnas

Se realizan las operaciones sobre algunas columnas y se almacena en la tabla REG_COLUMNAS, aplica para los dos tipos de operaciones que existen donde hay campos específicos para cada uno.

N°	Identificación campo	Tipo	Descripción
1	ID_COLUMNAS	INT	Número de identificación de la operación
2	NOMBRE_COLUMNA	VARCHAR(80)	Nombre de la operación
3	IDPROCESO	INT	Identificación del caso asociado
4	IDTABLA	INT	Identificación de la tabla utilizada
5	IDPRIMERACOLUMNA	INT	Nombre de la primera columna, se utiliza para tipo de operación sobre una colección
6	IDSEGUNDACOLUMNA	INT	Nombre de la segunda columna, se utiliza para tipos de operación con columnas calculadas.
7	ID_OPERACION	NUMERIC(18,0)	Identificación de la operación realizada
8	FILTRO_ENTRADA	VARCHAR(1000)	Indica el filtro de entrada
9	FILTRO_RESULTADO	VARCHAR(1000)	Indica el filtro completo
10	USUARIO	VARCHAR(30)	Identificación del usuario que realizó la operación
11	ORDEN	NUMERIC(18,0)	Indica el orden de la operación
12	ACTIVO	CHAR(1)	Indica si esta activo o no
13	FECHA_INGRESO	DATE	Fecha de ingreso al sistema
14	IDNOMPRIMERACOLUMNA	NUMERIC(18,0)	Nombre de la primera columna, se utiliza para tipo de operación sobre columnas calculadas
15	IDNOMBSEGUNDACOLUMNA	NUMERIC(18,0)	Nombre de la segunda columna, se utiliza para tipo de operación sobre columnas calculadas
16	ID_DESTINO	NUMERIC(18,0)	Columna donde se almacena el resultado de la operación
17	ESTADOPROCESO	NUMERIC(18,0)	Indica el estado del proceso para saber si ya se ejecuto
18	MES	VARCHAR(2)	Contiene el mes
19	ANNO	VARCHAR(20)	Contiene el año

Tabla 7: Registro de columnas

4.4.2.5 Registro de proceso

Existe una tabla llamada REG_PROCESO donde se almacena la información de los casos, con una descripción y un nombre, además registra si esta activo o no.

N°	Identificación campo	Tipo	Descripción
1	IDPROCESO	INT	Número de identificación del caso
2	DESCRIPCION	VARCHAR(255)	Descripción del caso
3	COD_PERIODICIDAD	VARCHAR(1)	Código de periodicidad
4	FECHA_INICIO	DATE	Fecha de inicio del caso
5	USUARIO	VARCHAR(30)	Nombre del usuario que creó el caso
6	NOMBRE	VARCHAR(50)	Nombre del caso
7	ACTIVO	VARCHAR(1)	Indica si esta activo o no

Tabla 8: Registro de procesos

4.4.2.6 Expediente

Para cada persona, de acuerdo a la información almacenada en REPOSITORIO ÚNICO, se relacionan domicilios, declaraciones y otro tipo de información que puede ser enlazada a un contribuyente mediante el identificador fiscal. En el expediente se almacena la información importante del contribuyente y de la tabla relacionada.

4.4.2.6.1 Secciones

Se presenta un catálogo o una división del expediente por las secciones que contiene, donde se incluye una descripción y además indica si esa sección está activa, esto con el fin de agrupar la información.

N°	Identificación campo	Tipo	Descripción
1	IDSECCION	INT	Número de identificación de la sección
2	DESCRIPCION	VARCHAR(80)	Descripción de la sección
3	ORDEN	INT	Orden de despliegue en el expediente
4	ACTIVO	BIT	Indica si la sección está disponible

Tabla 9: Secciones de expedientes

4.4.2.6.2 Categorías

Otra de las divisiones, es por las categorías que contiene el expediente, el cual se relaciona con una sección y una tabla específica. Al igual que las secciones muestran el orden de despliegue e indican si se encuentran activas.

Nº	Identificación campo	Tipo	Descripción
1	IDCATEGORIA	INT	Número de identificación de la categoría
2	DESCRIPCION	VARCHAR(120)	Descripción de la categoría
3	IDSECCION	INT	Código de la sección relacionado
4	ORDEN	INT	Orden de despliegue entre el expediente
5	IDTABLA	INT	Orden de despliegue en el expediente
6	ACTIVA	BIT	Indica si la sección está disponible

Tabla 10: Categorías de expedientes

4.4.2.6.3 Detalle

En la tabla EXP_DETALLE se almacena la información del expediente, donde indican las columnas que intervienen, la categoría, el orden de despliegue e indica si está activo.

Nº	Identificación campo	Tipo	Descripción
1	IDDETALLE	INT	Número de identificación del detalle
2	IDCOLUMNADESCRIPCION	INT	Identificación de la tabla columnas para obtener la descripción
3	IDCOLUMNAAGRUPA	INT	Identificación de la tabla columnas para obtener si agrupa o no
4	IDCATEGORIA	INT	Identificación Categoría relacionada
5	ORDEN	INT	Indica el orden de despliegue
6	ACTIVA	BIT	Indica si esta activa

Tabla 11: Detalles de expediente

4.4.2.7 Otras colecciones

4.4.2.7.1 Tipos de operación

El contenedor almacena los tipos de operación en una tabla llamada TIPO_OPERACIÓN, donde registra la descripción y un texto descriptivo.

Nº	Identificación campo	Tipo	Descripción
1	ID_TIPO	NUMERIC(18,0)	Identificación del tipo de operación
2	DESCRIPCION	VARCHAR(255)	Descripción del tipo de operación
3	TEXTO_DESCRIPTIVO	VARCHAR(500)	Formato o texto descriptivo del tipo de operación

Tabla 12: Tipos de operación

4.4.2.7.2 Tipo de datos

En la tabla TIPO_DATO se almacenan los diferentes tipos de datos utilizados para las columnas de una tabla en específico, almacena un código consecutivo y una pequeña descripción.

Nº	Identificación campo	Tipo	Descripción
1	COD_TIPODATO	INT	Identificación del tipo de dato
2	DESCRIPCION	VARCHAR(30)	Descripción del tipo de dato

Tabla 13: Tipo de datos

4.4.2.7.3 Catálogo de operadores

Se almacenan los diferentes tipos de operadores a manera de catálogo con un nombre físico y uno lógico que se utiliza en las operaciones. Se tiene un código por cada tipo de operador, el 1 para los operadores lógicos y el 2 para los operadores de equivalencia.

Nº	Identificación campo	Tipo	Descripción
1	CodigoOperador	INT	Identificación del operador
2	NombreLogico	VARCHAR(20)	Nombre lógico del operador
3	NombreFísico	VARCHAR(30)	Nombre físico del operador
4	Tipo	INT	Tipo de operador, 1 lógico y 2 equivalencia

Tabla 14: Catálogo de operadores

4.4.2.7.4 Periodicidades

Se maneja un catálogo de periodicidades que contiene una identificación específica, una descripción, el orden en que se muestran y también el desplazamiento según los meses. Existen desde periodos diarios hasta periodos anuales, y cada uno se identifica por una letra significativa.

N°	Identificación campo	Tipo	Descripción
1	COD_PERIODICIDAD	INT	Identificación de la periodicidad
2	DESCRIPCION	VARCHAR(30)	Descripción del tipo de periodicidad
3	ORDEN	NUMERIC(18,0)	Orden en que se muestran las periodicidades
4	DESPLAZAMIENTO	NUMERIC(18,0)	Desplazamiento de las periodicidades según los meses

Tabla 15: Periodicidades

4.4.2.7.5 Filtros de los casos

Con el fin de realizar una búsqueda de casos de una forma detallada, se crean algunos filtros. Cada filtro se relaciona a un caso, y siempre lleva un operador lógico y uno de equivalencia, donde se especifica que valores se van a comparar. En el caso de que sea el último filtro sobre un caso en específico el operador lógico se omite.

