



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

**ANÁLISIS Y DISEÑO DE UNA BASE DE DATOS MULTIDIMENSIONAL EN
APOYO A LA COORDINACIÓN NACIONAL DE SEGEPLAN DEL PROYECTO
DE INVENTARIO DEL AGUA DE GUATEMALA**

Andres Saenz Aguirre

Asesorado por el Ing. Jorge Armín Mazariegos Rabanales

Guatemala, julio de 2011

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**ANÁLISIS Y DISEÑO DE UNA BASE DE DATOS MULTIDIMENSIONAL EN APOYO A LA
COORDINACIÓN NACIONAL DE SEGEPLAN DEL PROYECTO DE INVENTARIO DEL AGUA
DE GUATEMALA**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

ANDRES SAENZ AGUIRRE

ASESORADO POR EL ING. JORGE ARMÍN MAZARIEGOS RABANALES

AL CONFERÍRSELE EL TÍTULO DE

INGENIERO EN CIENCIAS Y SISTEMAS

GUATEMALA, JULIO DE 2011

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. Murphy Olympto Paiz Recinos
VOCAL I	Ing. Alfredo Enrique Beber Aceituno
VOCAL II	Ing. Pedro Antonio Aguilar Polanco
VOCAL III	Ing. Miguel Ángel Dávila Calderón
VOCAL IV	Br. Juan Carlos Molina Jiménez
VOCAL V	Br. Mario Maldonado Muralles
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Ing. Murphy Olympto Paiz Recinos
EXAMINADORA	Inga. Floriza Felipa Ávila Pesquera De Medinilla
EXAMINADOR	Ing. Marlon Pérez Turk
EXAMINADORA	Inga. Sonia Castañeda
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

ANÁLISIS Y DISEÑO DE UNA BASE DE DATOS MULTIDIMENSIONAL EN APOYO A LA COORDINACIÓN NACIONAL DE SEGEPLAN DEL PROYECTO DE INVENTARIO DEL AGUA DE GUATEMALA

Tema que me fuera asignado por la Dirección de la Escuela de Ingeniería en Ciencias y Sistemas, con fecha febrero de 2009.

Andres Saenz Aguirre.

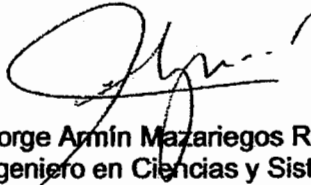
Guatemala 14 de Noviembre de 2009

Señores
Facultad de Ingeniería
Universidad de San Carlos de Guatemala
Guatemala, Ciudad

Por este medio hago de su conocimiento que he revisado el trabajo de graduación del estudiante Andres Saenz Aguirre que se identifica con carné 2002-19625, titulado **"Análisis y Diseño de una base de datos multidimensional en apoyo a la coordinación nacional de SEGEPLAN del proyecto de Inventario del Agua en Guatemala"**, y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo según el protocolo.

Sin otro particular, me suscribo de usted.

Atentamente,


Ing. Jorge Armin Mazariegos Rabanales
Ingeniero en Ciencias y Sistemas
Asesor de trabajo de EPS

Jorge Armin Mazariegos Rabanales
ING. EN CIENCIAS Y SISTEMAS
M.C. ADMINISTRADOR DE TECNOLOGIA
COLEGIADO 5547



Guatemala, 23 de marzo de 2011.
REF.EPS.DOC.506.03.2011.

Inga. Norma Ileana Sarmiento Zeceña de Serrano
Directora Unidad de EPS
Facultad de Ingeniería
Presente

Estimada Ingeniera Sarmiento Zeceña.

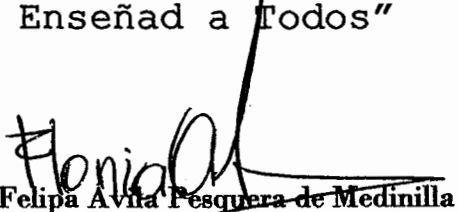
Por este medio atentamente le informo que como Supervisora de la Práctica del Ejercicio Profesional Supervisado, (E.P.S) del estudiante universitario de la Carrera de Ingeniería en Ciencias y Sistemas, **Andres Saenz Aguirre** Carné No. **200219625** procedí a revisar el informe final, cuyo título es **"ANÁLISIS Y DISEÑO DE UNA BASE DE DATOS MULTIDIMENSIONAL EN APOYO A LA COORDINACIÓN NACIONAL DE SEGEPLAN DEL PROYECTO DE INVENTARIO DEL AGUA DE GUATEMALA"**.

En tal virtud, **LO DOY POR APROBADO**, solicitándole darle el trámite respectivo.

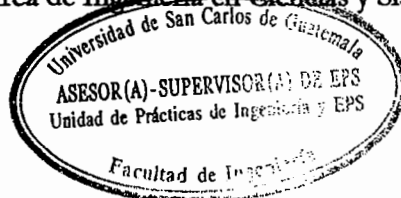
Sin otro particular, me es grato suscribirme.

Atentamente,

"Id y Enseñad a Todos"


Inga. Floriza Felipa Avila Pesquera de Medinilla
Supervisora de EPS
Área de Ingeniería en Ciencias y Sistemas

FFAPdM/RA





Guatemala, 23 de marzo de 2011.
REF.EPS.DOC.243.03.2011.

Ing. Marlon Antonio Pérez Turck
Director Escuela de Ingeniería Ciencias y Sistemas
Facultad de Ingeniería
Presente


Estimado Ingeniero Perez Turck.

Por este medio atentamente le envío el informe final correspondiente a la práctica del Ejercicio Profesional Supervisado, (E.P.S) titulado **“ANÁLISIS Y DISEÑO DE UNA BASE DE DATOS MULTIDIMENSIONAL EN APOYO A LA COORDINACIÓN NACIONAL DE SEGEPLAN DEL PROYECTO DE INVENTARIO DEL AGUA DE GUATEMALA”**, que fue desarrollado por el estudiante universitario **Andres Saenz Aguirre** Carné No. **200219625** quien fue debidamente asesorado por el Ing. Jorge Armin Mazariegos y supervisado por la Inga. Floriza Felipa Ávila Pesquera de Medinilla.

Por lo que habiendo cumplido con los objetivos y requisitos de ley del referido trabajo y existiendo la aprobación del mismo por parte del Asesor y la Supervisora de EPS, en mi calidad de Directora apruebo su contenido solicitándole darle el trámite respectivo.

Sin otro particular, me es grato suscribirme.

Atentamente,
“Id y Enseñad a Todos”


Inga. Norma Ileana Sarmiento Zeceña de Serrano
Directora Unidad de EPS

NISZ/ra





Universidad San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, 27 de Abril de 2011

Ingeniero
Marlon Antonio Pérez Turk
Director de la Escuela de Ingeniería
En Ciencias y Sistemas

Respetable Ingeniero Pérez:

Por este medio hago de su conocimiento que he revisado el trabajo de graduación-EPS del estudiante **ANDRES SAENZ AGUIRRE**, carné **2002-19625**, titulado: **"ANÁLISIS Y DISEÑO DE UNA BASE DE DATOS MULTIDIMENSIONAL EN APOYO A LA COORDINACIÓN NACIONAL DE SEGEPLAN DEL PROYECTO DE INVENTARIO DEL AGUA DE GUATEMALA"**, y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,


Ing. Carlos Alfredo Azurdia
Coordinador de Privados
y Revisión de Trabajos de Graduación



E
S
C
U
E
L
A

D
E

C
I
E
N
C
I
A
S

Y

S
I
S
T
E
M
A
S

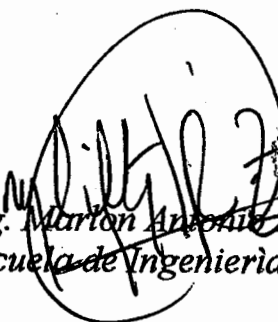
UNIVERSIDAD DE SAN CARLOS
DE GUATEMALA




FACULTAD DE INGENIERÍA
ESCUELA DE CIENCIAS Y SISTEMAS
TEL: 24767644

El Director de la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor con el visto bueno del revisor y del Licenciado en Letras, de trabajo de graduación titulado "ANÁLISIS Y DISEÑO DE UNA BASE DE DATOS MULTIDIMENSIONAL EN APOYO A LA COORDINACIÓN NACIONAL DE SEGEPLAN DEL PROYECTO DE INVENTARIO DEL AGUA DE GUATEMALA", presentado por el estudiante ANDRES SAENZ AGUIRRE, aprueba el presente trabajo y solicita la autorización del mismo.

"ID Y ENSEÑAD A TODOS"


Ing. Martín Antonio Pérez Turk
Director, Escuela de Ingeniería Ciencias y Sistemas



Guatemala, 05 de julio 2011



El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería en Ciencias y Sistemas, al trabajo de graduación titulado: **ANÁLISIS Y DISEÑO DE UNA BASE DE DATOS MULTIDIMENSIONAL EN APOYO A LA COORDINACIÓN NACIONAL DE SEGEPLAN DEL PROYECTO DE INVENTARIO DEL AGUA DE GUATEMALA**, presentado por el estudiante universitario **Andres Saenz Aguirre**, procede a la autorización para la impresión del mismo.

IMPRÍMASE.


Ing. Murphy Olimpo Paiz Reginos
DECANO



Guatemala, julio de 2011

/cc

ACTO QUE DEDICO A:

- Dios** Por darme el milagro de la vida.
- Mis padres** Edy y Lucrecia por su amor y apoyo de manera incondicional, no únicamente durante la carrera, sino durante toda la vida.
- Mis hermanas** Ana Renee y Anaely por su amor, paciencia y muchas veces compañía en las noches de desvelo.
- Mis asesores** Armín Mazariegos y Marco Vinicio De La Cruz por el tiempo y esfuerzo que invirtieron para la realización del proyecto de graduación. Y quiero hacer mención especial a Javier Gramajo que no aparece oficialmente como asesor del trabajo pero sí que lo fue.
- Mi novia** Flor por su amor, paciencia y comprensión.
- Mis amigos** Huber, Isaac, Adin, Paniagua, Joaquín, Danny, que más que mis compañeros de estudio, fueron personas que marcaron mi vida, y aunque estén lejos o cerca me han apoyado para terminar la carrera.

AGRADECIMIENTOS A:

- Dios** Por darme la vida y las aptitudes para poder terminar la carrera, y por la familia que me regaló.
- Mis padres** Por su amor y apoyo de manera incondicional, no únicamente durante la carrera, sino durante toda la vida.
- Mis asesores** Armin Mazariegos y Marco Morales por el esfuerzo y tiempo invertidos en este trabajo de graduación. Javier Gramajo López, por su amistad, apoyo y asesoría en muchos años de carrera.
- Mis amigos** Por su amistad y los recuerdos que hemos creado juntos.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES	XIX
LISTA DE SÍMBOLOS	XXI
GLOSARIO	XXV
RESUMEN	XXVII
OBJETIVOS.....	XXIX
INTRODUCCIÓN	XXXI
1. ALMACÉN DE DATOS	1
1.1. Descripción de un almacén de datos	1
1.2. Definiciones de almacén de datos	2
1.2.1. Definición de Bill Inmon.....	2
1.2.2. Definición de Ralph Kimball	3
1.2.3. Función de un almacén de datos	3
1.3. <i>Data Marts</i>	5
1.3.1. Dependencia de un <i>Data Mart</i>	5
1.3.2. Conceptos erróneos de los <i>data marts</i>	7
1.4. Cubos de información	7
1.4.1. Dimensiones y jerarquías.....	9
1.4.2. Dispersión en cubos OLAP	10
1.4.3. Acceso y cálculo de un cubo OLAP	10
1.4.4. Definición técnica	10
1.4.5. Variables	12
1.4.6. Tabla de hechos.....	12
1.4.7. Las métricas del negocio.....	13
1.4.8. Granularidad	13

1.4.9.	Agregación.....	14
1.4.10.	Tipos de datos	14
1.5.	Metadatos	15
1.5.1.	Manipulación de Metadatos	18
1.5.2.	Almacenamiento de Metadatos	18
1.5.3.	Codificación de Metadatos.....	19
1.5.4.	Vocabularios controlados y ontologías	19
1.5.5.	Formatos y estándares	19
1.6.	Proceso extracción, transformación y carga.....	20
1.6.1.	Extraer	21
1.6.2.	Transformar	22
1.6.3.	Carga	23
1.6.4.	Procesamiento paralelo	24
1.7.	Diseño de un almacén de datos	25
1.8.	Almacén de datos espacial	26
1.9.	Ventajas.....	28
1.10.	Desventajas	28
2.	SITUACIÓN ACTUAL.....	31
2.1.	Entendimiento de la problemática.....	31
2.2.	Inventario de recursos	39
2.3.	Riesgos y contingencias	39
2.3.1.	Atributos de los requerimientos	40
2.3.2.	Atributos del producto	41
2.3.3.	Atributos de la gente.....	42
2.3.4.	Atributos de la administración.....	43
2.3.5.	Atributos del ambiente	44
2.3.6.	Atributos del método de desarrollo	45
2.4.	Entendimiento de los datos.....	46

2.4.1.	Colección inicial de datos.....	46
2.4.2.	Boletas	48
2.4.3.	Exploración preliminar de los datos	80
2.4.4.	Verificar calidad de la data.....	81
3.	SOLUCIÓN PROPUESTA.....	83
3.1.	Metodología utilizada para resolver la problemática	83
3.1.1.	Entendimiento del negocio/organización.....	84
3.1.2.	Entendimiento de la data.....	85
3.1.3.	Preparación de los datos.....	85
3.1.4.	Construcción del modelo.....	85
3.1.5.	Evaluación.....	85
3.1.6.	Implementación.....	86
3.2.	Preparación de datos	87
3.2.1.	Selección de datos.....	87
3.3.	Proyecciones de las boletas sobre el modelo	97
3.4.	Construcción del modelo.....	100
3.5.	Descripción de cada una de las dimensiones y tablas de hechos.....	101
3.5.1.	Dim_UbicacionDeLaFuente	101
3.5.2.	Dim_Geografia	102
3.5.3.	Dim_Boleta.....	105
3.5.4.	Fact_SaneamientoYTratamientoDeAgua.....	107
3.5.5.	Dim_SistemaAbastecimientoAgua	109
3.5.6.	Fact_ResultadosDeLaboratorio.....	110
3.5.7.	Fact_FuentePozoMecanico.....	112
3.5.8.	Fact_UsoHidroElectrica.....	113
3.5.9.	Fact_UsoDomestico.....	114
3.5.10.	Fact_UsoIndustrialRecreativoPecuario	116

3.5.11. Dim_OrganizacionComunitaria	117
3.5.12. Dim_MiembrosOrganizacionComunitaria	119
3.5.13. Fact_SistemasFormalesDeAbastecimiento	120
3.5.14. Fact_SistemasInformalesDeAbastecimiento	120
3.5.15. Fact_FuentesDeAgua	121
4. EVALUACIÓN DE LA SOLUCIÓN PROPUESTA Y RECOMENDACIONES PARA FUTUROS DESARROLLOS.....	125
4.1. Evaluación del modelo.....	125
4.2. Recomendaciones para el modelo	126
4.3. Recomendaciones ETL.....	127
4.3.1. Pasos para un proceso de ETL	127
4.4. Recomendaciones para desarrollo de futuras aplicaciones informáticas	129
4.4.1. Consultas realizadas directamente en la base de datos...	129
4.4.2. Minería de datos	129
4.4.3. Reportes comunes.....	130
4.4.4. Aplicaciones para análisis.....	130
4.4.5. Cuadros de mandos y tarjetas de responsabilidades	130
4.4.6. Aplicaciones de Inteligencia de Negocios.....	131
4.4.7. Portal <i>Web</i>	131
4.4.8. Calendarización de los reportes	135
4.5. Recomendaciones de implementación	135
4.6. Recomendaciones post-implementación	136
4.6.1. Mantener el portal actualizado.....	136
4.6.2. Copias de respaldo y recuperación	137

CONCLUSIONES	141
RECOMENDACIONES	143
REFERENCIAS	145
BIBLIOGRAFÍA.....	149
ANEXOS	153

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Diagrama de fases de la metodología CRISP	84
2.	Tablas que corresponden a la boleta PARPA	98
3.	Tablas que corresponden a la boleta PROVIAGUA	99
4.	Tablas que corresponden a la boleta SAS	100
5.	Dimensión Dim_UbicacionDeLaFuente.....	101
6.	Dimensión Dim_Geografia	103
7.	Dimensión Dim_Boleta.....	105
8.	Tabla de hechos Fact_SaneamientoYTratamientoDeAgua.....	107
9.	Dimensión Dim_SistemaAbastecimientoAgua	109
10.	Tabla de hechos Fact_ResultadosDeLaboratorio	111
11.	Tabla de hechos Fact_FuentePozoMecanico	112
12.	Tabla de hechos Fact_UsoHidroelectrica.....	113
13.	Tabla de hechos Fact_UsoDomestico.....	115
14.	Tabla de hechos Fact_UsoIndustrialRecreativoPecuario	116
15.	Dimensión Dim_OrganizacionComunitaria.....	118
16.	Dimensión Dim_MiembrosOrganizacionComunitaria	119
17.	Tabla de hechos Fact_SistemasFormalesDeAbastecimiento	120
18.	Tabla de hechos Fact_SistemasInformalesDeAbastecimiento.....	121
19.	Tabla de hechos Fact_FuentesDeAgua	122

TABLAS

I.	Oferta hídrica de Guatemala.....	32
II.	Uso del agua en Guatemala	33
III.	Dimensión Dim_Boleta	88
IV.	Dimensión Dim_CalidadDelAgua.....	89
V.	Dimensión Dim_CalidadDelAgua.....	89
VI.	Dimensión Dim_MiembrosOrganizacionComunitaria	90
VII.	Dimensión Dim_OrganizacionComunitaria	90
VIII.	Dimensión Dim_SistemaAbastecimientoAgua.....	91
IX.	Dimensión Dim_UbicacionDeLaFuente	91
X.	Tabla de hechos Fact_CalidadDelAgua	92
XI.	Tabla de hechos Fact_FuentePozoMecanico.....	92
XII.	Tabla de hechos Fact_FuentesDeAgua	93
XIII.	Tabla de hechos Fact_ResultadosDeLaboratorio.....	93
XIV.	Tabla de hechos Fact_SaneamientoYTratamientoDeAgua.....	94
XV.	Tabla de hechos Fact_SistemasFormalesDeAbastecimiento.....	95
XVI.	Tabla de hechos Fact_UsoDomestico	95
XVII.	Tabla de hechos Fact_UsoHidroElectrica.....	96
XVIII.	Tabla de hechos Fact_UsoIndustrialRecreativoPecuario	96
XIX.	Tabla de hechos Fact_UsoIrrigacion	97