N°	Identificación campo	Tipo	Descripción
1	idFiltro	INT	Identificación del filtro
2	IDPROCESO	INT	Identificación del caso al que se relaciona
3	CodigoOperadorLogico	INT	Código del operador lógico
4	CodigoOperadorComparacion	INT	Código del operador de equivalencia
5	idColumna	INT	Identificación de la columna que se va a comparar
6	VALOR_COMPARAR	VARCHAR(50)	Valor que se desea comparar
7	STRDESCRIPCION	VARCHAR(100)	Descripción de los filtros

Tabla 16: Filtros de los casos

4.4.3 Diseño de aplicación

4.4.3.1 Casos de estudio

4.4.3.1.1 Casos de estudio creados

El sistema muestra los casos de estudio creados en una tabla, los casos de estudio son el primer paso para realizar los procesos con las columnas de búsqueda y las colecciones. De esta forma, la información desplegada en pantalla de forma tabulada muestra una descripción del caso de estudio con el nombre respectivo, esto con el fin de identificarlo. Además, muestra la periodicidad sobre la que se va a trabajar, esto para que exista proporcionalidad entre los valores comparados, y por último muestra un estado para indicar si está activo o inactivo.

Si se quiere filtrar aún más los datos que se van a mostrar dentro de la tabla que se indicó anteriormente, se puede utilizar un combo que contiene los estados, tanto activo como inactivo y muestra los casos filtrados para proporcionar facilidad al buscar los datos específicos.

La tabla de los casos de estudio, contiene una paginación lo que facilita la búsqueda de casos, agrupa y muestra diez registros en pantalla y se puede navegar por las diferentes opciones. Los datos de la tabla son los siguientes:

N°	Nombre del campo	Descripción
1	Nombre	Nombre del caso
2	Periodicidad	Periodicidad sobre la que se va a trabajar
3	Descripción	Descripción del caso
4	Activo	Indica si está activo o no
5	Privado	Indica si es privado

Tabla 17: Casos estudio

Adicionalmente la tabla contiene dos opciones más que permitirá editar los casos o ejecutar y finalmente se cuenta con un botón “Nuevo” utilizado para ingresar un caso.

Seleccione el estado: Activo

	Nombre	Periodicidad	Descripción	Activo
	▶ Caso 59	Anual	Caso de estudio 59	Sí
	▶ Caso 58	Anual	Caso de estudio 58	Sí
	▶ Caso 57	Anual	Caso de estudio 57	Sí
	▶ Caso 56	Anual	Caso de estudio 56	Sí
	▶ Caso 55	Anual	Caso de estudio 55	Sí
	▶ Caso 54	Mensual	Caso de estudio 54	Sí
	▶ Caso 53	Trimestral	Caso de estudio 53	Sí
	▶ Caso 52	Anual	Caso de estudio 52	Sí
	▶ Caso 51	Semestral	Caso de estudio 51	Sí
	▶ Caso 50	Anual	Caso de estudio 50	Sí

« Ant | 1 | 2 | 3 | 4 | 5 | 6 | Sig »



Figura 13: Casos de estudio creados
Tomado de: Elaboración propia

Si se desea ingresar un nuevo caso de estudio o modificar uno existente se debe presionar “Nuevo” y en la imagen de “Editar” respectivamente y se muestra la página de Mantenimiento de casos.

En esta opción se puede ingresar o cambiar el contenido del caso de estudio, si se quiere modificar se cargan los datos automáticamente en los campos respectivos y se realizan los cambios permitiendo inactivar o activar los casos. Si se desea ingresar un nuevo caso, los campos permanecerán vacíos para poderlos ingresar.

Los campos solicitados en la pantalla son los siguientes:

Proceso de selección de candidatos

MODIFICAR CASO DE ESTUDIO

Nombre:

Seleccione la periodicidad:

Descripción del proceso:

Activo: ☒

Privado: ☐

.....

 Guardar
  Limpiar
  Regresar

Figura 14: Mantenimiento de casos de estudio
Tomado de: Elaboración propia

Si se presiona el botón “Regresar”, muestra la pantalla anterior que carga los casos de estudio creados, si se presiona “Limpiar”, se vacían todos los campos para poder llenarlo nuevamente.

Al presionar el botón guardar, antes de almacenar los datos nuevos se realiza una validación de datos para que cumplan con los tipos adecuados. Seguidamente pasa a la página de crear filtros al caso.

4.4.3.1.2 Agregar filtros al caso

El siguiente paso es agregar filtros al caso que se creó anteriormente. Se muestra a manera de consulta; el nombre del caso de estudio, la periodicidad, si está activo o inactivo y si es privado o no es privado.

Se debe seleccionar el filtro deseado y seguidamente la actividad. La actividad se divide en partes, primero se debe establecer el operador de equivalencia, segundo el campo que se desea comparar, según el filtro seleccionado y por último el operador lógico. Una vez completa la actividad, se presiona el botón para agregar el filtro y lo muestra en una tabla o lista de filtros, que contiene los siguientes datos:

N°	Nombre del campo	Descripción
1	Filtro	Muestra el nombre del filtro
2	Condición	Operador de equivalencia
3	Valor	Nombre del campo
4	Operador	Operador lógico

Tabla 18 Tabla de filtros de casos

Se pueden agregar varios filtros, y además se pueden eliminar presionando el botón que se encuentra en la tabla de lista de filtros para cada registro.

Caso de estudio: Caso 59

Periodicidad: Anual

Activo: Sí

Seleccione el filtro: Actividad

Seleccione la actividad: Igual ACTIVIDAD NO ESPECIFICADA O

Lista de filtros				
	Filtro	Condición	Valor	Operador
X	Administración	Diferente	Administración de C...	Y
X	Provincia	Diferente	ALA JUELA	Y

Continuar

Figura 15: Mantenimiento de filtros de casos
Tomado de: Elaboración propia

Al terminar de agregar los filtros a un caso en específico se presiona el botón de continuar para pasar a la página de definición de operaciones.

4.4.3.1.3 Definición de operaciones

El siguiente paso es la definición de operaciones, se muestra una tabla con la información de las operaciones sobre las columnas respectivas. La tabla contiene un número para definir el orden de despliegue, este número tiene un hipervínculo que permite modificar las operaciones sobre la columna.

Otro campo es el nombre de la columna que tiene las operaciones, seguido de una descripción antes mencionada, es un conjunto de datos importantes. Contiene el nombre de la operación, el año de referencia, la colección o nombre de la tabla, el atributo o columna, además del criterio de ordenación también por columna y por último los filtros correspondientes a cada columna.

El siguiente campo indica el estado de la operación, dependiendo de los colores que muestre, si es de color gris indica que la operación está sin ejecutar, celeste indica que se está ejecutando, verde indica que ya se ejecutó y rojo que no se

ejecutó. La tabla contiene una paginación y agrupa las operaciones por cada diez registros.