LISTA DE SÍMBOLOS

Símbolo	Significado
CDC	<i>Change Data Capture</i>
CONAP	Consejo Nacional de Áreas Protegidas de Guatemala.
EPS	Ejercicio Profesional Supervisado
GPM	Galones Por Minuto
INE	Instituto Nacional de Estadística de Guatemala
INSIVUMEH	Instituto Nacional de Sismología, Vulcanología, Meteorología e Hidrología de Guatemala
*	Llamada a referencia al pie de la página
MSNM	Metros sobre el nivel del mar
ML	Mililitros
MAGA	Ministerio de Agricultura Ganadería y Alimentación de Guatemala

MARN	Ministerio de Ambiente y Recursos Naturales de Guatemala
MSPAS	Ministerio de Salud Pública y Asistencia Social
MDX	<i>Multi Dimensional eXpressions</i>
OLAP	<i>On Line Analytical Process</i>
%	Porcentaje
PARPA	Programa de Apoyo a la Reconversión Productiva Agroalimentaria
SEGEPLAN	Secretaría de Planificación y Programación de la Presidencia de Guatemala
SLA	<i>Service Level Agreement</i>
UTM	Sistema de Coordenadas Universal Transversal de Mercator (En inglés <i>Universal Transverse Mercator</i>)
SIAGua	Sistema Integrado de Información y Conocimiento del Agua de Guatemala
SQL	<i>Structured Query Language</i>
URI	<i>Uniform Resource Identifier</i>

WGS84

World Geodetic System 84

GLOSARIO

Aforo	Medida del caudal de una corriente de agua.
Campo	En bases de datos, es cada una de las columnas que forman la tabla. Contienen datos de tipo diferente a los de otros campos.
Caudal	Cantidad de agua de una corriente.
ETL	Proceso de Extracción, Transformación y Carga de los datos. (Por sus siglas en inglés).
GPS	Sistema de posicionamiento global (por sus siglas en inglés: <i>Global Positioning System</i>) o <i>NAVSTAR-GPS1</i> es un sistema global de navegación por satélite que permite determinar en todo el mundo la posición de un objeto, una persona o un vehículo con una precisión hasta de centímetros, aunque lo habitual son unos pocos metros de precisión. El sistema fue desarrollado, instalado y actualmente operado por el Departamento de Defensa de los Estados Unidos de América.
Hídrico	Del agua o relativo a ella.

Métrica	Proporciona una indicación cuantitativa de extensión, cantidad, dimensiones, capacidad y tamaño de algunos atributos de un proceso o producto.
Registro	En las bases de datos, es cada una de las filas que contiene una tabla. Cada registro contiene datos de los mismos tipos que los demás registros.
Relación	En las bases de datos, se refiere al vínculo que existe entre dos tablas en un modelo de base de datos relacional, por un campo en común, en ambos casos posee las mismas características.
Tabla	En las bases de datos, se refiere al tipo de modelado de datos, donde se guardan los datos recogidos por un programa. Su estructura general se asemeja a la vista general de un programa de hoja de cálculo.
Trigger	En las bases de datos Un <i>trigger</i> es un procedimiento que se ejecuta cuando se cumple una condición establecida al realizar una operación.

RESUMEN

Actualmente la Secretaria Técnica del Gabinete Específico del Agua que está a cargo de la Dirección de Recursos Hídricos de La Secretaría de Planificación y Programación de la Presidencia (SEGEPLAN), tiene a su cargo la coordinación del trabajo de todas las comisiones que se integren dentro del gabinete para realizar el Sistema Nacional de Información y Conocimiento del Agua "SIAgua".

El gabinete del agua tiene definidas seis medidas gubernamentales, pero para el ámbito de este proyecto, únicamente nos interesa la medida: "Agua e Información". Acorde con los objetivos de dicha medida "Agua e Información", algunas de las acciones clave se pueden interpretar desde la lógica de los sistemas de información. Dentro de esas acciones se pueden señalar las siguientes:

- Recolectar la información: recabando toda la información existente que se encuentra en distintas instituciones, y distintos formatos, para unificarla e identificar los vacíos existentes y que es necesario generar en campo y gabinete en acciones futuras.
- Análisis de la información: este paso es parte importante del sistema nacional de información del agua y es una acción acorde con su tercer objetivo específico; además del análisis de información de carácter experto (hidrológico, económico, ambiental), se busca que tanto este análisis como la información previamente recolectada/generada, pueda ser accesible para la toma de decisiones, por lo que también se debe contar con una

herramienta para la generación de distintos reportes en el ámbito de las comunidades, los municipios, los departamentos, las cuencas hidrográficas (demarcación técnica para la planificación del agua) y a nivel de país.

- Divulgación de la información: haciendo el mejor uso de las Tecnologías de la Información y la Comunicación (TICS). La divulgación de la información se hace con el propósito de poder realizar distintas publicaciones con los datos obtenidos para los distintos grupos que la requieran, que pueden ser Ministros, Alcaldes, y población en general; dicha acción es estratégica para dar cumplimiento al decreto 57-2008 de libre acceso a la información.

El alcance de este Ejercicio Profesional Supervisado se limita a proveer el análisis y diseño de la base de datos multidimensional, en la cual se van a poder almacenar todos los datos recabados, para su posterior análisis y generación de los reportes necesarios.

OBJETIVOS

General

Elaborar el diseño de la base de datos multidimensional para el Sistema Nacional de Información y Conocimiento del Agua (SIAgua) el cual permita almacenar la información de cada una de las instituciones involucradas en realizar el inventario nacional de fuentes de agua en Guatemala.

Específicos

1. Realizar una investigación documental que permita identificar al menos cinco instituciones responsables de la realización del inventario nacional de fuentes de agua por medio de un listado.
2. Identificar y recolectar al menos tres instrumentos de recopilación de información que permiten llevar el inventario en las instituciones identificadas.
3. Identificar de los instrumentos: datos, registros, campos y sus posibles valores; los cuales son relevantes para la construcción de un diccionario de datos que finalmente permita diseñar un modelo conceptual.
4. Validar que el diseño del modelo conceptual propuesto cumple con los requerimientos de las distintas instituciones involucradas en el inventario, por medio de reuniones de trabajo.

5. Realizar al menos dos presentaciones del informe final, una reunión con el gabinete específico del agua y una a la Secretaría Técnica del Gabinete, para que los participantes se apropien de la propuesta.

INTRODUCCIÓN

El desarrollo de esta investigación se ha organizado en cuatro capítulos. El primer capítulo “Almacén de datos” consiste en todo el marco teórico para sustentar todos los conceptos que se van a discutir a lo largo de todo el documento, se describen los términos almacén de datos, *data mart*, se clarifican varios conceptos erróneos de un *data mart*, definiciones de diferentes autores sobre el tema, una breve descripción del proceso de ETL.

En el segundo capítulo “Situación actual” contextualiza al lector con las políticas y los esfuerzos que se han hecho para tener el inventario nacional de fuentes de agua, así como las dificultades que se tienen para unificarlo, se identifican las instituciones involucradas con el inventario, una descripción de las herramientas que las instituciones utilizan actualmente para recabar el inventario, y el análisis de la calidad de los datos.

En el tercer capítulo “Solución propuesta” se describe la metodología CRISP, que es la que se propone para resolver la problemática, y una descripción de cada una de las fases de dicha metodología, se describen los datos seleccionados de las diferentes fuentes de datos, así como se identifica información duplicada, se definen los campos de cada una de las tablas que componen el modelo, así como el tipo de datos para almacenar la información. Se demuestra cómo en el modelo propuesto de datos se cubren todas las áreas de cada una de las fuentes de datos de origen.

En el cuarto capítulo “Evaluación de la solución propuesta y recomendaciones para futuros desarrollos” se describen los resultados obtenidos de las pruebas realizadas al modelo, y un conjunto de recomendaciones para la implementación del modelo propuesto, también para futuros desarrollos, dentro de las cuales podemos mencionar: recomendaciones para el modelo, ETL, aplicaciones para aprovechar el modelo de datos, y mantenimiento del sistema.

1. ALMACÉN DE DATOS

1.1. Descripción de un almacén de datos

En el contexto de la informática, un almacén de datos (del inglés *data warehouse*) es una colección de datos orientada a un determinado ámbito el cual es de interés para una empresa, organización, es un conjunto de datos integrado, no volátil y variable en el tiempo, que sirve de apoyo a la toma de decisiones en la entidad en la que se utiliza. Básicamente es un extenso conjunto de datos históricos de una organización, más allá de la información de las operaciones y transacciones diarias, dicho conjunto de datos es almacenado en una base de datos diseñada para favorecer el análisis y la publicación eficiente de datos especialmente OLAP¹.

El almacenamiento de los datos no debe usarse con datos de uso actual debido a que los almacenes de datos contienen a menudo grandes cantidades de información que se subdividen a veces en unidades lógicas más pequeñas dependiendo del subsistema de la entidad del que procedan o para el que sea necesario, ocasionando un incremento considerable en la respuesta de los servidores de datos.

¹ *On-Line Analytical Processing.*

1.2. Definiciones de almacén de datos

1.2.1. Definición de Bill Inmon

Bill Inmon fue uno de los primeros autores en escribir sobre el tema de los almacenes de datos, define un almacén de datos en términos de las características del repositorio de datos:

- Orientado a temas: los datos en la estructura de la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí;
- Variante en el tiempo: los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los reportes que se puedan generar reflejen esas variaciones;
- No volátil: la información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas;
- Integrado: la base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.

Inmon propone una metodología descendente [1] a la hora de diseñar un almacén de datos, ya que de esta forma se considerarán mejor todos los datos de la organización. En esta metodología los *Data Marts* se crearán después de haber terminado el almacén de datos completo de la organización.

1.2.2. Definición de Ralph Kimball

Este es otro reconocido autor en el tema de los almacenes de datos, define un almacén de datos como: "una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis" [2]. También fue Kimball quien definió que un almacén de datos no era más que: "la unión de todos los *Data Marts* de una entidad". Defiende por tanto una metodología ascendente [3] a la hora de diseñar un almacén de datos.

Las definiciones anteriores se basan en los datos en sí mismos. Sin embargo, los medios para obtener y analizar esos datos, para extraerlos, transformarlos y cargarlos, así como las diferentes formas para realizar la administración de datos son componentes básicos de un almacén de datos. Muchas referencias a un almacén de datos utilizan esta definición más amplia; por lo tanto, en esta definición se incluyen herramientas para la inteligencia empresarial, herramientas para extraer, transformar y cargar datos (ETL por sus siglas en inglés) en el almacén de datos, y herramientas para gestionar y recuperar los metadatos.

1.2.3. Función de un almacén de datos

En un almacén de datos lo que se quiere es recopilar datos que son de utilidad para una organización, es decir, que se utiliza como un repositorio para almacenar los datos para luego transformarlos en información de utilidad para el usuario. Un almacén de datos debe entregar la información correcta a la gente que corresponde de manera óptima y en un formato que sea entendible para el usuario. El almacén de datos debe suplir las necesidades de usuarios expertos, utilizando sistemas de soporte a decisiones [4], sistemas de información ejecutiva [5] o herramientas para hacer consultas o informes. Los usuarios

finales pueden hacer fácilmente consultas sobre sus almacenes de datos sin tocar o afectar el sistema operacional. En el funcionamiento de un almacén de los datos son muy importantes las siguientes ideas:

- Integración de los datos provenientes de bases de datos distribuidas por las diferentes unidades de la organización y que con frecuencia tendrán diferentes estructuras (fuentes de origen heterogéneas). Se debe facilitar una descripción global y un análisis comprensivo de toda la organización en el almacén de datos.
- Separación de los datos usados en operaciones diarias de los datos usados en el almacén de datos para los propósitos de divulgación, de ayuda en la toma de decisiones, para el análisis y para operaciones de control. Ambos tipos de datos no deben coincidir en la misma base de datos, ya que obedecen a objetivos muy distintos y podrían perjudicar el desempeño entre ambos sistemas.

Periódicamente se importan datos al almacén de datos de los distintos sistemas de planeamiento de recursos de la entidad [6] y de otros sistemas de *software* relacionados con el negocio para la transformación posterior. Es práctica común normalizar los datos antes de combinarlos en el almacén de datos mediante herramientas de extracción, transformación y carga. Estas herramientas leen los datos primarios, comúnmente de bases de datos transaccionales [7] de la organización, realizan el proceso de transformación al almacén de datos que puede incluir procesos como filtración, adaptación, cambios de formato, etc. y luego son cargadas en el almacén de datos.

1.3. Data Marts

Un *data mart* es subconjunto de datos del almacén de datos con el propósito de ayudar a que un área o departamento específico pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples maneras para que diferentes grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades.

El *data mart* es un sistema orientado a la consulta, en el que se producen procesos por lotes de carga de datos con una frecuencia conocida y no muy alta. Es consultado mediante herramientas OLAP que ofrecen una percepción multidimensional de la información. Sobre estas bases de datos se pueden construir EIS² y DSS³. Por otra parte también existe la minería de datos [8] que es el proceso no trivial de análisis de grandes cantidades de datos con el objetivo de extraer información útil, como pueden ser realizar clasificaciones o predicciones de tendencias.

1.3.1. Dependencia de un Data Mart

Según la tendencia marcada por Inmon sobre los almacenes de datos, un *data mart* dependiente es un subconjunto lógico o un subconjunto físico de un almacén de datos más grande, que se ha apartado por alguna de las siguientes razones:

² *Executive Information System*

³ *Decision Support System*

- Si se necesita para un esquema o modelo de datos espacial;
- Prestaciones: para descargar el *data mart* a una computadora independiente para mejorar la eficiencia o para obviar las necesidades de administrar todo el volumen del almacén de datos centralizado;
- Seguridad: para separar un subconjunto de datos de forma selectiva a los que queremos permitir o restringir el acceso;
- Conveniencia: la de poder pasar por alto las autorizaciones y requerimientos necesarios para poder incorporar una nueva aplicación en el almacén de datos principal de la empresa;
- Demostración: para demostrar la viabilidad y el potencial de una aplicación antes de migrarla al almacén de datos de la Empresa;
- Política: cuando se decide una estrategia para las TICs⁴ en situaciones en las que un grupo de usuarios tiene más influencia, para determinar si se financia dicha estrategia o descubrir si ésta no sería buena para el almacén de datos centralizado;
- Política: estrategia para los consumidores de los datos en situaciones en las que un equipo de almacén de datos no está en condiciones de crear un almacén de datos utilizable.

Según la definición de Inmon de almacenes de datos, entre las pérdidas inherentes al uso de *data marts* están la escalabilidad limitada, la duplicación de datos, la inconsistencia de los datos con respecto a otros almacenes de información y la incapacidad para aprovechar las fuentes de datos de la empresa.

⁴ Tecnologías de la información y las comunicaciones

1.3.2. Conceptos erróneos de los *data marts*

Al hablar de los *data marts* generalmente se hace la comparación con los almacenes de datos y al final se llega a la conclusión de que un *data mart* es como un almacén de datos pero en pequeño, hasta cierto punto este argumento es válido, pero esta idea suele hacer caer en conceptos erróneos sobre la implementación y funcionamiento de los *data marts*, por lo que es necesario aclarar lo siguiente:

- La implementación de un *data mart*, no es más simple ni rápida que la de un almacén de datos, es un proceso bastante similar, y debe proporcionar la misma funcionalidad que un almacén de datos;
- Los recursos que necesita un *data mart*, no son menores a los que necesita un almacén de datos;
- Las consultas no son más rápidas, dado el menor volumen de datos, el menor volumen de datos se debe a que no se tienen todos los datos de toda la empresa, pero si se tienen todos los datos de un determinado segmento de la organización, por lo que una consulta sobre dicho segmento tarda lo mismo si se hace sobre el *data mart* que si se hace sobre el almacén de datos;
- Actualizar el *Data Mart* desde el almacén de datos cuesta menos que actualizar el almacén de datos desde sus fuentes de datos primarias, porque los formatos de los datos son o suelen ser idénticos.

1.4. Cubos de información

OLAP es un término adoptado por *EF Codd & Associates*, que publicó un libro blanco [9] en 1993, solicitado por *Arbor Software* actualmente *Hyperion Solutions* [10], titulado "Prestación de OLAP para análisis de usuario: Un

mandato de las tecnologías de la información". Hay que remarcar que Codd cobró para escribir este libro blanco de apoyo a Essbase [11]. A menudo se solía pensar que todo lo que los usuarios pueden querer de un sistema de información se podría hacer de una base de datos relacional. Sin embargo Codd fue uno de los precursores de las bases de datos relacionales, por lo que sus opiniones fueron y son respetadas.

La propuesta de Codd consistía en realizar una disposición de los datos en vectores para permitir un análisis rápido. Estos vectores son llamados cubos. Disponer los datos en cubos evita una limitación de las bases de datos relacionales, que no son muy adecuadas para el análisis de instantáneas de grandes cantidades de datos. Las bases de datos relacionales son más adecuadas para registrar datos provenientes de transacciones (conocido como OLTP). Aunque existen muchas herramientas de generación de informes para bases de datos relacionales, éstas son lentas cuando debe explorarse toda la base de datos.

Nos referimos a cubos OLAP cuando hablamos de bases de datos multidimensionales, en las cuales el almacenamiento físico de los datos se realiza en vectores multidimensionales. Los cubos OLAP se pueden considerar como una ampliación de las dos dimensiones de una hoja de cálculo. Estos parámetros en función de los cuales se analizan los datos se conocen como dimensiones. Para acceder a los datos sólo es necesario indexarlos a partir de los valores de las dimensiones o ejes.

Almacenar físicamente los datos de esta forma tiene sus ventajas y sus desventajas. Por ejemplo, en estas bases de datos las consultas de selección son muy rápidas. Pero uno de los problemas más grandes de esta forma de almacenamiento es que una vez poblada la base de datos ésta no permite realizar cambios en su estructura. Para ello sería necesario rediseñar el cubo.

En un sistema OLAP puede haber más de tres dimensiones, por lo que los cubos OLAP también reciben el nombre de hipercubos. Las herramientas comerciales OLAP tienen diferentes métodos de creación y vinculación de estos cubos o hipercubos

1.4.1. Dimensiones y jerarquías

Cada una de las dimensiones de un cubo OLAP puede resumirse mediante una jerarquía. Por ejemplo, si se considera una escala o dimensión temporal “Mayo de 2005” se puede incluir en “Segundo Trimestre de 2005”, que a su vez se incluye en “Año 2005”. De igual manera, otra dimensión de un cubo que refleje una situación geográfica, las ciudades se pueden incluir en regiones, países o regiones mundiales; los productos podrían clasificarse por categorías, y las partidas de gastos podrían agruparse en tipos de gastos.

En cambio, el analista podría comenzar en un nivel muy resumido, como por ejemplo el total de la diferencia entre los resultados reales y lo presupuestado, para posteriormente descender en el cubo en sus jerarquías para poder observar con un mayor nivel de detalle que le permita descubrir en el cubo los lugares en los que se ha producido esta diferencia, según los productos y períodos.

1.4.2. Dispersión en cubos OLAP

Vincular o enlazar cubos es un mecanismo para superar la dispersión. Se le llama dispersión cuando sucede que no todas las celdas del cubo se rellenan con datos. El tiempo de procesamiento es tan valioso que se debe adoptar la manera más efectiva de sumar ceros que puede incluir los valores nulos o no existentes. En lugar de crear un cubo disperso, a veces es mejor crear otro distinto, pero vinculado, cubo en el que un subconjunto de los datos se puede analizar con gran detalle. La vinculación asegura que los datos de los dos cubos se mantienen coherentes.

1.4.3. Acceso y cálculo de un cubo OLAP

Los datos de los cubos pueden ser actualizados esporádicamente, tal vez por personas diferentes de modo concurrente. Para solucionar este problema a menudo es necesario bloquear partes de un cubo mientras otro usuario está escribiendo, para volver a calcular los totales en el cubo. Otras implementaciones añaden la posibilidad de mostrar una alerta que indique que los totales calculados previamente ya no son válidos tras los nuevos datos. También hay algunos productos que calculan los totales cuando se les necesita con los últimos datos producidos en el sistema.

1.4.4. Definición técnica

En teoría de bases de datos, un cubo OLAP es una representación abstracta de la proyección de una relación de un Sistema administrador de bases de datos relacionales. Dada una relación de orden N , se considera la posibilidad de una proyección que dispone de los campos X, Y, Z como clave de la relación y de W como atributo residual.

Categorizando esto como una función se tiene que:

$$W : (X, Y, Z) \rightarrow W,$$

Los atributos X, Y, Z se corresponden con los ejes del cubo, mientras que el valor de W devuelto por cada tripleta (X, Y, Z) se corresponde con el dato o elemento que se rellena en cada celda del cubo.

Debido a que los dispositivos de salida sólo cuentan con dos dimensiones, no pueden caracterizar fácilmente cuatro dimensiones, es más práctico proyectar "rebanadas" o secciones de los datos del cubo, tales como la siguiente expresión:

$$W : (X, Y) \rightarrow W$$

Aunque no se conserve la clave del cubo (al faltar el parámetro Z), puede tener algún significado semántico; sin embargo, también puede que una sección de la representación funcional con tres parámetros para un determinado valor de Z también resulte de interés.

La motivación que hay tras OLAP vuelve a mostrar de nuevo el paradigma de los informes de tablas cruzadas de los sistemas de gestión de base de datos de los años ochenta. Se puede desear una visualización al estilo de una hoja de cálculo, donde los valores de X se encuentran en la fila \$1, los valores de Y aparecen en la columna \$A, y los valores de W en función de X y Y se encuentran en las celdas individuales a partir de la celda \$B2 y desde ahí, hacia abajo y hacia la derecha. Si bien se puede utilizar el lenguaje de manipulación de datos de SQL para mostrar las tuplas (X, Y, W), este formato de salida no es tan deseable como la alternativa de tablas cruzadas. El primer método requiere que se realice una búsqueda lineal para cada par (X, Y) dado,

para determinar el correspondiente valor de W, mientras que el segundo permite realizar una búsqueda más convenientemente permitiendo localizar el valor W en la intersección de la columna X apropiada con la fila Y correspondiente.

Se ha desarrollado el lenguaje MDX [12] para poder expresar problemas OLAP de forma fácil. Aunque es posible traducir algunas sus sentencias a SQL tradicional, con frecuencia se requieren expresiones SQL poco claras, incluso para las sentencias más simples del MDX. Este lenguaje ha sido aceptado por la gran mayoría de los proveedores de OLAP y se ha convertido en el estándar para estos sistemas.

1.4.5. Variables

Representan algún aspecto cuantificable o medible de los objetos o eventos a analizar. Normalmente, las variables son representadas por valores detallados y numéricos para cada instancia del objeto o evento medido. También son llamadas “indicadores de gestión”; son los datos que se quieren analizar, forman parte de la tabla de hechos. En forma contraria, las dimensiones son atributos relativos a las variables, y son utilizadas para indexar, ordenar, agrupar o abreviar los valores de las mismas. Las dimensiones poseen una granularidad menor, tomando como valores un conjunto de elementos menor que el de las variables.

1.4.6. Tabla de hechos

En las bases de datos, y más concretamente en un almacén de datos, una tabla de hechos es la tabla central de un esquema dimensional y contiene los valores de los indicadores del negocio. Cada métrica se toma mediante la

intersección de las dimensiones que la definen, dichas dimensiones estarán reflejadas en sus correspondientes tablas de dimensiones que rodearán la tabla de hechos y estarán relacionadas con ella.

1.4.7. Las métricas del negocio

Las métricas más útiles para incluir en una tabla de hechos son las sumatorias, es decir, aquéllas que pueden ser sumadas son medidas numéricas que pueden calcularse con la suma de varias cantidades de la tabla. En consecuencia, los hechos a almacenar en una tabla de hechos van a ser casi siempre valores numéricos, enteros o reales.

1.4.8. Granularidad

Una característica importante que define a una tabla de hechos es el nivel de granularidad de los datos que en ella se almacenan, entendiéndose por 'granularidad' el nivel de detalle de dichos datos, es decir, la granularidad de la tabla de hechos representa el nivel más atómico por el cual se definen los datos. Por ejemplo, no es lo mismo contar el tiempo por horas que por semanas; o en el caso de los productos, se puede considerar cada variante de un mismo artículo como un producto o agrupar todos los artículos de una misma familia considerándolos como un único producto.

Como se puede observar, la granularidad afecta a la cardinalidad, tanto de las dimensiones como de la tabla de hechos, a mayor granularidad (grano más fino) mayor será el número de registros final de la tabla de hechos.

Cuando la granularidad es mayor, es frecuente que se desee disponer de subtotales parciales, es decir, si tenemos una tabla de hechos con las ventas

por días, podría interesar disponer de los totales semanales o mensuales, estos datos se pueden calcular haciendo sumas parciales, pero es frecuente añadir a la tabla de hechos registros donde se almacenan dichos cálculos para no tener que repetirlos cada vez que se requieran y mejorar así el rendimiento de la aplicación. En este caso se dispondrá en la misma tabla de hechos de datos de grano fino y de grano más grueso, aumentando aún más la cardinalidad [25] de la tabla.

1.4.9. Agregación

La agregación es un proceso de cálculo por el cual se resumen los datos de los registros de detalle, esta operación consiste normalmente en el cálculo de totales dando lugar a medidas de grano grueso. Cuando se resumen los datos, el detalle ya no está directamente disponible para el analista, ya que éste se elimina de la tabla de hechos.

Esta operación se realiza típicamente con los datos más antiguos de la empresa con la finalidad de seguir disponiendo de dicha información, aunque sea de manera resumida para poder eliminar registros obsoletos de la tabla de hechos para liberar espacio.

1.4.10. Tipos de datos

Es normal que las tablas de hechos almacenen muchos millones de registros por esta razón hay que procurar utilizar los tipos de datos adecuados, si una medida a almacenar puede guardarse en un campo de tipo entero, no debemos definir ese campo como de tipo entero largo o como tipo real. Del mismo modo, si una magnitud necesita decimales, será mejor utilizar un tipo real simple que un tipo real de doble precisión. Nótese que elegir uno u otro de

estos campos, en principio sólo supondría una diferencia de unos pocos bytes en un registro, pero dado que en una tabla de hechos estamos hablando de cientos de millones de registros, en realidad, esa diferencia es bastante significativa.

1.5. Metadatos

Son datos que describen otros datos. En general, un grupo de metadatos se refiere a un grupo de datos, llamado recurso. El concepto de metadatos es análogo al uso de índices para localizar objetos en vez de datos. Por ejemplo, en una biblioteca se usan fichas que especifican: autores, títulos, casas editoriales y lugares para buscar libros. Así, los metadatos ayudan a ubicar datos.

Para varios campos de la informática, como la recuperación de información o la web semántica, los metadatos en etiquetas son un enfoque importante para construir un puente sobre el intervalo semántico.

El término metadatos no tiene una definición única. Según la definición más difundida de metadatos es que son datos sobre datos. También hay muchas declaraciones como informaciones sobre datos, datos sobre informaciones e informaciones sobre informaciones.

La mayoría de las veces no es posible diferenciar entre datos y metadatos. Muchas veces, los datos son tanto datos como metadatos.

Debido a que los metadatos son datos en sí mismos, es posible crear metadatos sobre metadatos. Aunque, a primera vista, no parece tener lógica, los metadatos sobre metadatos pueden ser muy útiles. Por ejemplo, fusionando

dos imágenes y sus metadatos distintos puede ser muy importante deducir cuál es el origen de cada grupo de metadatos, registrando ello en metadatos sobre los metadatos.

El intervalo semántico plantea el problema de que el usuario y la computadora no se entiendan, porque esta última no comprenda el significado de los datos. Es posible que los metadatos posibiliten la comunicación declarando cómo están relacionados los datos. Por eso, la representación del conocimiento usa metadatos para categorizar informaciones. La misma idea facilita la inteligencia artificial al deducir conclusiones automáticamente.

Los metadatos facilitan el flujo de trabajo convirtiendo datos automáticamente de un formato a otro. Para eso es necesario que los metadatos describan contenido y estructura de los datos.

Otra idea de aplicación es la presentación variable de datos. Si hay metadatos señalando los detalles más importantes, un programa puede seleccionar la forma de presentación más adecuada. Por ejemplo, si un teléfono móvil sabe dónde está localizada una persona en una imagen, tiene la posibilidad de reducirlo a las dimensiones de su pantalla. Del mismo modo, un navegador puede decidir presentar un diagrama a su usuario ciego en forma táctil o leída.

Subdividir metadatos por su contenido es lo más común. Se puede separar los metadatos que describen el recurso mismo de los que describen el contenido del recurso. Es posible subdividir estos dos grupos más veces, por ejemplo, para separar los metadatos que describen el sentido del contenido de los que describen la estructura del contenido o los que describen el recurso mismo de los que describen el ciclo vital del recurso.

Se pueden crear metadatos manualmente, semiautomáticamente o automáticamente. El proceso manual puede ser muy laborioso, dependiente del formato usado y del volumen deseado, hasta un grado en el que los seres humanos no puedan superarlo. Por eso, el desarrollo de utillaje semiautomático o automático es más que deseable.

En la producción automática el *software* adquiere las informaciones que necesita sin ayuda externa. Aunque el desarrollo de algoritmos tan avanzados está siendo objeto de investigación actualmente, no es probable que la computadora vaya a ser capaz de extraer todos los metadatos automáticamente. En vez de ello, se considera la producción semiautomática más realista; aquí un servidor humano sostiene algoritmos autónomos con la aclaración de inseguridades o la proposición de informaciones que el *software* no puede extraer sin ayuda.

Hay muchos expertos que se encargan del diseño de herramientas para la creación de metadatos pero que ignoran cuestionar este proceso. Según los que no evitan el asunto, la generación no debe comenzar después de la terminación de un recurso sino que debe hacerse durante la fabricación: hay que archivar los metadatos tan pronto como se originan, con los conocimientos especiales del productor, para evitar una laboriosa reconstrucción posterior. Por eso, se tiene que integrar la producción de metadatos en el procedimiento de fabricación del recurso.

1.5.1. Manipulación de metadatos

Si los datos cambian, los metadatos tienen que cambiar también. Aquí se hace la pregunta quién va a adaptar los metadatos. Hay modificaciones que pueden ser manejadas sencilla y automáticamente, pero hay otras donde la intervención de un actor humano es indispensable.

La meta-producción es el reciclaje de partes de recursos para crear otros recursos demanda atención particular. La fusión de los metadatos afiliados no es trivial, especialmente si se trata de información con relevancia jurídica. En algunos casos es conveniente eliminar los metadatos junto con sus recursos, en otros es razonable conservar los metadatos.

1.5.2. Almacenamiento de metadatos

Hay dos posibilidades para almacenar metadatos: depositarlos internamente en el mismo medio que los datos, o depositarlos externamente en su mismo recurso. Inicialmente, los metadatos se almacenaban internamente para facilitar la administración.

Hoy, por lo general, se considera mejor opción la localización externa porque hace posible la concentración de metadatos para optimizar operaciones de busca. Por el contrario, existe el problema de cómo se liga un recurso con sus metadatos. La mayoría de los estándares usa URIs [13], la técnica de localizar documentos en la *World Wide Web* [14].

1.5.3. Codificación de metadatos

Los primeros y más simples formatos de los metadatos usaron texto no encriptado o la codificación binaria para almacenar metadatos en archivos de texto. En la actualidad es común codificar metadatos usando XML⁵. Así, son legibles tanto por seres humanos como por computadoras. Además, este lenguaje es aceptado ya como estándar mundial. Pero también hay inconvenientes: los datos necesitan más espacio de memoria que en formato binario y no está claro cómo convertir la estructura de árbol en un flujo de datos. Por eso, muchos estándares incluyen utilidades para convertir XML en codificación binaria y viceversa, de forma que se tienen las ventajas de los dos.

1.5.4. Vocabularios controlados y ontologías

Para garantizar la uniformidad y la compatibilidad de los metadatos, muchos sugieren el uso de un vocabulario controlado fijando los términos de un campo. Una ontología además, define las relaciones de los términos del vocabulario para que la computadora pueda evaluarlas automáticamente. Así es posible presentar una página web sobre un tema en específico aunque el usuario busque un tema en general; usando una ontología adecuada el buscador comprende cómo llegar del tema general a uno más específico.

1.5.5. Formatos y estándares

Hay dos grupos que impulsan el desarrollo de formatos de metadatos: la técnica multimedia y la web semántica. El objetivo de la técnica multimedia es describir un singular recurso de multimedia, el de la web semántica la descripción de recursos de cada tipo y además la gestión del conocimiento.

⁵ *eXtensible Markup Language*

1.6. Proceso extracción, transformación y carga

ETL son las siglas en inglés de extraer, transformar y cargar, es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos, depurarlos, y cargarlos en otra base de datos, *data mart*, o almacén de datos para analizar, o en otro sistema operacional para apoyar un proceso de negocio. Los procesos ETL también se pueden utilizar para la integración con sistemas heredados.

En un sistema operacional el rango de valores de los datos o la calidad de éstos pueden no coincidir con las expectativas de los diseñadores a la hora de especificarse las reglas de validación o transformación. Es recomendable realizar un examen completo de la validez de los datos del sistema de origen durante el análisis, para identificar las condiciones necesarias para que los datos puedan ser tratados adecuadamente por las reglas de transformación especificadas. Esto conducirá a una modificación de las reglas de validación implementadas en el proceso ETL.

Normalmente los almacenes de datos son alimentados de manera asíncrona desde distintas fuentes, que sirven a propósitos muy diferentes. El proceso ETL es clave para lograr que los datos extraídos asíncronamente de orígenes heterogéneos se integren finalmente en un entorno homogéneo.

La escalabilidad de un sistema de ETL durante su vida útil tiene que ser establecida durante el análisis. Esto incluye la comprensión de los volúmenes de datos que tendrán que ser procesados según los acuerdos de nivel de servicio [15], El tiempo disponible para realizar la extracción de los sistemas de origen podría cambiar, lo que implicaría que la misma cantidad de datos tendría que ser procesada en menos tiempo. Algunos sistemas ETL son escalados para

procesar varios terabytes de datos para actualizar un almacén de datos que puede contener decenas de terabytes⁶ de información. El aumento de los volúmenes de datos que pueden requerir estos sistemas puede hacer que los lotes que se procesaban a diario pasen a procesarse en micro-lotes o incluso a la integración con colas de mensajes o a la captura de datos modificados [16] en tiempo real para una transformación y actualización continua.

Es necesario considerar un *middleware* [17] en el contexto de los almacenes de datos para asegurar la conectividad entre todos los componentes de la arquitectura de un almacén de datos.

1.6.1. Extraer

La primera parte del proceso ETL consiste en extraer los datos desde los sistemas de origen. La mayoría de los proyectos de almacenamiento de datos fusionan datos provenientes de diferentes sistemas de origen. Cada sistema separado puede usar una organización diferente de los datos o formatos distintos. Los formatos de las fuentes normalmente se encuentran en bases de datos relacionales o archivos planos, pero pueden incluir bases de datos no relacionales u otras estructuras diferentes. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación.

Una parte intrínseca del proceso de extracción es la de analizar los datos extraídos, de lo que resulta un chequeo que verifica si los datos cumplen la pauta o estructura que se esperaba. De no ser así los datos son rechazados.

Un requerimiento importante que se debe exigir a la tarea de extracción es que ésta cause un impacto mínimo en el sistema origen. Si los datos a

⁶ 1 Terabyte = 1024 Gigabytes

extraer son muchos, el sistema de origen se podría ralentizar e incluso colapsar, provocando que éste no pueda utilizarse con normalidad para su uso cotidiano. Por esta razón, en sistemas grandes las operaciones de extracción suelen programarse en horarios o días donde este impacto sea nulo o mínimo.