Se establecen los criterios para agregar los registros utilizados en el procesamiento

#	Columna	Descripción
1	Columna01	<ul style="list-style-type: none"> Operación: 'Incluir' Año referencia: Año -3 Colección: 'D101' Atributo: 'Gasto ventas' Agrupado por: 'Cantón' Filtro: <ul style="list-style-type: none"> - COLUMNA 2 Mayor 500000 Y - COLUMNA 2 Menor 1000000
2	Columna02	<ul style="list-style-type: none"> Operación: 'Incluir' Año referencia: Año -2 Colección: 'D101' Atributo: 'Monto' Agrupado por: 'Cédula' Filtro: <ul style="list-style-type: none"> - COLUMNA 3 Igual 1500000
3	Columna03	<ul style="list-style-type: none"> Operación: 'Incluir' Año referencia: Año 0 Colección: 'D101' Atributo: 'Crédito contrato producción' Agrupado por: 'Código declaración' Filtro: <ul style="list-style-type: none"> - COLUMNA 3 Diferente 6542374657 O - COLUMNA 4 Diferente 4364732423

« Ant | 1 | Sig »

 Nuevo
  Ver resultados
  Regresar

Figura 16: Definición de operaciones

Existen diferentes opciones en pantalla según lo que se desea hacer, la opción de “Regresar” para volver a la página anterior. Y por último está la opción “Nuevo” o se presiona el enlace del número de orden de despliegue de la operación y aparece la pantalla de mantenimiento de operaciones.

Dependiendo del tipo de operación que se indiqué se despliegan diferentes opciones. Si se selecciona el tipo de operación sobre una colección, mostrará las colecciones cuya periodicidad corresponda a la del caso, y mostrará los campos por los cuales se puede realizar la operación.

Si se selecciona el tipo de operación “Entre columnas calculadas” se muestra un combo para seleccionar la operación, dos combos para indicar las dos columnas o atributos que se desean aplicar en la operación.

Para finalizar se selecciona la opción de “Guardar”, almacena los cambios, y seguidamente los envía a la pantalla de definición de operaciones. Si no se desea guardar se puede limpiar la pantalla e incluir nuevos datos o regresar a la página anterior y cancelar la operación.

A continuación se muestra el flujo de casos de estudio desde la creación hasta la definición de operaciones:



Figura 17: Flujo de casos de estudio
Tomado de: Elaboración propia

4.4.3.2 Dicionarios

4.4.3.2.1 Mantenimiento de tablas

El diccionario de tablas almacena la información de las colecciones para mantener la relación entre los campos físicos, como están definidos en las tablas, y los campos lógicos, como se podrían mostrar en pantalla, en el expediente y en los controles de usuario que correspondan. Para realizar la inclusión de una nueva tabla o darle mantenimiento se ingresa a la pantalla de tablas creadas. Donde muestra en una tabla todas las colecciones que existen.

Para filtrar la información hay un combo que contiene los dos estados, activa e inactiva y al seleccionar alguno se realiza una búsqueda por el criterio indicado. La

tabla utiliza paginación para agrupar por cada diez registros. Y muestra los siguientes campos:

N°	Nombre del campo	Descripción
1	Nombre Físico	Nombre que se encuentra en la tabla
2	Nombre Lógico	Nombre que se utiliza para mostrar en pantalla, expediente o controles de usuario.
3	Activa	Indica el estado de la tabla, si está activa o inactiva
4	Periodicidad	Muestra la periodicidad
3	Muestra expediente	Indica si ese dato se va a mostrar en el expediente

Tabla 19 Diccionario de Tablas

La tabla contiene dos opciones que sirven de hipervínculo. La primera es para editar la tabla y la segunda para editar las columnas que pertenecen a una tabla específica.



Seleccione el estado:		Seleccione				
		Nombre Físico	Nombre Lógico	Activa	Periodicidad	Muestra expediente
		REG_BUSQUEDAS	Carrifo	Si	Anual	No
		REG_COLUMNAS	Columnas	No	No aplica	No
		CONTRIBUYENTES	Contribuyentes	Si	Anual	No
		D101	D101	Si	Anual	Si
		D150	D150	No	Anual	Si
		D151	D151	Si	Anual	Si
		D152	D152	No	Anual	Si
		D153	D153	No	Anual	Si
		D158	D158	No	Anual	Si
		D160	D160	No	Anual	Si

« Ant | 1 | 2 | 3 | Sig »

Nuevo

Figura 18: Tablas creadas
Tomado de: Elaboración propia

Si se presiona el hipervínculo de editar o la opción de “Nuevo” el sistema muestra la pantalla de mantenimiento de tablas, si es modificar carga los datos en los campos respectivos y si es nuevo limpia los campos.

Entre los campos que se encuentran en la pantalla está un cuadro de texto donde se ingresa el nombre físico y otro para el nombre lógico, estos campos son

obligatorios. Aparece un combo que muestra la periodicidad y dos casillas de verificación para indicar si está activo y si se muestra en el expediente. Al ingresar los datos, si se desea almacenar se presiona la opción “Guardar”, por el contrario si se desea cancelar se presiona la opción “Regresar” o “Limpiar” para borrar los datos incluidos (Véase figura 19)



Formulario de mantenimiento de tablas con los siguientes campos:

- Nombre físico: REG_BUSQUEDAS
- Nombre lógico: Carrito
- Seleccione la periodicidad: Anual
- Activo: ☒
- Muestra expediente: ☐

Botones: Guardar, Limpiar, Regresar

Figura 19: Mantenimiento de tablas
Tomado de: Elaboración propia

Al realizar la inserción o modificación de un registro se muestra el resultado en una página con todos los datos de la tabla, como se muestra en la figura 20.



Mantenimiento de tablas

III. RESULTADO

Nombre Físico:	REG_BUSQUEDAS
Nombre Lógico:	Carrito
Periodicidad:	Anual
Activa:	Sí
Muestra expediente:	No

Botón: Regresar

Figura 20: Resultado tablas
Tomado de: Elaboración propia

4.4.3.2.2 Mantenimiento de columnas

Si se presiona el hipervínculo de editar columnas aparece la pantalla con las columnas creadas para la tabla específica, que al igual que las tablas posee filtros para poder facilitar la búsqueda. Son tres filtros, uno para indicar si se agrega como

columna, uno para indicar si se muestra en el expediente y el último para establecer si está activo o inactivo. Al seleccionar la columna indicada se despliega la información filtrada en la tabla.

Los campos de la tabla donde aparece la información son los siguientes:

N°	Nombre del campo	Descripción
1	Nombre Físico	Nombre que se encuentra en la tabla
2	Nombre Lógico	Nombre que se utiliza para mostrar en pantalla, expediente o controles de usuario.
3	Nombre de la tabla	Indica el estado de la tabla, si está activa o inactiva
4	Tipo de dato	Muestra la periodicidad
3	Agrupar	Indica si ese dato se va a mostrar en el expediente
2	Filtrar	Indica si filtra algún resultado
3	Agrega como columna	Indica si se agrega como columna
4	Expediente	Indica si se muestra en el expediente
3	Activa	Indica si está activa o inactiva

Tabla 20 Diccionario de columnas

La tabla posee paginación donde agrupa por diez registros las columnas y cuenta con una opción para poder modificar las columnas (Ver figura 21).