1.6.2. Transformar

La fase de transformación aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. Algunas fuentes de datos requerirán alguna pequeña manipulación de los datos. No obstante en otros casos pueden ser necesarias aplicar algunas de las siguientes transformaciones:

- Seleccionar sólo ciertas columnas para su carga
- Traducir códigos
- Codificar valores libres
- Obtener nuevos valores calculados
- Unir datos de múltiples fuentes
- Calcular totales de múltiples filas de datos
- Generación de campos clave en el destino
- Transponer múltiples columnas en filas o viceversa
- Separar el valor de una columna en varias

La aplicación de cualquier forma, simple o compleja, de validación de datos, y la consiguiente aplicación de la acción que en cada caso se requiera:

- Datos correctos: entregar datos a la siguiente etapa
- Datos erróneos: ejecutar políticas de tratamiento de excepciones

1.6.3. Carga

La fase de carga es el momento en el cual los datos de la fase de transformación son cargados en el sistema de destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes. En algunas bases de datos se sobrescribe la información antigua con nuevos datos. Los almacenes de datos mantienen un historial de los registros de manera que se pueda hacer una auditoría de los mismos y disponer de un rastro de toda la historia de un valor a lo largo del tiempo.

La fase de carga interactúa directamente con la base de datos de destino. Al realizar esta operación se aplicarán todas las restricciones y disparadores que se hayan definido en ésta. Estas restricciones y disparadores contribuyen a que se garantice la calidad de los datos en el proceso ETL, y deben ser considerados. Existen dos formas básicas de desarrollar el proceso de carga:

- Acumulación simple
- *Rolling*

1.6.3.1. Acumulación simple

La acumulación simple es la más sencilla y común, y consiste en realizar un resumen de todas las transacciones comprendidas en el período de tiempo seleccionado y transportar el resultado como una única transacción hacia el almacén de datos, almacenando un valor calculado que consistirá típicamente en una sumatoria o un promedio de la magnitud considerada.

1.6.3.2. Rolling

El proceso de *rolling* por su parte, se aplica en los casos en que se opta por mantener varios niveles de granularidad. Para ello se almacena información resumida a distintos niveles, correspondientes a distintas agrupaciones de la unidad de tiempo o diferentes niveles jerárquicos en alguna o varias de las dimensiones de la magnitud almacenada.

1.6.4. Procesamiento paralelo

Un avance reciente en el *software* ETL es la aplicación de procesamiento paralelo. Esto ha permitido desarrollar una serie de métodos para mejorar el rendimiento general de los procesos ETL cuando se trata de grandes volúmenes de datos. Hay tres tipos principales de paralelismos que se pueden implementar en las aplicaciones ETL:

- De datos: consiste en dividir un único archivo secuencial en pequeños archivos de datos para proporcionar acceso paralelo;
- De segmentación: permitir el funcionamiento simultáneo de varios componentes en el mismo flujo de datos. Un ejemplo de ello sería buscar un valor en el registro número 1 a la vez que se suman dos campos en el registro número 2;
- De componente: consiste en el funcionamiento simultáneo de múltiples procesos en diferentes flujos de datos en el mismo puesto de trabajo.

Estos tres tipos de paralelismo no son excluyentes, sino que pueden ser combinados para realizar una misma operación ETL.

Una dificultad adicional es asegurar que los datos que se cargan sean relativamente consistentes. Las múltiples bases de datos de origen tienen diferentes ciclos de actualización. En un sistema de ETL será necesario que se puedan detener ciertos datos hasta que todas las fuentes estén sincronizadas. Del mismo modo, cuando un almacén de datos tiene que ser actualizado con los contenidos en un sistema de origen, es necesario establecer puntos de sincronización y de actualización.

1.7. Diseño de un almacén de datos

Para construir un almacén de datos se necesitan herramientas para ayudar a la migración y a la transformación de los datos hacia el almacén. Una vez construido, se requieren medios para manejar grandes volúmenes de información. Se diseña su arquitectura dependiendo de la estructura interna de los datos del almacén y especialmente del tipo de consultas a realizar. Con este criterio los datos deben ser repartidos entre numerosos *data marts*. Para abordar un proyecto de almacén de datos es necesario hacer un estudio de algunos temas generales de la organización o empresa, los cuales se describen a continuación:

- Situación actual de partida: cualquier solución propuesta de almacén de datos debe estar muy orientada por las necesidades del negocio y debe ser compatible con la arquitectura técnica existente y planeada de la compañía;
- Tipo y características del negocio: es indispensable tener el conocimiento exacto sobre el tipo de negocios de la organización y el soporte que representa la información dentro de todo su proceso de toma de decisiones;

- Entorno técnico: se debe incluir tanto el aspecto del *hardware* así como aplicaciones y herramientas. Se dará énfasis a los sistemas de soporte a decisiones, si existen en la actualidad, cómo operan, etc;
- Expectativas de los usuarios: un proyecto de almacén de datos no es únicamente un proyecto tecnológico, es una forma de vida de las organizaciones y como tal, tiene que contar con el apoyo de todos los usuarios y su convencimiento sobre su bondad;
- Etapas de desarrollo: con el conocimiento previo, ya se entra en el desarrollo de un modelo conceptual para la construcción del almacén de datos;
- Prototipo: un prototipo es un esfuerzo designado a simular tanto como sea posible el producto final que será entregado a los usuarios;
- Piloto: el piloto de un almacén de datos es el primer resultado generado de forma iterativa que se harán para llegar a la construcción del producto final deseado;
- Prueba del concepto tecnológico: es un paso opcional que se puede necesitar para determinar si la arquitectura especificada del almacén de datos funcionará finalmente como se espera.

1.8. Almacén de datos espacial

Un almacén de datos espacial es una colección de datos orientados al tema, integrados, no volátiles, variantes en el tiempo y que añaden la geografía de los datos, para la toma de decisiones. Sin embargo, la componente geográfica no es un dato agregado, sino que una dimensión o variable en la tecnología de la información, de tal manera que permita modelar todo el negocio como un ente holístico, y que a través de herramientas de procesamiento analítico en línea OLAP, no solamente se posea un alto desempeño en consultas

multidimensionales, sino que adicionalmente se puedan visualizar espacialmente los resultados.

El almacén de datos espacial forma el núcleo de un extensivo sistema de información geográfica para la toma de decisiones, éste al igual que los GIS⁷, permiten que un gran número de usuarios accedan a información integrada, a diferencia de un simple almacén de datos que está orientado al tema, el almacén de datos espacial [18] adicionalmente es geo-relacional, es decir que en estructuras relacionales combina e integra los datos espaciales con los datos descriptivos. Actualmente es geo-objetos, esto es que los elementos geográficos se manifiestan como objetos con todas sus propiedades y comportamientos, y que adicionalmente están almacenados en una única base de datos Objeto-Relacional.

Los almacenes de datos espaciales son aplicaciones basadas en un alto desempeño de las bases de datos, que utilizan arquitecturas cliente-servidor para integrar diversos datos en tiempo real. Mientras los almacenes de datos trabajan con muchos tipos y dimensiones de datos, muchos de los cuales no referencian ubicación espacial, a pesar de poseerla intrínsecamente, y sabiendo que un 80% de los datos poseen representación y ubicación en el espacio, en los almacenes de datos espaciales, la variable geográfica desempeña un papel importante en la base de información para la construcción del análisis, y de igual manera que para un almacén de datos, la variable tiempo es imprescindible en los análisis, para los almacenes de datos espaciales la variable geográfica debe ser almacenada directamente en ella.

⁷ *Geographic Information Systems*

1.9. Ventajas

Existen muchas ventajas por las cuales es recomendable usar un almacén de datos. Algunas de ellas son:

- Los almacenes de datos hacen más fácil el acceso a una gran variedad de datos a los usuarios finales;
- Facilitan el funcionamiento de las aplicaciones de los sistemas de apoyo a la decisión tales como informes de tendencia;
- Los almacenes de datos pueden trabajar en conjunto y, por lo tanto, aumentar el valor operacional de las aplicaciones empresariales, en especial la gestión de relaciones con clientes.

1.10. Desventajas

Utilizar almacenes de datos también plantea algunos inconvenientes, algunos de ellos son:

- A lo largo de su vida los almacenes de datos pueden suponer altos costos. El almacén de datos no suele ser estático. Los costos de mantenimiento son elevados;
- Los almacenes de datos se pueden quedar obsoletos relativamente pronto
- A veces, ante una petición de información éstos devuelven una información de baja calidad, que también supone una pérdida para la organización;
- A menudo existe una delgada línea entre los almacenes de datos y sistemas operativos. Hay que determinar qué funcionalidades de éstos se pueden aprovechar y cuáles se deben implementar en el almacén de

datos, resultaría costoso implementar operaciones no necesarias o dejar de implementar alguna que sí vaya a necesitarse.

2. SITUACIÓN ACTUAL

2.1. Entendimiento de la problemática

Guatemala es un país que por la característica de sus recursos naturales tiene un enorme potencial para el desarrollo sostenible de su gente a partir del agua. Las Naciones Unidas proponen dentro de los objetivos del milenio para Guatemala [19], específicamente el objetivo número 7, "Garantizar la sostenibilidad del medio ambiente", meta numero 10: "Proporción de la población con acceso sostenible a mejores fuentes de abastecimiento de agua potable", las instituciones encargadas de velar por que se cumpla esta meta son la OMS y UNICEF, en el informe presentado donde se dice que el 78% de los hogares guatemaltecos posee cobertura de agua potable. Dicho indicador puede parecer positivo, pero en realidad no se mide el tiempo en que la población realmente tiene el acceso a esta fuente de agua [20].

Se estima que en Guatemala los cuerpos de agua poseen un caudal de agua que totalizan 3,190 metros cúbicos por segundo (ver tabla I). Equivalente a 84,991 millones de metros cúbicos de agua al año, para fines de comparación, este volumen de agua equivale a 300 veces al lago de Amatitlán [21].

Tabla I. **Oferta Hídrica de Guatemala**

Vertiente	Área km²	Porcentaje respecto al país	Cuencas (número)	Caudal promedio (m³/seg.)
Pacífico	24,237	22	18	808
Caribe	33,799	31	10	1010
Golfo de México	50,852	47	10	1372
Total	108,889	100	38	3,190

Fuente: elaboración propia. Datos tomados de <http://www.infoiarna.org.gt/guateagua/>, febrero 2009.

La principal fuente de agua es la precipitación pluvial. Las lluvias en Guatemala están fuertemente influenciadas por las cadenas montañosas, así la Sierra Madre delimita un sistema de laderas que corren paralelas al litoral del pacífico. Las cordilleras montañosas del Norte delimitan otro sistema de laderas, la denominada Franja Transversal del Norte. Estas laderas actúan como un mecanismo de ascenso para los vientos cargados de humedad provenientes del mar Caribe y el golfo de México.

En estas regiones la lluvia se mezcla con la temporada de frentes o de olas de frío. En la parte posterior de estas laderas, en donde están las depresiones formadas por las cuencas de los ríos Cuilco, Chixoy y Motagua, se produce un fuerte efecto de sombra pluviométrica, que hace que en esta zona se registren los menores volúmenes de lluvia. La precipitación promedio anual en el país es de aproximadamente 2,000 mm, con variaciones que van desde 500 mm en las regiones secas del Oriente⁸, hasta 5,600 mm en la zona Norte y Occidente⁹.

⁸ Jalapa, Jutiapa, Chiquimula y Zacapa

⁹ Huehuetenango, Quiché, Alta Verapaz y Baja Verapaz

Entre los ríos más caudalosos del país destacan el Usumacinta con un caudal aproximado de 1,800 m³/seg; el Motagua con 240 m³/seg; el Sarstún con 172 m³/seg; el Suchiate con 28 m³/seg en la Costa Sur. Sin embargo, el 55% del territorio guatemalteco forma parte de cuencas internacionales, y del caudal nacional se estima que el 47.5% drena hacia México, el 7% hacia El Salvador, el 6% hacia Belice y el 0.5% hacia Honduras [22]. En el caso de las aguas subterráneas se estima que el potencial de agua es de 33,699 millones de metros cúbicos. Siendo los acuíferos de la Costa del Pacífico los de mayor rendimiento [23].

Existe poca información actualizada sobre estadísticas de los diferentes usos del agua. Es por ello que se recurre a estimaciones gruesas de los usos más importantes del recurso hídrico en el país. El siguiente cuadro resume las estimaciones del uso actual y a futuro del recurso agua en Guatemala.

Tabla II. **Uso del agua en Guatemala**

Demanda del Agua	Cantidad usada al año 2005 (millones de m³)	Proyección de uso al año 2025 (millones de m³)
Agua Potable	284	1,211
Riego	2,200	10,200
Industria	850	3,625
Energía	2,883	15,000
Usos Total	6,217	30,036
Excedente/Déficit Hídrico	23,530	(289)

Fuente: elaboración propia. Datos tomados de <http://www.infoiarna.org.gt/guateagua/>, febrero de 2009.

Se estima que el consumo de agua potable se aproxima a los 284 millones de metros cúbicos anuales lo cual representa el 1% del agua que tiene el país, aunque las proyecciones para el año 2025 indican un crecimiento hacia el 4%, es decir 1,211 millones de metros cúbicos por año. En la actualidad el mayor consumidor de agua lo representa el sector agrícola con un 6%. El suministro de agua potable de las 331 municipalidades, se abastece del 70% con aguas superficiales y 30% con aguas subterráneas, un 66% usa sistemas de gravedad, 18.5% utilizan bombeo y 15.2% son sistemas mixtos [24].

En riego se estima que se utilizan 2,200 millones de metros cúbicos anuales, basándose en un consumo promedio equivalente a 1.10 litros por segundo por hectárea. Para producción de energía, se estima que se usan 2,283 millones de metros cúbicos de agua anuales aunque el agua después de pasar por la turbina queda disponible nuevamente para su uso. El 51% de la potencia eléctrica instalada en el país es generado por hidroeléctricas [25].

En la industria no existen datos confiables sobre el uso del agua por este sector. Se estima una utilización de 425 millones de metros cúbicos (80% del volumen producido) en el proceso de destilación, rectificación y mezcla de bebidas espirituosas, productos vinícolas, fabricación de cerveza y la producción de bebidas gaseosas, jugos y otras. Es importante hacer notar que hay otros usos considerables de agua en la industria alimenticia, así como la limpieza de equipo e instalaciones, la alimentación de calderas, y el agua utilizada para refrigeración y enfriamiento.

Existen otros usos como pesca, turismo, transporte acuático, que aunque no consumen agua directamente, si requieren del uso del recurso en cantidad y calidad. Además el agua también es un receptor de desechos, siendo este uso el más nocivo, porque se reduce la disponibilidad futura del recurso.

La temporalidad se produce porque en Guatemala tenemos dos estaciones muy marcadas: la lluviosa y la seca. La lluvia se concentra en los meses de junio y septiembre, con una canícula o período de menor precipitación, entre julio y agosto. En las regiones secas, la estación sin lluvias es de seis meses, que comprenden de noviembre a abril, mientras que para las regiones más húmedas, se reduce a dos o tres meses [26]. La agricultura, el principal usuario del agua en términos de volumen, requiere del agua precisamente en la época seca.

En el caso del agua contaminada, sólo por descargas industriales y municipales, el volumen es equivalente al de 120 veces el volumen del lago de Amatitlán contaminados por año.

En la actualidad se han realizado varios esfuerzos en Guatemala para realizar el sistema del inventario del agua, desde una perspectiva sectorial, es decir, con una visión parcial de la información del agua según el ámbito institucional (léase ambiente, agricultura, energía, salud, saneamiento, etc.) lo que se busca es tener un sistema integrado y centralizado porque aporta la base de conocimiento necesaria para la toma de decisiones y publicación de la misma.

Entre las entidades que han realizado algún tipo de esfuerzo, se pueden mencionar El Ministerio de Ambiente y Recursos Naturales MARN, el Ministerio de Agricultura Ganadería y Alimentación MAGA, el Instituto Nacional de Sismología, Vulcanología, Meteorología e Hidrología INSIVUMEH, el Consejo Nacional de Áreas Protegidas CONAP, la Universidad Rafael Landívar, el Ministerio de Salud Pública y Asistencia Social MSPAS, entre otros.

El Gobierno Central de Guatemala reconoce la importancia del agua en el desarrollo sostenible del país, creando el Gabinete Específico del Agua (GEA)¹⁰, para coordinar los esfuerzos gubernamentales en la temática del agua.

Se pretende que el Estado Guatemalteco recupere su papel rector en el tema del agua, y para esto se han lanzado seis medidas de carácter político gubernamental establecidas en el Gabinete del Agua en el año 2008. Dentro de la medida “Agua e Información”, se tiene contemplado coordinar el diseño y operación del “Sistema Integrado de Información y Conocimiento del Agua de Guatemala (SIAGua)” con participación multisectorial.

Se han planteado como objetivos específicos de la medida “Agua e Información” los siguientes:

- a) Coordinar la mejora de la red guatemalteca de monitoreo del agua;
- b) Coordinar el Inventario Nacional del Agua;
- c) Coordinar la modernización de metodologías de análisis de información del agua;
- d) Coordinar el sistema eficaz y eficiente de acceso a la información y conocimiento del agua.

El inventario nacional del agua con una perspectiva integral, como la aportada desde la SEGEPLAN¹¹, es un tema fundamental para avanzar hacia la gestión integrada del agua, porque provee ciertos conocimientos necesarios para la toma de decisiones. Y como parte de esa visión integral también es fundamental que dicho inventario de aguas esté dentro de la lógica del Sistema Nacional de Información del Agua de Guatemala. En este sentido, es necesario

¹⁰ Véase anexo A.

¹¹ Secretaría de Planificación y Programación de la Presidencia.

que dicho sistema nacional de información del agua sea coherente con los avances tecnológicos de punta y de acceso a la información que la sociedad guatemalteca demanda.

Actualmente la Secretaria Técnica del Gabinete Específico del Agua, que está a cargo de la dirección de recursos hídricos de La Secretaría de Planificación y Programación de la Presidencia, tiene a su cargo la coordinación del trabajo de todas las comisiones que se integren dentro del gabinete para realizar el Sistema Nacional de Información y Conocimiento del Agua “SIAGua”

El gabinete del agua tiene definidas seis medidas gubernamentales, pero para el ámbito de este proyecto, únicamente nos interesa la medida: “Agua e Información”. Acorde con los objetivos de dicha medida “Agua e Información”, algunas de las acciones clave se pueden interpretar desde la lógica de los sistemas de información. Dentro de esas acciones se pueden señalar las siguientes:

- Recolectar la información: recabando toda la información existente que se encuentra en distintas instituciones, y distintos formatos, para unificarla e identificar los vacíos existentes y que es necesario generar en campo y gabinete en acciones futuras;
- Análisis de la información: este paso es parte importante del sistema nacional de información del agua y es una acción acorde con su tercer objetivo específico; además del análisis de información de carácter experto (hidrológico, económico, ambiental), se busca que tanto este análisis como la información previamente recolectada/generada, pueda ser accesible para la toma de decisiones, por lo que también se debe contar con una herramienta para la generación de distintos reportes en el ámbito de las comunidades, los municipios, los departamentos, las

cuenas hidrográficas (demarcación técnica para la planificación del agua) y a nivel de país;

- Divulgación de la información: haciendo el mejor uso de las tecnologías de la información y la comunicación (TICs). Esto se hace con el propósito de poder realizar distintas publicaciones con los datos obtenidos para los distintos grupos que la requieran, que pueden ser ministros, alcaldes, y población en general; dicha acción es estratégica para dar cumplimiento al Decreto 57-2008 de la República de Guatemala, de libre acceso a la información y con la filosofía de "El poder no lo tiene quien tiene la información, sino quien analiza y genera conocimiento comunitario".

Con un análisis de las estadísticas presentadas al inicio del capítulo se puede notar que la disponibilidad de agua en Guatemala supera el uso actual, basados en la diferencia de la oferta hídrica menos el agua que se consume, se demuestra que actualmente Guatemala podría ser un país exportador de agua. Sin embargo, continuamente se escucha hablar de escasez y falta de agua. El principal objetivo de este proyecto, es realizar el análisis y diseño para una base de datos multidimensional, que pueda almacenar información proveniente de cualquiera de las tres boletas analizadas durante este ejercicio, las cuales son las boletas de PARPA¹², PROVIAGUA¹³ y la boleta SAS de cobertura. El modelo que se propone en el capítulo 3 ayudará a proporcionar la información a las personas o instituciones encargadas de la toma de decisiones que involucran al recurso hídrico, para que se puedan basar en información actualizada y unificada a la hora de tomar las mismas, teniendo un repositorio centralizado e integrado.

¹² Programa de Apoyo a la Reconversión Productiva Agroalimentaria

¹³ Programa Nacional de Vigilancia de la Calidad de Agua para Consumo Humano

2.2. Inventario de recursos

Dentro de las instituciones que se identificaron que cuentan con cierto conocimiento, o tienen parcialmente información del inventario nacional de fuentes de agua se encuentran:

- Dirección General de Límites y Aguas Internacionales
- Ministerio de Salud Pública y Asistencia Social
- Ministerio de Agricultura, Ganadería y Alimentación
- Ministerio de Ambiente y Recursos Naturales
- Secretaría de Seguridad Alimentaria y Nutricional

2.3. Riesgos y contingencias

A continuación podemos encontrar los riesgos que se tomaron en consideración para este ejercicio con su respectiva ponderación, separados en las siguientes categorías:

- Atributos de los requerimientos
- Atributos del producto
- Atributos de la gente
- Atributos de la administración
- Atributos del ambiente
- Atributos del método de desarrollo

2.3.1. Atributos de los requerimientos

Mínimo	Complejidad
	¿Es el requerimiento excesivamente complejo?
	¿Existen diferentes partes del proyecto que interactúen de maneras indeseadas?
	¿Existen restricciones con el tiempo de entrega?
Mínimo	Dificultad
	¿Existen algunas partes de los requerimientos difíciles de realizar?
Mínimo	Viabilidad
	¿Se tiene conocimiento de aquellas partes de los requerimientos que pueden ser realizadas?
	¿Existen partes de los requerimientos que se encuentran más allá de la tecnología actual?
Mínimo	Novedad
	¿El desarrollador del proyecto tiene experiencia previa implementando este tipo de sistemas?
Mínimo	Verificabilidad
	¿Es difícil de verificar que se han realizado ciertas partes de los requerimientos?
	¿Existen partes de los requerimientos que no son cuantificables o subjetivas?
Mínimo	Volatilidad
	¿Los requerimientos cambian constantemente, o son nuevos requerimientos agregados frecuentemente?
	¿Es probable que los requerimientos continúen cambiando durante la implementación?

2.3.2. Atributos del producto

No aplica	Integración
	¿Existen un plan para la integración de los subsistemas en un producto completo?
	¿Están bien definidas las relaciones con los sistemas involucrados?
Mínimo	Mantenimiento
	¿Serán las partes del producto difíciles de modificar?
	¿Existen límites en el posible crecimiento del producto?
Moderado	Desempeño
	¿Es posible que el desempeño sea inapropiado?
	¿Existen límites en el desempeño del producto?
Mínimo	Confiabilidad
	¿Existen dudas sobre la confiabilidad del sistema o de una parte crítica?
Mínimo	Seguridad
	¿Existen dudas acerca de la seguridad del sistema o de una parte crítica?
	¿Están todos los componentes críticos de seguridad claramente identificados?
Mínimo	Escala
	¿Es el producto significativamente más grande de lo que cualquier equipo ha construido?
Mínimo	Comprobabilidad
	¿Será difícil de probar el sistema, o cualquier componente?
	¿Existe un plan de pruebas para cada componente, especificación y función?
No aplica	Usabilidad
	¿Será el producto difícil de usar en el ambiente de operación para los usuarios finales?
	¿Existe un plan para realizar un prototipo, evaluar o probar en el campo la interfaz del usuario?
Moderado	Criticidad
	¿Es la aplicación crítica para el negocio del cliente?

2.3.3. Atributos de la gente

Mayor	Comunicación
	¿Existe falta de comunicación entre los grupos funcionales?
	¿Existe falta de comunicación entre los trabajadores y el administrador?
	¿Es posible que la cantidad de grupos afectados obstaculice los esfuerzos de comunicación?
Mínimo	Cooperación
	¿Existen grupos que no valoran el rol y la contribución de otros?
	¿Existen en el ambiente de trabajo actitudes y comportamientos de falta de colaboración?
Mínimo	Conocimiento del ámbito
	¿El líder del proyecto o el equipo en general no tienen el conocimiento suficiente del ámbito del proyecto?
	¿Son muy pocos los miembros en el equipo que han utilizado un producto similar?
Mínimo	Experiencia
	¿Son el proceso de desarrollo, las herramientas o métodos desconocidos para el equipo?
	¿Son la mayoría de los miembros del equipo nuevos para la organización?
Mayor	Conocimiento técnico
	¿El equipo no tiene conocimiento suficiente de la solución o alguno de los componentes?
	¿Tiene el equipo las habilidades clave?

2.3.4. Atributos de la administración

Mínimo	Cooperación
	¿Hay ocasiones en que el proceso de administración impide cooperación?
	¿EL proceso de administración hace difícil la libre comunicación con los administradores?
Mínimo	Relaciones con el cliente
	¿Tiene la administración suficiente acceso a todos los niveles de clientes?
	¿El cliente actúa algunas veces de una manera que impide el proceso de administración?
Mayor	Experiencia
	¿Es el administrador inexperimentado con la estructura de administración?
	¿Es el administrador inexperimentado con el proceso de desarrollo?
Mínimo	Habilidades humanas
	¿Tiene el administrador problemas para manejar asuntos personales del personal?
	¿Algunas veces los buenos trabajadores no son reconocidos y recompensados?
Mínimo	Proceso de administración
	¿Es el proceso de administración no permite desarrollar habilidades personales?
	¿Permite el proceso de administración el surgimiento y discusión de problemas?
	¿Es el proceso muy lento para la respuesta de eventos inesperados?
	¿Permite el proceso abandonar políticas o prácticas inapropiadas o ineficientes?
Mínimo	Estructura
	¿Es difícil de decir que administrador es responsable por hacer las decisiones críticas?
	¿Los administradores se encuentran separados geográficamente de los sitios claves de trabajo?

2.3.5. Atributos del ambiente

Mínimo	Presupuesto
	¿Es el presupuesto inadecuado para las tareas requeridas?
	¿Es el presupuesto cambiado frecuentemente sin importar el trabajo que se ha realizado?
Mínimo	Restricciones externas
	¿Contiene el sistema componentes externos cuya exactitud es incierta?
	¿El proceso de desarrollo requiere que se usen procedimientos o herramientas no apropiadas?
	¿Existen fechas límites que restrinjan el proyecto?
Moderado	Políticas
	¿Existen interferencias con detalles del proyecto provenientes de Fuentes externas de control?
	¿Existen decisiones impuestas al proyecto por razones políticas no importando sacrificar la calidad?
Mayor	Recursos
	¿Son los recursos inadecuados para la tarea requerida?
Mínimo	Calendarización
	¿Es la calendarización del desarrollo realista?
	¿Existen muchos traslapes o paralelismo en la calendarización?

2.3.6. Atributos del método de desarrollo

Moderado	Capacidad
	¿Hay dudas de que el sistema de desarrollo pueda hacer frente a un producto de este tamaño?
	¿Puede ser que el sistema de desarrollo sea demasiado lento ante la carga esperada?
Moderado	Documentación
	¿Existe muy poca documentación del sistema de desarrollo?
	¿Existen herramientas o componentes que estén escasamente documentados?
Mínimo	Conocimiento
	¿El equipo de desarrollo es familiar al método de desarrollo?
	¿En ocasiones, se pierde el trabajo por errores en el método de desarrollo?
No aplica	Robustez
	¿El sistema de desarrollo se encuentra muchas veces fuera de línea debido a fallas del equipo?
	¿En ocasiones, se pierde el trabajo por errores en el método de desarrollo?
No aplica	Soporte de herramientas
	¿Es el soporte para las herramientas utilizadas el adecuado?
	¿Se realiza demasiado esfuerzo manual en tareas de rutina o en tareas de oficina?
No aplica	Usabilidad
	¿Es el método de desarrollo demasiado difícil de usar?
	¿Es fácil cometer errores y muy difícil de corregirlos?

2.4. Entendimiento de los datos

A continuación podemos encontrar los riesgos que se tomaron en consideración para este ejercicio con su respectiva ponderación, separados en las siguientes categorías:

2.4.1. Colección Inicial de datos

Para la colección inicial de datos se utilizaron los obtenidos de las boletas PARPA, y SAS, provenientes de los departamentos de Chiquimula, Huehuetenango, Quiché y San Marcos. En total se obtuvieron datos de noventa y dos municipios.

Chiquimula		
Camotán	Chiquimula	Concepción Las Minas
Esquipulas	Ipala	Jocotán
Olopa	Quezaltepeque	San Jacinto
San José La Arada	San Juan La Ermita	

Huehuetenango		
Aguacatán	Barillas	Chiantla
Colotenango	Concepción Huista	Cuilco
Huehuetenango	Ixtahuacán	Jacaltenango
La Democracia	La Libertad	Malacantancito
Nentón	San Antonio Huista	San Gaspar Ixchil
San Juan Atitán	San Juan Ixcoy	San Mateo Ixtatán
San Miguel Acatán	San Pedro Necta	San Rafael La Independencia

Continúa tabla Huehuetenango.

San Rafael Petzal	San Sebastián	San Sebastián Coatán
Santa Ana Huista	Santa Bárbara	Santa Eulalia
Santiago Chimaltenango	Soloma	Tectitán
Todos Santos		

Quiché		
Canillá	Chajul	Chicamán
Chiché	Chichicastenango	Chinique
Cunén	Ixcán	Joyabaj
Nebaj	Pachalum	Patzité
Sacapulas	San Andrés Sajcabajá	San Antonio Ilotenango
San Bartolomé	San Juan Cotzal	San Pedro Jocopilas
Santa Cruz del Quiché	Uspantán	Zacualpa

San Marcos		
Ayutla	Catarina	Comitancillo
Concepción Tutuapa	El Quetzal	El Rodeo
El Tumbador	Esquipulas Palo Gordo	Ixchiguán
La Reforma	Malacatán	Nuevo Progreso
Ocos	Pajapita	Río Blanco
San Antonio	San Cristóbal Cucho	San José Ojetenán
San Lorenzo	San Marcos	San Miguel Ixtahuacán
San Pablo	San Pedro Sacatepéquez	San Rafael Pie de la Cuesta
Sibinal	Sipacapa	Tacaná
Tajumulco	Tejutla	

2.4.2. Boletas

Para el ámbito de este ejercicio se puede resumir el inventario de fuentes de agua, en tres boletas, las cuales son utilizadas por las diferentes instituciones para recabar datos, estas boletas son la boleta PARPA (Programa de Apoyo a la Reconversión Productiva Agroalimentaria) de catastro de fuentes de agua, SAS de cobertura, y la de PROVIAGUA, a continuación se describe con mayor detalle cada una de las secciones de las distintas boletas.

2.4.2.1. Boleta PARPA¹⁴

Encabezado

Nombre del encuestador: llenar con el nombre completo de la persona que está recopilando la información.

Institución responsable: institución a la que pertenece la persona que está recopilando la información.

Fecha: fecha en la que se llenó la boleta en formato DD/MM/AA.

Hora de inicio: hora en la que se inicio a llenar la boleta.

Departamento/unidad: departamento o unidad a la que pertenece la persona dentro de la institución.

¹⁴Véase anexo D.

Sección A

Parte I: datos generales

Departamento: nombre del departamento donde está ubicada la fuente de agua.

Municipio: nombre del municipio donde está ubicada la fuente de agua.

Comunidad/finca: nombre de la comunidad o finca donde la fuente está ubicada.

Sección B

Parte I: ubicación de la fuente

Coordenadas UTM:

X: coordenada X proporcionada por el GPS en la ubicación de la fuente.

Y: coordenada Y proporcionada por el GPS en la ubicación de la fuente.

Modelo de GPS: modelo del GPS utilizado durante las mediciones.

Altitud: altitud de la fuente proporcionada por el GPS o un altímetro en la ubicación de la fuente, indicado en metros sobre el nivel del mar (msnm).

Nombre de la fuente: indicar si la fuente de agua es conocida por algún nombre en específico.

Parte II: datos del tipo de fuente

Tipo de la fuente: seleccionar el tipo de la fuente, de las opciones siguientes:

- Manantial
- Nacimiento
- Río
- Lago
- Pozo perforado con maquina
- Laguna
- Pozo excavado
- Embalse

Método de aforo: se refiere a cómo el caudal de agua es medido en la fuente, se debe seleccionar de las siguientes opciones:

- Flotador
- Molinete
- Cubeta
- Otros

Distancia: distancia medida a la hora de realizar el aforo, indicada en metros.

Ancho: ancho medido a la hora de realizar el aforo, indicada en metros.

Tiempo: tiempo transcurrido entre las distintas mediciones.

Profundidad: profundidad medida a la hora de realizar el aforo, indicado en metros.

Cubeta: chequear esta casilla si el método de aforo fue realizado con cubeta.

Tiempo: tiempo transcurrido entre las distintas mediciones. Seleccionar minutos o segundos.

Volumen: volumen capturado por la cubeta, seleccionar galones o litros.

Observaciones: anotar aquí cualquier observación general que realizó durante el trabajo de campo, que crea relevante dar a conocer.

Caudal promedio: calcular y anotar el valor del caudal promedio en las mediciones, seleccionar la dimensional correspondiente m^3/seg , Lt/seg, GPM¹⁵.

Tipo de usuario: seleccionar el tipo de usuario que utilizan la fuente de las siguientes opciones:

- Privado
- Particular
- Comunal
- Municipal
- Estatal

¹⁵ Galones por minuto

Estructura de protección en la fuente: detallar si la fuente cuenta con algún tipo de estructura de protección.

¿Tiene cobertura vegetal?: seleccionar si la fuente tiene o no cobertura vegetal.

¿Actualmente la fuente está en uso?: seleccionar si la fuente se encuentra en uso o no.

Estado legal de la fuente: seleccionar el estado legal en que se encuentra la fuente, de las siguientes opciones:

- Escritura pública
- Escritura registrada
- Acta municipal
- Convenios verbales

Condiciones de la fuente: seleccionar el estado actual de la fuente, de las siguientes opciones

- Buena
- Regular
- Mala

Parte III: si la fuente es un pozo mecánico

Volumen de extracción: indicar la capacidad que tiene el pozo mecánico para extraer agua, indicar dimensional.

Potencia de la bomba: indicar la potencia que tiene la bomba en caballos de fuerza (HP).

Horas de uso: número de horas al día que se usa la bomba para extraer agua.

La bomba es: seleccionar el tipo de bomba de las siguientes opciones:

- Eléctrica
- Gasolina
- Diesel

Sección C

Tomas y caudales

Tipo de captación: se refiere a cómo es capturada el agua desde la fuente, seleccionar de las siguientes opciones:

- Embalse
- Toma
- Caja
- Tanque

Cuadro No. 1: éste tiene tres secciones con los siguientes campos, para tres tomas de agua de la fuente.

Toma #: correlativo asignado a la fuente toma medida.

Coordenadas de salida:

X: coordenada X que marca el GPS en el punto donde se toma el agua.

Y: coordenada Y que marca el GPS en el punto donde se toma el agua.

Elevación de la salida: altitud del punto donde se toma el agua.
Medido en msnm.

Tipo de sección: seleccionar entre las siguientes opciones:

- Tubería
- Canal semicircular
- Canal rectangular
- Canal trapezoidal

Materiales: materiales usados para la toma de agua.

Observaciones: anotar aquí cualquier observación general que realizó durante el trabajo de campo, que crea relevante dar a conocer.

Sección D

Uso doméstico

Nombre de la población beneficiada: indicar el nombre de la población que se beneficia del agua de la fuente.

Número de población beneficiada: indicar el número de personas beneficiadas en la población.

Tipo de servicio: seleccionar el tipo de servicio que se presta con el agua obtenida de la fuente, puede ser:

- Domiciliar
- Predial
- Publica

¿Se descarga el agua utilizada?: indicar si se descarga el agua utilizada o no.