Mantenimiento de columnas

I. COLUMNAS CREADAS

Agregar como columna:

Muestra en expediente:

Seleccione el estado:

Seleccione

Seleccione

Seleccione

	Nombre Físico	Nombre Lógico	Nombre Tabla	Tipo de dato	Agrupar	Filtrar	Agrega como columna	Expediente	Activa
	ACTIVO	Activo	Carrito	CHAR	No	No	Si	No	Si
	IDPROCESO	Código Proceso	Carrito	INT	No	No	Si	No	Si
	MONTO_0	COLUMNA 0	Carrito	NUMBER	No	No	Si	No	Si
	MONTO_1	COLUMNA 1	Carrito	NUMBER	No	No	Si	No	Si
	MONTO_10	COLUMNA 10	Carrito	NUMBER	No	No	Si	No	Si
	MONTO_100	COLUMNA 100	Carrito	NUMBER	No	No	Si	No	Si
	MONTO_101	COLUMNA 101	Carrito	NUMBER	No	No	Si	No	Si
	MONTO_102	COLUMNA 102	Carrito	NUMBER	No	No	Si	No	Si
	MONTO_103	COLUMNA 103	Carrito	NUMBER	No	No	Si	No	Si
	MONTO_104	COLUMNA 104	Carrito	NUMBER	No	No	Si	No	Si

« Ant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | Sig »

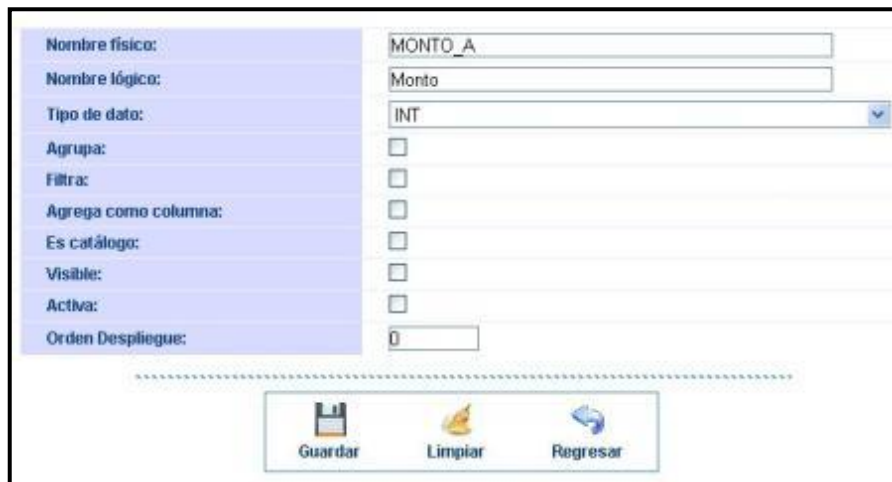
Nuevo

Regresar

Figura 21: Columnas creadas
Tomado de: Elaboración propia

Si se presiona la opción de “Nuevo” o el hipervínculo de modificar se despliega una página para el mantenimiento de columnas, donde se deberá ingresar la información solicitada. La página contiene un cuadro de texto para ingresar el nombre físico que se encuentra en la tabla y otro para el nombre lógico que es el que se muestra en pantalla, estos dos campos son de carácter obligatorio.

Además, se encuentra un combo para seleccionar el tipo de dato, y aparecen seis cuadros de verificación, para indicar si agrupa, si filtra alguna columna, el siguiente es para saber si agrega como columna, para indicar si es catálogo, si está visible y por último si está activa o inactiva. Por último se encuentra un cuadro de texto para agregar un número para el orden de despliegue, como se muestra en la figura 22.



El formulario de mantenimiento de columnas presenta los siguientes campos:

- Nombre físico:** Campo de texto con el valor "MONTO_A".
- Nombre lógico:** Campo de texto con el valor "Monto".
- Tipo de dato:** Menú desplegable con el valor "INT".
- Agrupar:** Casilla de verificación desactivada.
- Filtrar:** Casilla de verificación desactivada.
- Agrega como columna:** Casilla de verificación desactivada.
- Es catálogo:** Casilla de verificación desactivada.
- Visible:** Casilla de verificación desactivada.
- Activa:** Casilla de verificación desactivada.
- Orden Despliegue:** Campo de texto con el valor "0".

En la parte inferior del formulario hay tres botones: "Guardar" (icono de disco), "Limpiar" (icono de pila) y "Regresar" (icono de flecha curva).

Figura 22: Mantenimiento de columnas
Tomado de: Elaboración propia

Una vez que se llenan o cambian los datos se presiona “Guardar” y almacena los nuevos datos. Si se desea cancelar la operación se presiona “Regresar” o “Limpiar” para vaciar los campos y poder seleccionar nuevamente. Al guardar los datos se muestra una pantalla de resultado en modo de lectura para informar cuales fueron los nuevos valores (ver figura 23).

Nombre Tabla:	Carrito
Nombre Físico:	MONTO_A
Nombre Lógico:	Monto
Tipo de dato:	NUMBER
Agrupar:	No
Filtrar:	No
Agrega como columna:	Sí
Es catálogo:	No
Muestra en expediente:	No
Activa:	Sí
Orden de despliegue:	0



 Regresar

Figura 23: Resultado de columnas

En la figura 24 puede observarse el flujo de los diccionarios de columnas y tablas.

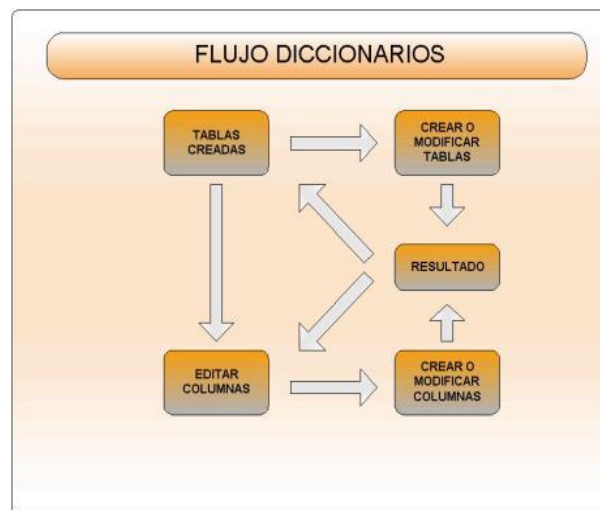


Figura 24: Flujo de diccionarios
Tomado de: Elaboración propia

4.4.3.3 Ejecución de operaciones

4.4.3.3.1 Casos de estudio creados

Se muestra los casos de estudio creados en una tabla que contiene la información importante del caso, como lo es, el nombre del caso, la periodicidad, la descripción y además indica si está activo. Además, cuenta con un filtro por estado para que muestre los casos que estas activos o inactivos. Y en la tabla existe otra opción para ejecutar el caso de estudio especificado (Ver figura 25).



Nombre	Periodicidad	Descripción	Activo
▶ Caso 66	Anual	Caso de estudio 66	Sí
▶ Caso 65	Trimestral	Caso de estudio 65	Sí
▶ Caso 64	Semestral	Caso de estudio 64	Sí
▶ Caso 63	Semestral	Caso de estudio 63	Sí
▶ Caso 62	Anual	Caso de estudio 62	Sí
▶ Caso 61	Semestral	Caso de estudio 61	Sí
▶ Caso 60	Diaria	Caso de estudio 60	Sí
▶ caso-prueba	Semestral	Prueba casos de estu...	Sí
▶ caso-prueba	Semestral	Prueba casos de estu...	Sí
▶ Caso 59	Anual	Caso de estudio 59	Sí

Figura 25: Casos de estudio para ejecución
Tomado de: Elaboración propia

Al presionar el botón de ejecutar sobre algún registro se carga la página de ejecución donde se muestra el detalle del caso de estudio seleccionado, muestra el nombre del caso, la descripción, la periodicidad y también indica si está activo o inactivo.

Seguidamente se muestra el detalle de las operaciones que tiene el caso de estudio en una tabla con el nombre de la operación y la descripción, la cual está compuesta por varias definiciones, entre ellas está, la operación que realiza, el año

de referencia, la colección o tabla que afecta, el atributo o columna específica, la columna por la que se quiere agrupar y los filtros que tiene la operación.

Por último se debe seleccionar el año de referencia que se desea aplicar en la operación y presionar el botón de ejecutar. (Véase figura 26)

Casos de estudio
II. EJECUTAR OPERACIONES

Nombre:	Caso 61
Periodicidad:	Semestral
Descripción:	Caso de estudio 61
Activo:	Sí

Columna	Descripción
Columna1	<ul style="list-style-type: none"> Operación: 'Incluir' Año referencia: Año 0 Colección: 'D101' Atributo: 'Depreciación' Agrupado por: 'Compras' Filtro: - No posee filtros.

Año de referencia: 2012

Tamaño máximo de población:

Ejecutar Regresar

Figura 26: Ejecución de operaciones
Tomado de: Elaboración propia

Al presionar ejecutar se realiza una búsqueda de las personas que cumplan con los filtros que contiene ese caso, y si es mayor a mil registros lanza un mensaje de error indicando que no se puede ejecutar la operación. Si por el contrario, la cantidad de personas es menor a mil, los registros son ingresados en el “Carrito de casos” y se despliega una muestra de los resultados.

4.4.3.4 Búsqueda de personas

4.4.3.4.1 Búsqueda por número de identificación

La pantalla permite realizar una búsqueda de personas físicas o jurídicas según el número de identificación o el nombre de la persona. Si se selecciona la búsqueda por identificación se debe ingresar el número de cédula en un cuadro de texto especificado que permite caracteres alfanuméricos y guiones.

Sin embargo, el campo no se habilita hasta que se seleccione el tipo de persona, ya sea jurídica o física como se indicó anteriormente. Los campos para ingresar el nombre con apellidos incluidos se deshabilitan. De este modo, al ingresar el número de cédula se presiona el botón consultar y se realiza la búsqueda donde se muestra el resultado en una tabla con la información de la persona que se detallará más adelante. Finalmente, si no existen registros que coincidan con los parámetros indicados se muestra un mensaje de información.

4.4.3.4.2 Búsqueda por nombre y apellidos

Si se selecciona la búsqueda por nombre y apellidos se deshabilita el cuadro de texto de la identificación, a continuación se debe indicar el tipo de persona. Del mismo modo, si es de tipo jurídica se habilita el campo de texto para el nombre y se deshabilitan los campos de apellidos, si por el contrario es de tipo física se habilitan todos los campos. Y es obligatorio el nombre de la persona con al menos un apellido para realizar la búsqueda, si no cumple con lo indicado se muestra un mensaje de error.

Estos campos están creados de manera que solo acepte letras y espacios. Al presionar el botón de consultar se muestra el resultado en una tabla de lo contrario se especifica que no hay registros que coincidan con los parámetros establecidos.

4.4.3.4.3 Resultado de la búsqueda

Al realizar la búsqueda ya sea por número de identificación o por el nombre con los apellidos se despliega el resultado en una tabla donde muestra el detalle de la persona. Los campos que se muestran en la tabla son los siguientes:

N°	Nombre del campo	Descripción
1	Tipo de identificación	Tipo de identificación de la persona
2	Identificación	Número de identificación de la persona
3	Nombre	Nombre de la persona con apellidos

Tabla 21 Resultado de búsqueda

Además de estos datos, en la tabla se encuentra un botón para cada registro que al presionarlo aparece el detalle del expediente de dicha persona. (Ver figura 27)

Realizar la búsqueda por:

☐ Identificación.

☒ Nombre.

Tipo de persona:

☒ Física.

☐ Jurídica.

Búsqueda por el nombre

Búsqueda por el número de identificación

Identificación:

Nombre:

Primer apellido:

Segundo apellido:

Tipo identificación	Identificación	Nombre
FÍSICA	<input type="text"/>	ARLEY MIGUEL <input type="button" value="Detalle"/>
FÍSICA	<input type="text"/>	ARLEY MIGUEL <input type="button" value="Detalle"/>

< Ant | 1 | Sig >

Figura 27: Búsqueda de personas
Tomado de: Elaboración propia

A partir de los resultados generados es posible seleccionar una persona para ver los datos relacionados con el cliente. Una vez seleccionados los mismos podría presentarse una pantalla en la que se consolide la información personal de la

persona, con datos como direcciones, correos, declaraciones asociadas y otros, como se muestra en la figura 28.

The screenshot shows a web interface titled "Proceso de selección de candidatos". Below the title is a section "DATOS DEL CONTRIBUYENTE". It contains two input fields: "Identificador fiscal:" with the value "50600" and "Nombre:" with a redacted black bar. Below this is a section "Seleccione para mostrar el detalle:" which contains four expandable sections: "Domicilios:" with a link "Lista domicilios registrados", "Teléfonos:" with a link "Teléfonos del contribuyente", "Actividades económicas:" with a link "Lista de actividades económicas", and "Declaraciones:" with links "D101" and "D151".

Figura 28: Datos del contribuyente
Tomado de: Elaboración propia

En cada uno de las secciones sería posible hacer clic ya sea para expandir a mayor detalle o para ir a la información de los datos de la persona como puede observarse en la figura 29.

This screenshot shows the "Seleccione para mostrar el detalle:" section expanded. It lists the same categories as Figure 28: "Domicilios:", "Teléfonos:", "Actividades económicas:", and "Declaraciones:". The "Declaraciones:" section is expanded to show a list of years: 2012, 2011, 2010, 2009 (which is highlighted with a blue border), and 2008.

Figura 29: Desglose de datos del contribuyente

Al seleccionar alguna de las propiedades del expediente de la persona, por ejemplo la información de una de las declaraciones, se presenta el detalle de los campos, según estén definidos en el diccionario correspondiente (Ver figura 30).

Proceso de selección de candidatos

DETALLE DE LA DECLARACIÓN RENTA

Identificación:

Nombre:

THE VERSE CREATIONS LLC DBA VERSE

Ingresos gravables	868995
Ingresos no gravables	0
Ingresos totales	868994
Total costos gastos	434492
Ventas	0.00
Depresiación	0.00
Total de costos	217250.00
Impuesto bruto	0.00
Impuesto neto	65175.00

Regresar

Figura 30: Detalle de los datos de la persona
Tomado de: Elaboración propia

Capítulo V. Conclusiones y recomendaciones

5.1 Conclusiones

- Se definieron como formularios de ejemplo los relacionados a venta y renta, por cuanto estos representan declaraciones con diferentes niveles de periodicidad y con el fin de acotar el ámbito de estudio entre todos los tipos existentes.
- De acuerdo con las fuentes disponibles, como el CIAT, se identificaron los procesos que forman parte de una estrategia de inteligencia tributaria. De esta manera, se definen elementos tanto técnicos como operacionales, requeridos para aplicar la estrategia y que la misma sea de utilidad no solo a usuarios técnicos, sino también a la contraparte funcional del sistema.
- Se preparó un prototipo como laboratorio para el modelo que se propone. De este modo, se pudieron ejecutar pruebas con casos de estudio utilizando los modelos propuestos, las estructuras de datos, los procesos de carga al repositorio unificado y la extracción y visualización para los analistas. Lo anterior permitió validar las estructuras y datos que componen el modelo.
- Con la información recopilada en cuanto a estructura de los formularios de renta y venta, se crearon las estructuras de datos que permitan representar, en un repositorio unificado, los tipos de declaración utilizados en los laboratorios de forma tabular. De esta manera, se tiene una estructura plana de un formulario de declaración por niveles, con los valores calculados y los datos centralizados para los analistas y sus herramientas.
- De acuerdo con la información recopilada en los documentos del CIAT, artículos en Internet y entrevistas con analistas, se elaboró una propuesta base para una estrategia de Inteligencia Tributaria. Inicialmente, para la

investigación, se consideraron declaraciones de renta y venta. No obstante, el modelo puede ser extendido a otros tipos de declaraciones.