Coordenadas UTM de dónde descarga:

X: coordenada X que marca el GPS en el punto donde se descarga el agua

Y: coordenada Y que marca el GPS en el punto donde se descarga el agua

¿Tiene planta de tratamiento?: indicar si al descargar el agua pasa por una planta de tratamiento al ser descargada.

Sección E

Uso irrigación

Nombre del usuario: nombre de la persona que usa el agua de la fuente para irrigar.

Área total de riego: área superficial que se irriga.

Unidades: dimensional utilizada al calcular el área superficial.

Tipo de riego utilizado: seleccionar el método de riego que se realiza con el agua de las opciones siguientes:

- Cubeta
- Manguera
- Goteo
- Gravedad
- Aspersión
- Otro, si la opción seleccionada es esta, especifique: si es algún otro método de riego el que se realiza.

Cuadro No. 2: éste tiene 2 secciones como las siguientes:

Nombre (s) cultivo(s): indica el nombre del cultivo que se riega

Período de riego anual: indica el número de meses al año que se riega dicho cultivo.

Turnos de riego: indica los turnos de riegos si existiesen.

Duración: duración del riego, medido en horas al día.

Frecuencia: frecuencia con que se realiza el riego, cada cuantos días, semanas, etc., durante los meses del periodo de riego anual.

Sección F

Uso hidroeléctrica

Nombre del usuario: nombre de la persona que usa el agua de la fuente para generar electricidad.

Potencia de generación: potencia la cual se puede generar con el agua proveniente de esta fuente de agua.

Caudal de generación: caudal necesario para la potencia de generación indicada en el inciso anterior.

¿Existe curva de descarga?: indica si existe curva de descarga o no.

Tipo de operación: indica el tipo de operación, las opciones son:

- Filo de agua
- Embalse horario
- Diario
- Mensual
- Anual

¿Usa embalse?: indicar si se usa embalse o no para la generación de energía eléctrica. Si su respuesta es afirmativa, conteste lo siguiente:

Volumen del embalse: volumen de agua que puede contener el embalse.

Altura de la presa: altura del muro de la presa.

Longitud del embalse: longitud en la parte superior del muro de la presa.

Sección G

Uso industrial, recreativo, pecuario y otros

Nombre del usuario: nombre de la persona que usa el agua de la fuente.

Volumen de agua utilizado en el proceso: volumen de agua que se usa, indicar también la dimensional de volumen utilizada.

Período de tiempo que el agua es utilizada: número de horas al día en las que se hace uso del agua.

Época del año que es utilizada el agua: indica los meses del año que el agua es utilizada. Seleccionar el campo “todo el año” si el agua es utilizada durante todo el año.

¿Se descarga el agua utilizada?: indica si el agua utilizada es descargada o no.

Coordenadas UTM de donde descarga:

X: coordenada X que marca el GPS en el punto donde se descarga el agua.

Y: coordenada Y que marca el GPS en el punto donde se descarga el agua

¿Tiene planta de tratamiento?: indicar si al descargar el agua pasa por una planta de tratamiento al ser descargada.

Tipo de ganado: seleccionar el tipo de ganado que se beneficia del agua de la fuente, puede ser:

- Vacuno
- Porcino
- Ovino
- Aviar
- Caballar
- Otro: si la respuesta es otro especifique

Número de animales: número de cabezas de los animales beneficiados por el agua de la fuente.

Hora final: indica la hora en que se terminó de recolectar la información.

Tiempo: tiempo que se utilizó para recolectar la información.

Código de fotografía: si existe una fotografía, indicar el código con el cual se identifica.

2.4.2.2. Boleta PROVIAGUA¹⁶

Boleta para identificar muestras de agua

Fecha toma de muestra: indica la fecha en la que la muestra de agua fue recolectada.

¹⁶ Véase anexo C.

Hora toma de muestra: indica la hora en la que la muestra de agua fue recolectada.

Fecha y hora de llegada al laboratorio: indica la fecha y hora en que la muestra llegó al laboratorio, la cual no es igual a la fecha de toma de muestra, por diversas razones.

Número de muestra: correlativo asignado a la muestra.

Nombre de quién tomó la muestra: llenar con el nombre completo de la persona que está recopilando la información.

Resultado de laboratorio: indica si la muestra está contaminada o no.

Tipo de fuente: indica el tipo de fuente de donde la muestra fue tomada. Se debe seleccionar de las siguientes opciones:

- Manantial captado el ojo
- Pozo
- Lago/laguna
- Río/riachuelo
- Quebrada
- Manantial captado aguas abajo
- Canal de riego
- Red pública
- Otro: si es otro, indicar en el campo de texto

Lugar de muestreo: indica el lugar donde fue realizado el muestreo en la fuente de agua, este puede ser:

- Tanque de almacenamiento
- Caja de reunión
- Conexión domiciliar
- Otro: si es otro, indicar en el campo de texto

Sistema de desinfección: indica si el sistema cuenta con sistema de cloración o no.

Método de cloración utilizado: identifica el método de cloración utilizado si existiese, estos métodos pueden ser:

- Cloro gas
- Hipoclorito de sodio
- Otro: si es otro, indicar en el campo de texto

¿Cuentan con stock de cloro?: indica si se cuenta con inventario de cloro en el lugar de la toma de la muestra para el tratamiento de la misma.

Procedencia: sección que indica la ubicación geográfica de la procedencia del agua.

Centro poblado: centro poblado de donde proviene el agua.

Sector: sector de donde proviene el agua.

Municipio: nombre del municipio de donde proviene el agua.

Departamento: nombre del departamento de donde proviene el agua.

Comunidad: nombre de la comunidad de donde proviene el agua.

¿Se cobra por el servicio?: indica si se cobra por el servicio del agua o no.

Tarifa mensual: en caso afirmativo a la pregunta anterior, indica el monto que se cobra por el servicio del agua.

Quién cobra por el servicio: en caso afirmativo, indica quién cobra por el servicio del agua.

2.4.2.3. Boleta por comunidad "PROVIAGUA-SIS-001"¹⁷

Fecha de la inspección: indica la fecha en que se realizó la inspección.

Responsable: sección con los datos de la persona que realizó la inspección.

Nombre: nombre completo de la persona que realizó la inspección.

Cargo: cargo que ocupa dentro de la institución responsable de la institución.

Servicio de salud: servicio de salud al que pertenece la persona que realizó la inspección.

¹⁷Véase anexo C.

Comunidad: sección con los datos de la comunidad donde se realizó la inspección. Está compuesta de los siguientes campos:

Nombre: indica el nombre de la comunidad.

Tipo: en este campo se describe el tipo de la comunidad inspeccionada, por ejemplo aldea, caserío, etc.

Área de salud: área de salud a la que compete la inspección.

Municipio: nombre del municipio al que pertenece la comunidad.

Población aproximada: número aproximado de personas en la comunidad.

Número aproximado de familias: número aproximado de familias en la comunidad.

¿Existen sistemas formales?: indica si existen sistemas formales de abastecimiento en la comunidad o no.

¿Existen sistemas informales?: indica si existen sistemas informales en la comunidad o no.

Sistemas formales de abastecimiento

Nombre: nombre con que se identifica al sistema formal de abastecimiento.

Tipo: indica el tipo del sistema, puede ser:

- Gravedad
- Bombeo
- Mixto

Sistemas informales de abastecimiento

Nombre: nombre con que se identifica al sistema informal de abastecimiento.

Tipo: indica el tipo del sistema, puede ser:

- Gravedad
- Bombeo
- Mixto

2.4.2.4. Boleta por sistema "PROVIAGUA-SIS-001"¹⁸

Fecha de la Inspección: indica la fecha en que se realizó la inspección.

Responsable: sección con los datos de la persona que realizó la inspección.

Nombre: nombre completo de la persona que realizó la inspección.

¹⁸Véase anexo C.

Cargo: cargo que ocupa dentro de la institución responsable de la institución.

Servicio de salud: servicio de salud al que pertenece la persona que realizó la inspección.

Comunidad: sección con los datos de la comunidad donde se realizó la inspección.

Nombre: indica el nombre de la comunidad.

Tipo: en este campo se describe el tipo de la comunidad inspeccionada.

Área de salud: área de salud a la que compete la inspección.

Municipio: nombre del municipio al que pertenece la comunidad.

Población aproximada: número aproximado de personas en la comunidad.

Número aproximado de familias: número aproximado de familias en la comunidad.

Nombre o identificación: nombre con que se identifica al sistema de abastecimiento.

Tipo: indica si el sistema es formal o informal.

Fuente: nombre de la fuente de agua que alimenta el sistema.

Ubicación: ubicación geográfica con la fuente de agua.

¿Desinfectan el agua?: indica si utilizan algún método para desinfectar el agua, puede ser:

- Si
- No
- Parcial

Indique el método utilizado: método utilizado para desinfectar el agua en el sistema de abastecimiento. Si la respuesta a la pregunta anterior es no y/o parcial, indique la razón o descripción de por qué se cree o sabe que el agua no es desinfectada.

Conexiones

Intra-domicilares: indica si existen conexiones intra-domicilares con el sistema de abastecimiento, puede ser sí o no.

Prediales: indica si existen conexiones prediales con el sistema de abastecimiento, puede ser sí o no.

Llena-cántaros: indica si existen conexiones llena-cántaros con el sistema de abastecimiento, puede ser sí o no.

¿Qué entidad fue la responsable?: nombre de la entidad responsable de la construcción del sistema de abastecimiento.

Año de construcción: año en que fue hecha la construcción.

Administrado por: nombre de la persona o entidad encargada de la administración del sistema.

¿Funciona el sistema?: Indica si el sistema funciona correctamente, puede ser sí, no, o irregular.

2.4.2.5. Boleta por calidad "PROVIAGUA-CAL-001"¹⁹

Fecha de la inspección: indica la fecha en que se realizó la inspección.

Responsable: sección con los datos de la persona que realizó la inspección. Está compuesta de los siguientes campos:

Nombre: nombre completo de la persona que realizó la inspección.

Cargo: cargo que ocupa dentro de la institución responsable de la información.

Servicio de salud: servicio de salud al que pertenece la persona que realizó la inspección.

Comunidad: sección con los datos de la comunidad donde se realizó la inspección. Está compuesta de los siguientes campos:

Nombre: indica el nombre de la comunidad.

¹⁹Véase anexo C.

Tipo: en este campo se describe el tipo de la comunidad inspeccionada, por ejemplo aldea, caserío, etc.

Área de salud: área de salud a la que compete la inspección.

Municipio: municipio al que pertenece la comunidad.

Población aproximada: número aproximado de personas en la comunidad.

Número aproximado de familias: número aproximado de familias en la comunidad.

Nombre del sistema de abastecimiento: nombre con que se identifica al sistema de abastecimiento.

Calidad del agua en la fuente: sección con los datos de la calidad del agua que se encontró en la fuente de agua. Está compuesta de los siguientes campos:

Fuente: nombre de la fuente de agua que alimenta el sistema.

Coliformes fecales: número de coliformes fecales encontradas en la muestra, está dada en colonias por 100ml.

Calidad del agua en el sistema: sección con los datos de la calidad del agua que se encontró en el sistema. Está compuesta de los siguientes campos:

Tanque: existen tres mediciones para tres tanques diferentes, este es el identificador para cada tanque, del tanque 1 al tanque 3. Está compuesta de los siguientes campos:

Cloro residual: indica el cloro residual indicado en el tanque, expresado en miligramos por litro.

Coliformes fecales: coliformes fecales encontrados en el tanque, expresado en colonias por 100ml.

Vivienda: existen tres mediciones para tres viviendas diferentes, este es el identificador, de la vivienda 1 a la vivienda 3.

Cloro residual: indica el cloro residual indicado en el tanque, expresado en miligramos por litro.

Coliformes fecales: coliformes fecales encontrados en el tanque, expresado en colonias por 100ml.

Calidad del agua en otros puntos: sección con los datos de la calidad del agua que se encontró en otros puntos que no sean la fuente de agua ni el sistema. Está compuesta de los siguientes campos:

Camión cisterna: existe una medición realizada a un camión cisterna, de cloro residual y coliformes fecales.

Cloro residual: indica el cloro residual indicado en el camión, expresado en miligramos por litro.

Coliformes fecales: coliformes fecales encontrados en el camión, expresado en colonias por 100ml.

Fábrica de Hielo: existe una medición realizada a una fábrica de hielo, de cloro residual y coliformes fecales.

Cloro residual: indica el cloro residual indicado en la fábrica, expresado en miligramos por litro.

Coliformes fecales: coliformes fecales encontrados en la fábrica, expresado en colonias por 100ml.

Envasadora de agua: existe una medición realizada a una envasadora de agua, de cloro residual y coliformes fecales.

Cloro residual: indica el cloro residual indicado en la envasadora, expresado en miligramos por litro.

Coliformes fecales: coliformes fecales encontrados en la envasadora, expresado en colonias por 100ml.

2.4.2.6. Boleta SAS de cobertura²⁰

Encabezado

Nombre: nombre completo de la persona que está recopilando la información.

²⁰ Véase anexo B.

Fecha: fecha en que se realizó el trabajo de campo.

Institución: institución a la que pertenece la persona que realizó el trabajo de campo.

Departamento/unidad: departamento/unidad responsable que está recopilando información.

Datos generales de la comunidad

Comunidad: indicar el nombre del centro poblado en donde se está efectuando la obtención de la información.

Municipio: indicar el nombre del municipio al que pertenece el centro poblado.

Departamento: indicar el nombre del departamento donde se ubica el municipio.

Código INE: código con que identifica el Instituto nacional de estadística a la comunidad.

Coordenadas geográficas de la comunidad

Latitud: latitud que marca el GPS en la fuente de agua.

Longitud: longitud que marca el GPS en la fuente de agua.

Modelo GPS: identificar el modelo del GPS con el que está trabajando.

Datum: indicar el *datum* que está utilizando el GPS. Verificar que sea WGS84.

Altitud: escribir la altura en metros sobre el nivel del mar a la que se encuentra la fuente, ya sea con la elevación que proporcione el GPS o con un altímetro.

Tipo de comunidad

Rural: esta opción debe marcarse si la comunidad se encuentra fuera del área del casco urbano. Si es rural deberá además señalar en la casilla correspondiente si se trata de una aldea, caserío, parajes, fincas, otros.

Urbana: esta opción debe marcarse si el sistema de agua se encuentra en el casco urbano.

Marginal: es cuando la población está en el límite de la zona urbana pero con característica urbano/marginal.

Demografía de la comunidad

Habitantes: indicar el número total de personas que habitan en la comunidad en donde se está obteniendo la información.

Total de viviendas: número de viviendas que se encuentran en la comunidad en donde se está obteniendo la información.

Idioma predominante de la comunidad: el idioma que más habla la gente de la comunidad en donde se está obteniendo la información.

Energía eléctrica de la comunidad: indique si hay energía eléctrica en la comunidad en donde se está obteniendo la información.

Sistema de abastecimiento de agua: campo que indica si la comunidad tiene un sistema de abastecimiento de agua o no. Si cuenta con sistema de agua. En este caso, debe además indicar el nombre de la institución responsable o persona individual que construyó el sistema y el año de dicha construcción. Si la comunidad no cuenta con ningún sistema de agua debe indicar cómo se abastece de agua la comunidad.

Tipo de sistema: Indicar si el sistema funciona por:

- Gravedad: significa que el sistema que funciona aprovechando la pendiente del terreno.
- Bombeo: el sistema que funciona con bombas eléctricas, diesel o gasolina.
- Mixta: el sistema que funciona con la combinación de las dos descripciones anteriores.
- Otros: por favor especificar qué tipo de sistema se utiliza.

Cantidad y tipo de conexiones de agua

Número de conexiones: indicar para cada tipo de conexión el número de viviendas con conexión y con agua y el número de viviendas que no cuentan con ningún servicio, según su tipo:

- Domiciliar
- Predial

- Llena cántaro
- Otro

Funcionamiento del sistema de abastecimiento de agua

Continuidad del servicio: señalar las características del horario de abastecimiento de agua en el lugar de las siguientes opciones:

- Intermitente: significa si se recibe agua por períodos ya sean estos horas, días o semanas.
- Continuo: significa si el servicio lo reciben las 24 horas de todos los días.

Mantenimiento: indicar si se le da o no mantenimiento al servicio de agua.

Existe fontanero a cargo del servicio: marcar esta opción si la comunidad cuenta con alguna persona encargada de la operación y mantenimiento del sistema de agua y si esta persona tiene conocimientos de fontanería.

Responsable del sistema: identificar si el responsable del funcionamiento del sistema es la comunidad, la municipalidad u otro.

Se clora el agua: indicar si la comunidad clora el agua del sistema de abastecimiento:

- La de todo el sistema: significa que se clora el agua en todas las partes del sistema.

- La de una parte del sistema: significa que se aplica el cloro sólo en algunas partes del sistema.
- No se clora el agua: significa que no se aplica cloro al sistema de agua.

Pagan alguna tarifa por el consumo de agua: marcar esta opción si en la comunidad se realiza algún pago por el consumo de agua. En caso afirmativo indicar la cantidad en quetzales que pagan al mes por el servicio de agua.

Saneamiento existente

La comunidad cuenta con saneamiento para la disposición de excretas: marcar esta opción si las viviendas poseen letrinas, o servicio de drenajes, otro etc. No marcar esta opción si las viviendas realizan su disposición de excretas al aire libre.

Si la comunidad cuenta con saneamiento básico: indicar el número total de letrinas según el “tipo de letrina” con que se cuente. Además señalar el estado de cada tipo de letrina con que se cuente:

- Bueno: significa que las letrinas se encuentran en buen estado, poseen caseta, bote de basura e higiene adecuada.
- Malo: si no posee ninguna de las características mencionadas anteriormente.

Muestreo de la calidad del agua en la comunidad

Fecha de muestreo: indicar día, mes y año en que se tomaron las muestras en campo.

Fecha de análisis: indicar día, mes y año en que se analizaron las muestras tomadas en campo. Si utilizó un equipo portátil de análisis, la fecha de muestreo y análisis es la misma.

Realizado por: indicar el nombre y apellido de la persona que realizó el muestreo.

No. de muestras: indicar el número correlativo de la muestra.

Volumen: la cantidad de agua a analizar (100ml.).

Conteo: es el resultado del análisis del agua (colonias).

Cantidad: $\text{conteo} * 100 / \text{volumen}$

Ubicación de la fuente de agua

Coordenadas geográficas: escribir las coordenadas geográficas (latitud y longitud) de la fuente.

Modelo de GPS: identificar el modelo del GPS con el que está trabajando

Datum: asegurarse que utiliza el *datum* WGS84 en su GPS.

Altura: escribir la altura en metros sobre el nivel del mar a la que se encuentra la fuente, ya sea con la elevación que proporcione el GPS o con un altímetro.

Nombre de la fuente: indicar el nombre con el que las personas de la comunidad identifican la fuente de agua.

Datos del tipo de fuente: indicar cuándo se tome agua desde:

- Manantial (nacimiento u ojo de agua)
- Río
- Lago
- Laguna
- Embalse (lugar, natural o artificial, usado para el almacenamiento, regulación y control de los recursos hídricos)
- Pozo (pozo excavado mecánicamente con uso de máquinas o sin usarlas que luego se entuba para extraer agua desde cualquier profundidad)

Método de aforo de la fuente

En esta parte de la boleta se realiza el aforo y dependiendo del tipo de fuente y condición se pueden utilizar diversos métodos, ya sean estos:

- Flotador
- Molinete
- Volumétrico
- Otro

Usos de la fuente

Deberá indicarse si en la actualidad la fuente de agua está en uso.

En esta parte de la boleta anotará si el uso de la fuente es:

- Doméstico
- Riego
- Hidroeléctrico
- Industrial
- Recreativo
- Pecuario
- Otro

Si la fuente es un pozo mecánico o excavado

Si el pozo es mecánico, debe averiguarse el caudal de extracción del mismo (en metros cúbicos por segundo, litros por segundo o galones por minuto). Si el pozo es excavado o artesanal, debe averiguarse la profundidad del agua (en metros) durante la época lluviosa y durante la época seca.

Otra información de la fuente

Deberá indicarse si en la actualidad la fuente de agua tiene cobertura vegetal a su alrededor o no la tiene. Observar y establecer las condiciones de la fuente:

- Mala: las obras de captación/toma están seriamente dañadas, el agua está contaminada o las condiciones generales de la fuente no son favorables para su aprovechamiento.
- Regular: la fuente presenta signos de contaminación leve, la obra de captación/toma no está en perfectas condiciones pero puede ser reparada sin excesivo esfuerzo, y el agua puede ser aprovechada fácilmente.
- Buena: la obra de captación/toma se encuentra en buen estado, no presenta signos de contaminación y presenta condiciones favorables para su aprovechamiento.

Preguntar el estado legal de la fuente. Éste puede ser:

- Escritura pública
- Escritura registrada
- Acta municipal
- Convenios verbales

Observaciones generales

Anotar aquí cualquier observación general que realizó durante el trabajo de campo, que considere relevante documentar.

Visto bueno de la organización comunitaria

Para contar con el visto bueno de los pobladores de la comunidad/centro poblado, es necesario conocer si existe o no una organización comunitaria (comité, asociación, etc.) en el lugar de colecta de la información. En caso exista dicha organización comunitaria, es necesario

conocer el nombre de la misma, y si ésta se dedica o si se podría dedicar a actividades relacionadas con agua y saneamiento.

Miembros de la organización comunitaria: indicar el nombre o nombres de los integrantes de la organización comunitaria.

Información contacto: es importante conocer un medio para contactarlo (teléfono, dirección, correo electrónico) frente a cualquier duda o eventualidad.

2.4.3. Exploración preliminar de los datos

En una exploración preliminar a las boletas, y al conjunto inicial de datos se puede observar una cantidad de métricas bastante interesantes, como por ejemplo, población beneficiada por un sistema de agua, cuántas comunidades con sistema de distribución de agua, no tienen letrina, promedio de cuántas personas hay por sistema de distribución de agua en una comunidad, caudal promedio de una fuente de agua, caudal promedio de las fuentes de agua de un departamento, así como jugar de maneras muy interesantes, a través de las jerarquías de las dimensiones, como por ejemplo, ver información segmentada por país, región, departamento, municipio, comunidad, y ver como varían las métricas en dichos cambios, también priorizar que departamentos necesitan agua, a quienes les sobra, y poder planificar y tomar decisiones con base en esta información.

2.4.4. Verificar calidad de la data

Luego de revisar el conjunto inicial de datos, se puede observar que hay muchos vacíos para poder llegar a cubrir todas las situaciones que se esperan evaluar en el modelo a construir, si se tuvieran conjuntos de datos de las 3 boletas que están dentro del ámbito de este proyecto.

La mayor cantidad de información que se encuentra disponible es proveniente de las boletas SAS de cobertura, pero se encuentra de manera resumida, es aconsejable solicitar al INFOM que proporcione los resultados directos de las boletas, sin haber sido procesadas.

En las boletas PARPA se encuentra la misma situación, información parcial, que no cubre todos los aspectos que se deberían de la boleta, pero si se obtiene otro conjunto de datos diferentes a los de la boleta SAS, con lo cual se pueden lograr nuevos datos y evaluación de nuevas métricas esto permitiría la evaluación de otro segmento del modelo.

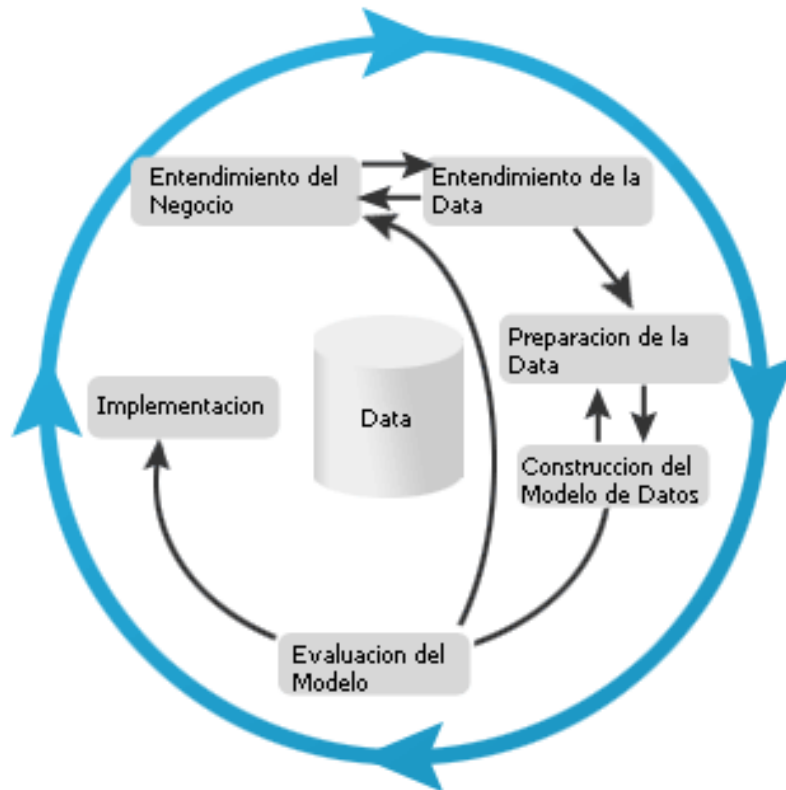
3. SOLUCIÓN PROPUESTA

3.1. Metodología utilizada para resolver la problemática

La metodología utilizada para plantear la solución de este problema, es una adaptación de la metodología CRISP-DM, que es una metodología utilizada para definir procesos de minería de datos, adaptándola al ámbito de este proyecto. Cabe aclarar lo siguiente:

La secuencia de las fases no es estricta, moverse adelante y atrás entre las diferentes fases es muy frecuente. Depende en su gran mayoría de lo que se obtenga de cada fase cual es la siguiente fase, o cual es la siguiente tarea. Las flechas indican el orden más frecuente e importante de las fases, el círculo exterior simbólicamente la naturaleza cíclica que se puede tomar para un proceso de modelado de datos. Una vez que el proyecto es implementado, puede no estar completado, puede requerir de varias iteraciones para llegar a un modelo de datos maduro y estable. Algunos resultados obtenidos durante el proceso y de la solución implementada, pueden generar nuevos requerimientos, comúnmente más enfocados en preguntas de la organización. Las iteraciones posteriores se verán beneficiadas por la experiencia obtenida en las primeras iteraciones.

Figura 1. Diagrama de fases de la metodología CRISP



Fuente: Cross-Industry Standard Process for Data Mining. <http://www.crisp-dm.org>, marzo 2009.

3.1.1. Entendimiento del negocio/organización

Esta es la fase inicial, se concentra principalmente en entender los objetivos del proyecto, desde la perspectiva del cliente, luego convirtiendo este conocimiento en la definición de un problema de almacenes de datos, y un plan preliminar diseñado para cumplir dichos objetivos.

3.1.2. Entendimiento de la data

Esta fase inicia recolectando una colección inicial de datos, para luego seguir con actividades para familiarizarse con los datos, para identificar problemas de calidad, para hacer los primeros hallazgos en los datos, y para detectar segmentos de mayor interés en los datos para formar hipótesis para las posibles métricas de la información.

3.1.3. Preparación de los datos

Esta fase cubre todas las actividades para construir el conjunto final de datos, dichos datos serán alimentados al modelo de datos en su debido momento. Las tareas de esta fase, comúnmente son ejecutadas repetidas veces, sin un orden en específico, estas tareas incluyen la selección de tablas, registros y atributos, así como la transformación y limpieza de los datos.

3.1.4. Construcción del modelo

En esta fase se seleccionan las posibles técnicas para el modelo, y son aplicadas. Usualmente hay varias técnicas para poder solucionar un problema. Algunas de estas técnicas tienen requerimientos específicos en cuanto al formato de la data, por consiguiente retroceder a la fase anterior es comúnmente necesario.

3.1.5. Evaluación

En esta etapa del proyecto, ya se debe haber construido uno o varios modelos, que parecen tener una gran aceptación para poder almacenar los datos correctamente. Antes de proceder a la implementación final del modelo,

es importante de evaluar exhaustivamente el modelo, y revisar los pasos ejecutados para construir el modelo para estar seguros que satisface los objetivos propuestos.

Un objetivo clave es determinar si existe alguna parte importante de los requerimientos que no ha sido considerada lo suficientemente. Al final de esta fase se debe decidir si se va a utilizar este modelo, o se debe retroceder en las fases anteriores para satisfacer los objetivos.

3.1.6. Implementación

La creación del proyecto generalmente no es el final del mismo. Incluso si el propósito del modelo es el de incrementar el conocimiento en los datos, el conocimiento obtenido necesita ser organizado y representado en una manera en que el cliente pueda utilizarlo.

Normalmente involucra varios modelos a los procesos de toma de decisiones de una organización. Sin embargo, dependiendo de los requerimientos, la fase de implementación puede ser tan simple como generar reportes, o tan compleja como implementar un proceso repetible de carga de datos históricos a través de toda la organización. En muchos casos es el cliente, no el analista de datos el que se encarga de los procesos de implementación. Es importante que el cliente entienda claramente que acciones se necesitan para poder hacer uso del o los modelos creados.

3.2. Preparación de datos

3.2.1. Selección de datos

Con base en el análisis realizado en el capítulo anterior, se han seleccionado los datos que se muestran en la siguiente tabla para la construcción del modelo de datos, identificando aquellos datos que se encuentran en común entre las 3 boletas, así como también datos que pueden existir solo para una boleta pero son relevantes para los objetivos de este proyecto, campos que son importantes para la toma de decisiones del inventario nacional de fuentes de agua.

Para seleccionar los datos, también fue necesaria una depuración de los mismos, para obtener un nivel de calidad aceptable de la información obtenida. Para eso fue necesario tomar un subconjunto de la data, así como la inserción de valores por defecto, también se analizó la necesidad de crear estimaciones por medio de modelado para datos que estuvieran ausentes.

En la tabla que se muestra continuación se muestra la correspondencia entre los atributos de las boletas a los campos en las tablas del modelo. El prefijo DIM_ se refiere a las tablas que representan una dimensión, y las el prefijo FACT_ se refiere a las tablas que representa una tabla de hechos (*Fact Table*). En un almacén de datos o un sistema OLAP, la construcción de Cubos OLAP requiere de una tabla de hechos y varias tablas de dimensiones, éstas acompañan a la tabla de hechos y determinan los parámetros (dimensiones) de los que dependen los hechos registrados en la tabla de hechos.

Tabla III. **Dimensión Dim_Boleta**

Nombre Del Atributo	Tipo	Boleta de Origen
IDFuente	int	AUTOGENERADO
NumeroBoleta	int	PARPA/PROVIAGUA/ SAS
NombreEncuestador	varchar(255)	PARPA/PROVIAGUA/ SAS
InstitucionResponsable	datetime	PROVIAGUA/SAS
FechaInicio	datetime	PARPA/PROVIAGUA/ SAS
HoraFinal	datetime	PARPA/PROVIAGUA/ SAS
Unidad	varchar(255)	PROVIAGUA/SAS
CodigoDeFotografia	varchar(50)	PARPA
FechaLlegadaalLaboratorio	datetime	PROVIAGUA
Cargo	varchar(10)	PROVIAGUA
ServicioDeSalud	varchar(255)	PROVIAGUA

Fuente: elaboración propia.

Tabla IV. Dimensión Dim_CalidadDelAgua

Nombre Del Atributo	Tipo	Boleta de Origen
IDFuente	int	AUTOGENERADO
Nombre	varchar(255)	PROVIAGUA
ColiformesFecales	real	PROVIAGUA
CloroResidual	real	PROVIAGUA
CamionCisternaCloroResidual	real	PROVIAGUA
CamionCisternaColiformesFecales	real	PROVIAGUA
FabricaDeHieloCloroResidual	real	PROVIAGUA
FabricaDeHieloColiformesFecales	real	PROVIAGUA
EnvasadoraCloroResidual	real	PROVIAGUA
EnvasadoraColiformesFecales	real	PROVIAGUA

Fuente: elaboración propia.

Tabla V. Dimensión Dim_CalidadDelAgua

Nombre Del Atributo	Tipo	Boleta de Origen
IDFuente	int	AUTOGENERADO
Departamento	varchar(255)	PARPA/PROVIAGUA/SAS
Municipio	varchar(255)	PARPA/PROVIAGUA/SAS
Comunidad	varchar(255)	PARPA/PROVIAGUA/SAS
TipoDeComunidad	varchar(255)	PARPA/PROVIAGUA/SAS
AccesoALaComunidad	varchar(255)	PARPA/PROVIAGUA/SAS
IdiomaPredominante	varchar(255)	PARPA/PROVIAGUA/SAS
ExisteEnergiaElectrica	bit	PARPA/PROVIAGUA/SAS
AreaDeSalud	varchar(255)	PARPA/PROVIAGUA/SAS
PoblacionAproximada	int	PARPA/PROVIAGUA/SAS
NumeroAproximadoDeFamilias	Int	PARPA/PROVIAGUA/SAS
Codigolne	Int	PARPA/PROVIAGUA/SAS

Fuente: elaboración propia.

Tabla VI. **Dimensión Dim_MiembrosOrganizacionComunitaria**

Nombre Del Atributo	Tipo	Boleta de Origen
ID_OrganizacionComunitaria	int	AUTOGENERADO
Nombre	varchar(255)	PARPA/SAS
Sexo	char(2)	PARPA/SAS
Cargo	varchar(255)	PARPA/SAS
Teléfono	varchar(255)	PARPA/SAS

Fuente: elaboración propia.

Tabla VII. **Dimensión Dim_OrganizacionComunitaria**

Nombre Del Atributo	Tipo	Boleta de Origen
ID	int	PARPA/SAS
ExisteOrganizacionComunitaria	bit	PARPA/SAS
Nombre	varchar(255)	PARPA/SAS
OrganizacionDedicadaAguaY Saneamiento	bit	PARPA/SAS
PodriaDedicarseAguaY Saneamiento	bit	PARPA/SAS
IDComunidad	int	PARPA/SAS

Fuente: elaboración propia.

Tabla VIII. **Dimensión Dim_SistemaAbastecimientoAgua**

Nombre Del Atributo	Tipo	Boleta de Origen
IDFuente	int	AUTOGENERADO
DesinfectanElAgua	bit	PROVIAGUA
MetodoUtilizado	varchar(255)	PROVIAGUA
ConexionesIntradomiciliars	bit	PROVIAGUA
ConexionesPrediales	bit	PROVIAGUA
ConexionesLlenaCantaros	bit	PROVIAGUA
QueEntidadFueLaResponsableDeLaConstruccion	varchar(255)	PROVIAGUA
AñoDeConstruccion	int	PROVIAGUA
Administrador	varchar(255)	PROVIAGUA
FuncionaElSistema	bit	PROVIAGUA

Fuente: elaboración propia.

Tabla IX. **Dimensión Dim_UbicacionDeLaFuente**

Nombre Del Atributo	Tipo	Boleta de Origen
IDFuente	int	AUTOGENERADO
CoordenadaX	real	PARPA/PROVIAGUA/SAS
CoordenadaY	real	PARPA/PROVIAGUA/SAS
ModeloDeGPS	varchar(255)	PARPA/PROVIAGUA/SAS
Altitud	int	PARPA/PROVIAGUA/SAS
NombreDeLaFuente	varchar(255)	PARPA/PROVIAGUA/SAS

Fuente: elaboración propia.

Tabla X. **Tabla de hechos Fact_CalidadDelAgua**

Nombre Del Atributo	Tipo	Boleta de Origen
IDFuente	int	AUTOGENERADO
Nombre	varchar(255)	PROVIAGUA
ColiformesFecales	real	PROVIAGUA
CloroResidual	real	PROVIAGUA
CamionCisternaCloroResidual	real	PROVIAGUA
CamionCisternaColiformesFecales	real	PROVIAGUA
FabricaDeHieloCloroResidual	real	PROVIAGUA
FabricaDeHieloColiformesFecales	real	PROVIAGUA
EnvasadoraCloroResidual	real	PROVIAGUA
EnvasadoraColiformesFecales	real	PROVIAGUA

Fuente: elaboración propia.