- Se investigó el modelo de declaración con información disponible en las fuentes públicas acerca de la forma actual en que son presentadas las declaraciones de renta y ventas en los últimos años, así como de su administración. Es así como la presentación de las declaraciones en papel genera mayor trabajo a nivel de digitadores, así como de calidad de los datos. De la misma forma, la automatización y la recepción electrónica de las declaraciones es una oportunidad que puede ser aprovechada para mejorar los mecanismos de control tributario.
- Con la información recopilada acerca de la situación actual de la presentación y manejo de declaraciones, las recomendaciones de instituciones como el CIAT en sus trabajos interregionales, las características de las declaraciones tributarias y las experiencias de analistas en proyectos con usuarios de inteligencia, se diseñó la base de una estrategia de Inteligencia Tributaria aplicable al Ministerio de Hacienda. La aplicación de una estrategia de este tipo debe considerar las características propias del Ministerio de Hacienda y las necesidades de sus usuarios, de forma que pueda ser adaptada y aplicada. Asimismo, esta es una base que puede ser extendida para considerar distintos elementos de Inteligencia que la puedan enriquecer y fortalecer.

5.2 Recomendaciones

- La variedad de impuestos y formularios disponibles hacen necesario delimitar los objetos de estudio e ir incorporando paulatinamente nuevos formularios con sus cálculos asociados dentro de los elementos involucrados en la estrategia.
- Los procesos de Inteligencia son muy similares en las distintas agencias tributarias de diferentes países. Por lo tanto, un inicio en la resolución de la estrategia es valorar experiencia y recomendaciones de organizaciones regionales como CIAT. Una vez con estos elementos recopilados se debe buscar la forma en que estos se relacionan con la realidad propia del Ministerio de Hacienda.
- Los prototipos facilitan la retroalimentación entre usuario y desarrollador para verificar que están de acuerdo en la parte conceptual de la herramienta. Se recomienda establecer a nivel de usuarios funcionales, un equipo para que valore y plantee los requerimientos por parte del área de inteligencia al área de tecnología. De la misma forma, en el área técnica es necesario conformar un equipo para filtrar y atender estos requerimientos. De esta manera, la Inteligencia Tributaria pasa a ser un objetivo conjunto entre usuarios funcionales y técnicos, de forma que ambos se sientan corresponsables y den el impulso necesario para salir adelante.
- Al validar los modelos de datos, la creación de las estructuras debe adaptarse no solamente a criterios meramente técnicos, sino también funcionales, que permitan a los analistas no técnicos explorar los datos de forma ágil y sencilla, y sin necesidad de un alto nivel técnico acerca de las herramientas. El repositorio unificado es un elemento de gran relevancia en el proceso porque

permite la estandarización de los datos y se tiene un punto único de acceso para las consultas.

- Para la definición de una estrategia de inteligencia la experiencia de personas e instituciones involucradas en procesos similares es importante para determinar cuáles pasos han sido exitosos y cuales han fracasado. Asimismo, en este proceso el involucramiento de los usuarios funcionales en el proceso es fundamental para impulsarlo como un proyecto en conjunto de la institución y no como un proyecto meramente tecnológico. Las recomendaciones de expertos y usuarios que han participado en proyectos de análisis de datos, así como las recomendaciones de los organismos especializados, son insumos importantes para poder elaborar una propuesta base a partir de los elementos que deben incorporarse.
- Es necesario extender la obligatoriedad de la presentación de declaraciones por medios electrónicos, de forma que se tenga de forma automatizada la recepción y validación de cada declaración en el momento de ser presentada. Además, a nivel institucional, se requieren protocolos para la generación de datos, así políticas de comunicación calidad y validación de los datos que serán enviados al Ministerio de Hacienda, con el fin de tener a disposición datos completos y útiles en el proceso de inteligencia.
- La definición de una estrategia para elaborar un sistema de inteligencia tributaria requiere un estudio progresivo de los procesos existentes para la captura de datos, así como el seguimiento y análisis de datos realizado por los usuarios funcionales. Lo anterior, por cuanto este debe ser un proceso integral que involucre tanto aspectos tecnológicos como los operacionales seguidos por los analistas para la detección de casos sospechosos. A partir de esta base ir incrementando las opciones disponibles y agregando nuevas funcionalidades.

Referencias

- Bornak, K. (18 de Octubre de 2012). *What is Ad Hoc Reporting?* Obtenido de JReport: <http://blog.jinfony.com/what-is-ad-hoc-reporting/>
- Camacho, C. (12 de 09 de 2011). *Inteligencia fiscal*. Obtenido de La Nación: <http://www.nacion.com/2011-09-12/Economia/inteligencia-fiscal.aspx>
- Carrasquilla, U. (2010). Recuperado el 12 de Marzo de 2012, de Benchmarking Algorithms for detecting anomalies in large Datasets: <http://cala.net/publications/anomalies.pdf>
- Castillo, D. (20 de diciembre de 2012). *BCCR estima un déficit fiscal de 5% para el 2013*. Recuperado el 06 de enero de 2013, de CRHOY: <http://www.crhoy.com/bccr-estima-un-deficit-fiscal-de-5-para-el-2013>
- CentralAmericaData.com. (15 de agosto de 2011). *Costa Rica: Licitan sistema de inteligencia tributaria*. Recuperado el 10 de julio de 2012, de CentralAmericaData: http://www.centralamericadata.com/es/article/home/Costa_Rica_Licitan_consultora_en_informtica
- Chimphlee, W. (2005). Recuperado el 01 de Marzo de 2012, de Unsupervised Anomaly Detection with Unlabeled Data Using Clustering.: http://eprints.utm.my/3356/1/Mohd_Noor_-_Unsupervised_Anomaly_Detection_with_Unlabeled_Data.pdf
- CIAT. (2006). *Manual CIAT sobre Inteligencia Fiscal*. Obtenido de Centro Interamericano de Administraciones Tributarias: http://www.ciat.org/index.php?option=com_content&task=view&id=150&Itemid=213

- Curto, J. (19 de Noviembre de 2007). *Diseño de un data warehouse: estrella y copo de nieve*. Recuperado el 01 de Marzo de 2013, de Information Management: <http://informationmanagement.wordpress.com/2007/11/19/disenio-de-un-data-warehouse-estrella-y-copo-de-nieve/>
- DGT. (2010). *Dirección General de Tributación*. Obtenido de Instructivo para el llenado del formulario D-101: <http://dgt.hacienda.go.cr/infotributaria/Tipos%20de%20impuestos/Brochoure%20N%206%20Instructivo%20formulario%20impuesto%20sobre%20la%20renta%20D-101-2010.doc>
- Eckerson, W. (2012). *The Secrets of Self-Service BI*. Obtenido de Blog: Wayne Eckerson: http://www.b-eye-network.com/blogs/eckerson/archives/2011/01/the_secrets_of.php
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2006). *Metodología de la investigación* (Cuarta Edición ed.). México: McGraw-Hill.
- Leitón, P. (28 de julio de 2012). *Más declaraciones de impuestos migran a Internet*. Recuperado el 01 de diciembre de 2013, de La Nacion: <http://www.nacion.com/2012-07-28/Economia/Mas-declaraciones-de-impuestos-migran-a-Internet.aspx>
- Leitón, P. (29 de enero de 2013). *Nuevo plan fiscal incluirá ingreso, gasto y deuda*. Recuperado el 29 de enero de 2013, de La Nacion: <http://www.nacion.com/2013-01-29/Economia/nuevo-plan--fiscal-incluirea--ingreso--gasto-y-deuda.aspx>
- Margineantu, D., Bay, S., Chan, P., & Lane, T. (21 de Agosto de 2005). Recuperado el 26 de Marzo de 2012, de International Workshop on Data Mining Methods for Anomaly Detection.: <http://www.dmarginantu.net/AD-KDD05/DMMAD-2005.WorkshopNotes.pdf>

Ministerio de Hacienda. (Setiembre de 2009). *Plan Nacional de prevención y lucha contra el fraude fiscal*. Recuperado el 02 de Marzo de 2013, de Ministerio de Hacienda:

<http://www.hacienda.go.cr/cifh/sidovih/uploads/Publicacion/Plan%20Nacional%20de%20Prevenci%C3%B3n%20y%20Lucha%20contra%20el%20Fraude%20Fiscal.pdf>

Ministerio de Hacienda. (12 de julio de 2010). *Misión, Visión y Valores del Ministerio de Hacienda*. Obtenido de Ministerio de Hacienda: <https://www.hacienda.go.cr/msib21/Templates/InformaciondelMinisterioContentTemplate.aspx?NRMODE=Published&NRORIGINALURL=%2fMsib21%2fEspa%C3%B1ol%2fAcerca%20del%20Ministerio%2fmision%2by%2bvision%2bmh.htm&NRNODEGUID={A94152ED-1EC2-4C5D-B148-382BD1AD8A30}&NRCACH>

Münz, G., Li, S., & Carle, G. (s.f.). Recuperado el 01 de Febrero de 2012, de Traffic Anomaly Detection Using K-Means Clustering: <http://www.net.in.tum.de/fileadmin/TUM/members/muenz/documents/muenz07k-means.pdf>

Olmedo, Y. (2012). *ColumnStore Index–Índices columnares en SQL*. Obtenido de Blogs Solidq: <http://blogs.solidq.com/bicorner/post.aspx?id=136&title=columnstore+index%E2%80%93%C3%8Dndices+columnares+en+sql>

Rodríguez, O. (27 de Febrero de 2013). *La Nación*. Obtenido de Tributación fortalece su sistema de pago en línea: http://www.nacion.com/archivo/Tributacion-fortalece-sistema-pago-linea_0_1326267463.html

Ruiz Ramón, G. (04 de diciembre de 2011). *Tributación Digital quedó en limbo*. Recuperado el 15 de marzo de 2012, de El Financiero: http://www.elfinancierocr.com/ef_archivo/2011/diciembre/04/economia2984372.html

- Servicio Nacional Integrado de Administración Aduanera y Tributaria. (19 de Julio de 2007). *Diseño de herramientas informáticas de apoyo a la comprobación y a la auditoría*. Recuperado el 01 de Marzo de 2013, de Eurosociat: http://www.eurosociatfiscal.org/uploads/documentos/20070719_110717_Herramientas_de_apoyo_a_la_comprobacion_y_auditoria_-_VENEZUELA.ppt
- Siu, M. (10 de julio de 2012). *Evasión en impuesto de ventas y renta ronda los 1.200 billones de colones*. Recuperado el 20 de diciembre de 2012, de CRHOY: <http://www.crhoy.com/evasion-en-impuesto-de-ventas-y-renta-ronda-los-1-200-billones-de-colones/>
- Tamberi, F. (26 de Junio de 2007). Recuperado el 05 de Febrero de 2012, de Anomaly Detection Data Mining Techniques: <http://www.cli.di.unipi.it/~tamberi/old/docs/tdm/anomaly-detection.pdf>
- Thiprungsri, S., & Vasarhelyi, M. (Julio de 2011). *Universidad de Huelva*. Recuperado el 28 de Marzo de 2012, de Cluster Analysis for Anomaly Detection in Accounting Data: An Audit Approach: http://www.uhu.es/ijdar/10.4192/1577-8517-v11_4
- Tomar, A. S. (2008). *ETC Infotech*. Recuperado el 28 de Marzo de 2012, de Tax Decision Support System: <http://etcinfotech.com/phocadownload/taxdssopen.pdf>
- Vitt, L., & Cameron, S. (Mayo de 2012). *Elegir una experiencia de modelado tabular o multidimensional en SQL Server 2012 Analysis Services*. Obtenido de Microsoft: <http://msdn.microsoft.com/es-es/library/jj695860.aspx>

Apéndices

A.1 Entrevista

Durante el proceso de trabajo de campo se definió un cuestionario que permitiera conocer los criterios que tienen distintos profesionales en cuanto a proyectos de inteligencia de acuerdo a sus experiencias. Para lo anterior se entrevistó de manera informal a analistas que han estado involucrados en diferentes proyectos de inteligencia. A continuación se presentan las preguntas y sus resultados.

1. ¿Cuál considera que es el principal elemento tecnológico para el inicio de un sistema de inteligencia?

- a. *La base de datos*
- b. *Herramientas de consulta*
- c. *Algoritmos de análisis*
- d. *Otro*

Respuestas:

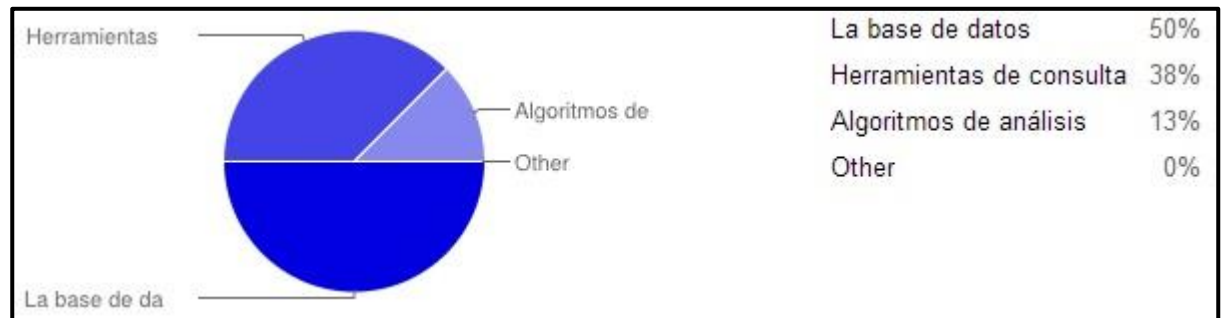


Figura 31: Principal elemento de un sistema de Inteligencia
Tomado de: Elaboración propia

Entre los consultados se ha podido observar como la mitad considera que la estructura de la base de datos es un elemento de relevancia para el inicio de procesos de inteligencia. Y un 13% indica que los algoritmos son la base del proceso.

De acuerdo con lo anterior, la experiencia de los involucrados señala la importancia de definir los modelos de datos para poder continuar con los procesos de inteligencia posteriores.

2. ¿Cuál fue el principal factor para el fracaso de proyectos de inteligencia?

- a. *Complejidad de las herramientas para los usuarios*
- b. *Limitaciones tecnológicas*
- c. *Resistencia a nuevos sistemas.*
- d. *Poco involucramiento de usuarios funcionales*
- e. *Otro*

Respuestas:

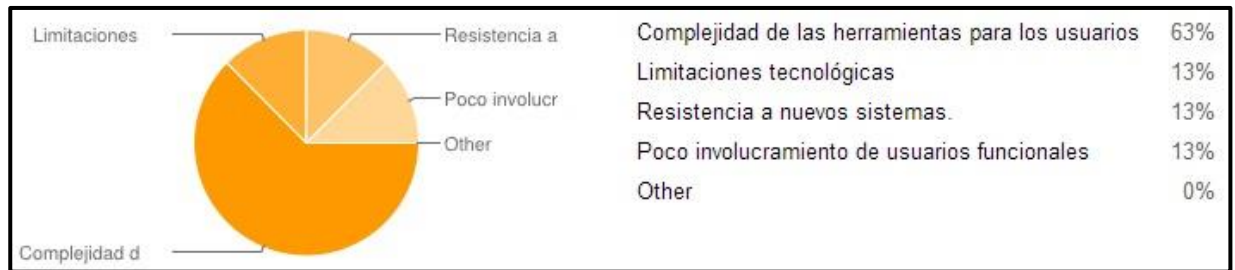


Figura 32: Principal factor de fracaso
Tomado de: Elaboración propia

Al consultar a los analistas acerca de las razones por las cuales los proyectos de inteligencia han fracasado a nivel de los usuarios, la mayoría coincide en que la complejidad para los usuarios en el uso de las herramientas ha dificultado su asimilación. Esto dado que algunas de las estructuras o conceptos presentados en estas herramientas no son fácilmente asociados con elementos que sean naturales de acuerdo a sus funciones, lo cual complica su asimilación.

3. ¿Cuál es el principal inconveniente al cargar datos de fuentes externas?

- a. *Calidad de los datos*
- b. *Formato de los archivos*
- c. *Protocolos de comunicación*



Figura 33: Principal inconveniente en la carga de datos
Tomado de: Elaboración propia

Respuestas:

Los procesos de inteligencia involucran la carga de datos de distintas fuentes para establecer una base de datos única para el análisis. Se consultó acerca de cuál es el principal inconveniente al momento de realizar procesos de extracción transformación y carga. El 75% coincidió en que la calidad de los datos es un elemento importante. En algunos casos, la calidad de los datos ha provocado que no se pueda cargar de una fuente o los datos obtenidos sean desechados.

4. ¿Cuáles son los principales inconvenientes al realizar los cruces de información?

- a. *Datos incompletos*
- b. *Calidad de los datos*
- c. *Complejidad de la extracción de datos por parte usuarios*

Respuestas:



Figura 34: Inconveniente al realizar las cargas
Tomado de: Elaboración propia

Al momento de consultar acerca de la principal dificultad al realizar cruces entre datos, la mitad indicó que no tener los datos completos afectaba la operación debido a que los cruces podían dar resultados sesgados. Por otra parte, la otra mitad indicó que la complejidad de la obtención de datos de las fuentes por parte de los usuarios y determinar cómo realizar los cruces son limitantes a la hora del diseño y ejecución de los cruces y análisis.

5. De acuerdo a su experiencia en proyectos anteriores. ¿Cuál modelo prefiere para el desarrollo de las aplicaciones de inteligencia en su empresa?

- a. *Desarrollo con recursos internos*
- b. *Desarrollo con recursos externos*
- c. *Desarrollo mixto (recursos internos y externos)*

Respuesta:



Figura 35: Modelo de desarrollo preferido
Tomado de: Elaboración propia

En cuanto al desarrollo de proyectos de inteligencia, la mitad de los entrevistados señalan la preferencia por un proceso mixto. Es decir, involucrar a recursos internos que puedan asimilar el conocimiento y consejos de consultores externos especializados. Esto para generar un balance entre el conocimiento del negocio de los colaboradores internos con el conocimiento técnico de los consultores externos para compartir el conocimiento.

6. ¿Qué herramienta es la más utilizada por los analistas?

- a. Herramientas a la medida
- b. Hojas de cálculo
- c. Otras herramientas de terceros

Respuesta:



Figura 36: Herramienta más utilizada por los analistas
Tomado de: Elaboración propia

Se preguntó, de acuerdo a la experiencia en sus labores, cuál herramienta es la más usada por parte de los usuarios. El 75% indicó que las hojas de cálculo. Esto debido a que los analistas han desarrollado experticia en el uso y explotación de datos a través de las hojas de cálculo, por lo que su uso se extendió y muchos de ellos las prefieren frente a herramientas más sofisticadas de inteligencia que les requieren mayor tiempo de entrenamiento y asimilación.

Figura 37: Formulario D101 (Renta)
Tomado de: <http://costarica.eregulations.org>

A.3 Formulario venta

ESTA FORMA NO USA PAPEL CARBON-ESCRIBA CON FIRMEZA

D. 104

REPÚBLICA DE COSTA RICA
MINISTERIO DE HACIENDA
DIRECCIÓN GENERAL DE TRIBUTACIÓN

DECLARACIÓN JURADA DEL IMPUESTO GENERAL SOBRE LAS VENTAS - LIQUIDACIÓN POR EL SISTEMA TRADICIONAL

IDENTIFICACIÓN	ADHIERA ETIQUETA		DECLARACIÓN NÚMERO		1	
	COMPLETE LAS CASILLAS RESPECTIVAS SI NO TIENE ETIQUETA		PERÍODO		2	
			CÓDIGO ADMINISTRACIÓN		3	
			NÚMERO DECLARACIÓN QUE CORRIGE		5	

CEDULA 4

CONTIENE SU NÚMERO, TAL Y COMO APARECE EN SU CédULA DE PERSONA FÍSICA O JURÍDICA Y SI ES EXTRANJERO CONTIENE EL NÚMERO ASIGNADO POR LA ADMINISTRACIÓN TRIBUTARIA.

6 **NOMBRE, RAZA N O DENOMINACIÓN SOCIAL COMPLETA**
SI ES SOCIEDAD O APELLIDOS Y NOMBRES SI ES PERSONA FÍSICA

No escriba sobre los espacios sombreados.
No incluya decimales en ninguno de los montos.

I. VENTAS DEL PERÍODO			
VENTAS POR EXPORTACIONES	20		
VENTAS EXENTAS Y AUTORIZADAS SIN IMPUESTO	21		
VENTAS GRAVADAS	22		
OTROS RUBROS A INCLUIR EN LA BASE IMPONIBLE	23		
BASE IMPONIBLE (SUME RENGLOS 22 Y 23)		24	
II. COMPRAS E IMPORTACIONES			
IMPORTACIONES EXENTAS Y AUTORIZADAS SIN IMPUESTO	25		
IMPORTACIONES GRAVADAS	26		
COMPRAS Y SERVICIOS EXENTOS Y AUTORIZADOS SIN IMPUESTO (NACIONALES)	27		
COMPRAS Y SERVICIOS GRAVADOS (NACIONALES)	28		
III. DETERMINACIÓN DEL IMPUESTO			
IMPUESTO GENERADO POR OPERACIONES GRAVADAS (RENGLO 24 POR TARIFA DE IMPUESTO)	29		
CREDITO POR IMPORTACIONES	30		
CREDITO POR COMPRAS Y SERVICIOS NACIONALES	31		
TOTAL CREDITOS (RENGLO 30 MAS RENGLO 31)		32	
IMPUESTO NETO DEL PERÍODO (SI RENGLO 29 ES MAYOR QUE RENGLO 32 ANOTE DIFERENCIA)	33		
SALDO A FAVOR DE ESTE PERÍODO (SI RENGLO 29 ES MENOR QUE RENGLO 32 ANOTE DIFERENCIA)	34		
SALDO A FAVOR DE PERÍODOS ANTERIORES	35		
SUBTOTAL DEL IMPUESTO		37	
MENOS RETENCIONES PAGO A CUENTA (LEY DE CONTINGENCIA FISCAL)		38	
IMPUESTO DEL PERÍODO		40	
IV. LIQUIDACIÓN DEUDA TRIBUTARIA			
Nº SOLICITUD DE COMPENSACIÓN	75		MONTO PENDIENTE DE APROBACIÓN
Nº SOLICITUD DE COMPENSACIÓN	77		MONTO PENDIENTE DE APROBACIÓN
IMPUESTO A PAGAR		79	
INTERESES		82	
TOTAL DEUDA TRIBUTARIA		83	

CONTRIBUYENTE O REPRESENTANTE LEGAL		FECHA DE RECEPCIÓN		TRÁMITE DE PAGO	
1er APELLIDO		88		89 C	
2do APELLIDO				90 C	
NOMBRES				91 C	
NÚMERO DE CEDULA				92 C	
FIRMA		FIRMA Y SELLO		93 C	
				94 C	
				95 C	
				96 C	
				97 C	
				98 C	
				99 C	

ORIGINAL - ADMINISTRACIÓN 1era COPIA - CONTRIBUYENTE 2da COPIA - BANCO

FOLIO 99

Figura 38: Formulario D104 (Ventas)
Tomado de: <http://dgt.hacienda.go.cr/>