Tabla XI. **Tabla de hechos Fact_FuentePozoMecanico**

Nombre Del Atributo	Tipo	Boleta de Origen
IDFuente	int	
VolumenDeExtraccion	real	PARPA
PotenciaDeLaBomba	int	PARPA
HorasDeUso	int	PARPA
LaBombaEs	varchar(255)	PARPA

Fuente: elaboración propia.

Tabla XII. **Tabla de hechos Fact_FuentesDeAgua**

Nombre Del Atributo	Tipo	Boleta de Origen
IDFuente	int	AUTOGENERADO
TipoDeLaFuente	varchar(255)	PARPA/SAS
MetodoDeAforo	varchar(255)	PARPA/SAS
CaudalPromedio	real	PARPA/SAS
TipoDeUsuario	varchar(255)	PARPA/SAS
EstructuraDeProteccionEnLaFuente	varchar(255)	PARPA/SAS
TieneCoberturaVegetal	bit	PARPA/SAS
ActualmenteLaFuenteEstaEnUso	bit	PARPA/SAS
EstadoLegalDeLaFuente	varchar(255)	PARPA/SAS
CondicionesDeLaFuente	varchar(255)	PARPA/SAS
IDComunidad	int	PARPA/SAS
IDBoleta	int	AUTOGENERADO

Fuente: elaboración propia.

Tabla XIII. **Tabla de hechos Fact_ResultadosDeLaboratorio**

Nombre Del Atributo	Tipo	Boleta de Origen
IDFuente	int	AUTOGENERADO
MuestraContaminada	bit	PROVIAGUA
LugarDeMuestreo	varchar(255)	PROVIAGUA
MetodoDeCloracion	varchar(255)	PROVIAGUA
SeCobraPorElServicio	bit	PROVIAGUA
TarifaMensual	money	PROVIAGUA
QuienCobraPorElServicio	Varchar(255)	PROVIAGUA

Fuente: elaboración propia.

Tabla XIV. **Tabla de hechos Fact_SaneamientoYTratamientoDeAgua**

Nombre Del Atributo	Tipo	Boleta de Origen
Hay Saneamiento para Disposición De Excretas	bit	PROVIAGUA
Letrinas De Hoyo Buen Estado	int	PROVIAGUA
Letrinas De Hoyo Mal Estado	int	PROVIAGUA
Letrina Hoyo Seco Ventilada Buen Estado	int	PROVIAGUA
Letrina Hoyo Seco Ventilada Mal Estado	int	PROVIAGUA
Letrina Abonera Seca Familiar Buen Estado	int	PROVIAGUA
Letrina Abonera Seca Familiar Mal Estado	int	PROVIAGUA
Letrina Sello Hidráulico Lavable Buen Estado	int	PROVIAGUA
Letrina Sello Hidráulico Lavable Mal Estado	int	PROVIAGUA
Con Sistema	int	PROVIAGUA
Con Tren	int	PROVIAGUA
Con Letrinas	int	PROVIAGUA
ID Comunidad	int	PROVIAGUA
ID Boleta	int	PROVIAGUA

Fuente: elaboración propia.

Tabla XV. **Tabla de hechos**
Fact_SistemasFormalesDeAbastecimiento

Nombre Del Atributo	Tipo	Boleta de Origen
IDComunidad	int	PROVIAGUA/SAS
Nombre	varchar(255)	PROVIAGUA/SAS
Tipo	varchar(255)	PROVIAGUA/SAS
IDComunidad	int	PROVIAGUA/SAS
Nombre	varchar(255)	PROVIAGUA/SAS
UbicacionX	real	PROVIAGUA/SAS
UbicacionY	real	PROVIAGUA/SAS

Fuente: elaboración propia.

Tabla XVI. **Tabla de hechos Fact_UsoDomestico**

Nombre Del Atributo	Tipo	Boleta de Origen
IDFuente	int	PARPA
NombreDeLaPoblacionBeneficiada	varchar(255)	PARPA
NumeroDePoblacionBeneficiada	int	PARPA
TipoDeServicio	nchar(10)	PARPA
SeDescargaElAguaUtilizada	bit	PARPA
CoordenadaXDondeSeDescarga	real	PARPA
CoordenadaYDondeSeDescarga	nchar(10)	PARPA
TienePlantaDeTratamiento	bit	PARPA

Fuente: elaboración propia.

Tabla XVII. **Tabla de hechos Fact_UsoHidroElectrica**

Nombre Del Atributo	Tipo	Boleta de Origen
IDFuente	int	PARPA
NombreDelUsuario	varchar(255)	PARPA
PotenciaDeGeneracion	int	PARPA
CaudalDeGeneracion	real	PARPA
ExisteCurvaDeDescarga	bit	PARPA
TipoDeOperacion	varchar(255)	PARPA
UsaEmbalse	bit	PARPA
VolumenDelEmbalse	real	PARPA
AlturaDeLaPresa	real	PARPA
LongitudDelEmbalse	real	PARPA

Fuente: elaboración propia.

Tabla XVIII. **Tabla de hechos Fact_UsoIndustrialRecreativoPecuario**

Nombre Del Atributo	Tipo	Boleta de Origen
IDFuente	int	PARPA
NombreDelUsuario	varchar(255)	PARPA
VolumenDeAguaUtilizado	real	PARPA
PeriodoDeTiempoQueEsUtilizada	real	PARPA
EpocaDelAño	int	PARPA
SeDescargaElAguaUtilizada	bit	PARPA
CoordenadaXDondeSeDescarga	real	PARPA
CoordenadaYDondeSeDescarga	real	PARPA
TienePlantaDeTratamiento	bit	PARPA
TipoDeGanado	varchar(255)	PARPA
NumeroDeAnimales	varchar(255)	PARPA

Fuente: elaboración propia.

Tabla XIX. **Tabla de hechos Fact_Usolrriacion**

Nombre Del Atributo	Tipo	Boleta de Origen
IDFuente	int	PARPA
NombreUsuario	varchar(255)	PARPA
AreaTotalDeRiego	real	PARPA
TipoDeRiego	varchar(255)	PARPA
NombreCultivo	varchar(255)	PARPA
PeriodoRiegoAnual	int	PARPA
TurnosDeRiego	varchar(255)	PARPA
Duracion	real	PARPA
Frecuencia	int	PARPA
EstructuraHidraulica	varchar(255)	PARPA

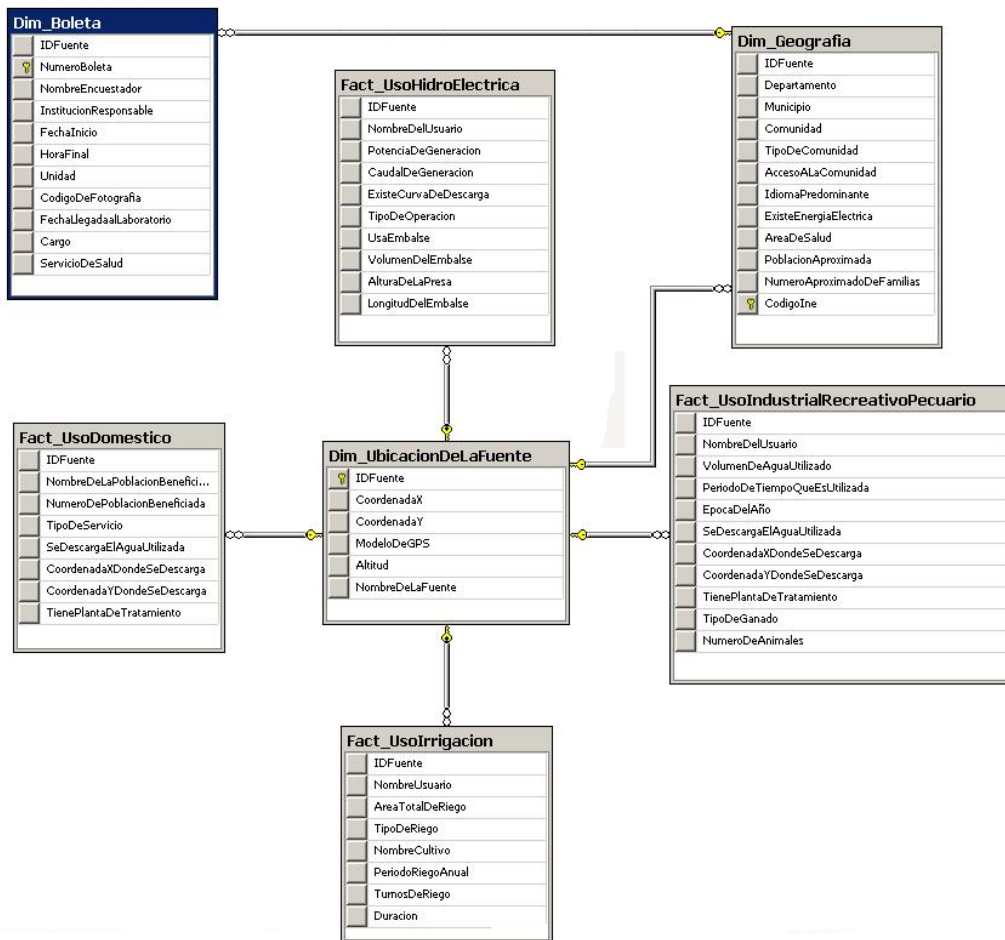
Fuente: elaboración propia.

3.3. Proyecciones de las boletas sobre el modelo

A continuación se muestran las gráficas de cómo cada una de las boletas se ven reflejadas en el modelo de datos, es decir el modelo contiene la información que se analizó de las 3 boletas, por lo que no todos los datos estarán presentes, ya que hay algunas dimensiones que son provenientes de una boleta en específico, lo cual no significa que los datos que vienen en una boleta no estén considerados en el modelo, con lo que se puede decir que cualquier información que venga de cualquiera de las boletas está considerada en el modelo de datos, y podrá ser almacenada para su posterior análisis.

A continuación vemos la gráfica de las dimensiones que contienen la boleta PARPA.

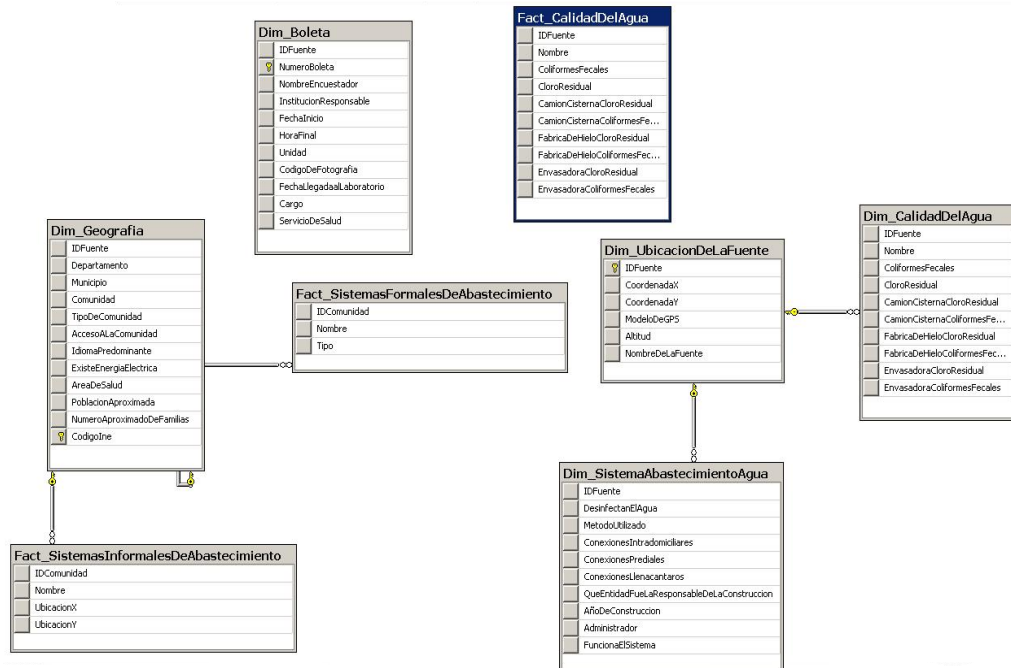
Figura 2. Tablas que corresponden a la boleta PARPA



Fuente: elaboración propia.

A continuación vemos la gráfica de las dimensiones que contienen la boleta PROVIAGUA.

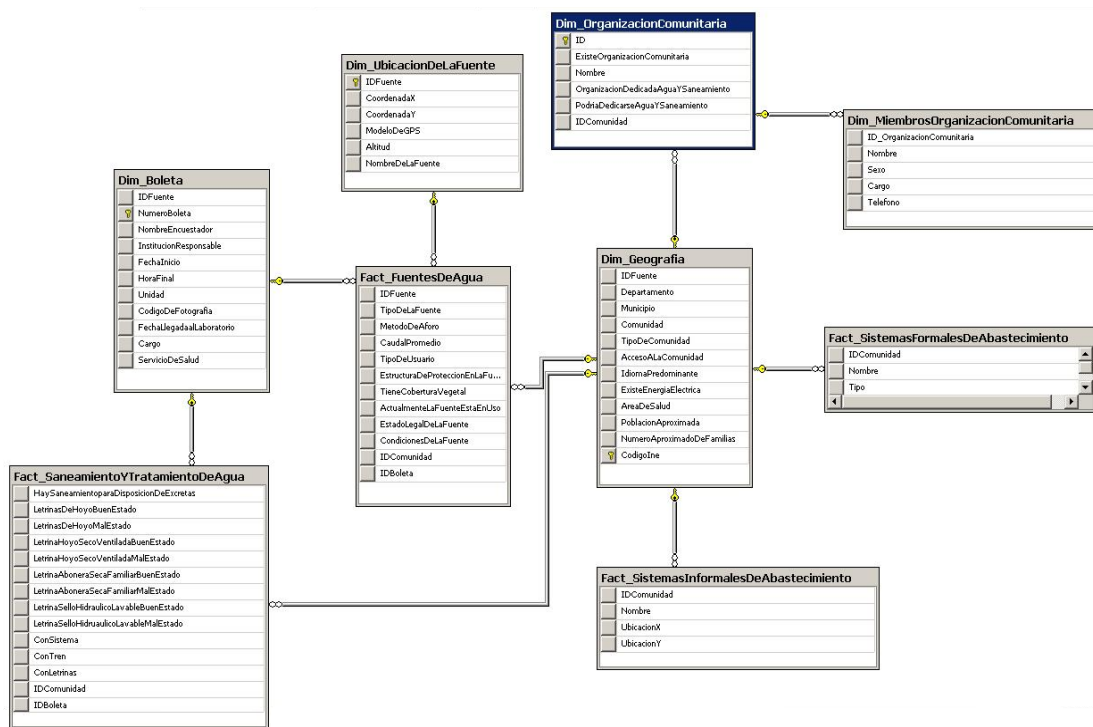
Figura 3. Tablas que corresponden a la boleta PROVIAGUA



Fuente: elaboración propia.

A continuación vemos la gráfica de las dimensiones que contienen la boleta SAS.

Figura 4. Tablas que corresponden a la boleta SAS



Fuente: elaboración propia.

3.4. Construcción del modelo

El modelo lógico de datos que se decidió utilizar para este proyecto está estructurado en un esquema de estrella, que tiene cada dimensión limitada a una sola tabla (tablas planas). A continuación algunas de las razones principales por las cuales se escogió utilizar este esquema contra un conjunto de datos más normalizado:

- Usuarios que realizan consultas directamente a la base de datos relacional tienen mayor facilidad de poder navegar por el modelo;
- Consultas directas a la base de datos relacional tienen mejor desempeño en esta estructura;
- Las tablas planas normalmente son más fáciles de manejar en el proceso de ETL, con menor cantidad de tablas y llaves foráneas.

A continuación, una descripción de cada una de las dimensiones que componen el modelo:

3.5. Descripción de cada una de las dimensiones y tablas de hechos

3.5.1. Dim_UbicacionDeLaFuente

En esta dimensión se almacenan todos los datos de la geo-referencia [27] de la fuente de agua.

Figura 5. Dimensión Dim_UbicacionDeLaFuente



Dim_UbicacionDeLaFuente	
IDFuente	
CoordenadaX	
CoordenadaY	
ModeloDeGPS	
Altitud	
NombreDeLaFuente	

Fuente: elaboración propia.

A continuación una descripción de cada uno de los campos de la tabla:

IDFuente: código único que identifica a la fuente de agua.

CoordenadasX: coordenada X proporcionada por el GPS en la ubicación de la fuente.

CoordenadaY: coordenada Y proporcionada por el GPS en la ubicación de la fuente.

Modelo de GPS: modelo del GPS utilizado durante las mediciones.

Altitud: altitud de la fuente proporcionada por el GPS o un altímetro en la ubicación de la fuente, indicado en metros sobre el nivel del mar (msnm).

Nombre de la fuente: indicar si la fuente de agua es conocida por algún nombre en específico.

3.5.2. Dim_Geografia

En esta dimensión se almacenan todos los datos de la geografía relacionados con una fuente de agua, esta dimensión es una de las más importantes a la hora de hacer los análisis, es la que nos permite crear las jerarquías para dichos análisis pudiendo seleccionar por región la información a analizar, y cambiar de manera dinámica el ámbito en el que se desea presentar la información.

Figura 6. **Dimensión Dim_Geografía**



Dim_Geografía	
IDFuente	
Departamento	
Municipio	
Comunidad	
TipoDeComunidad	
AccesoALaComunidad	
IdiomaPredominante	
ExisteEnergiaElectrica	
AreaDeSalud	
PoblacionAproximada	
NumeroAproximadoDeF...	
CodigoTne	

Fuente: elaboración propia.

A continuación una descripción de cada uno de los campos de la tabla:

IDFuente: código único que identifica a la fuente de agua.

Departamento: nombre del departamento donde está ubicada la fuente de agua.

Municipio: nombre del municipio donde está ubicada la fuente de agua.

Comunidad: nombre de la comunidad o finca donde está ubicada la fuente.

TipoDeComunidad: almacena el tipo de comunidad de donde proviene la información. Ésta puede ser:

- Rural: esta opción debe marcarse si la comunidad se encuentra fuera del área del casco urbano. Si es rural deberá además señalar en la casilla correspondiente si se trata de una aldea, caserío, parajes, fincas, otros;
- Urbana: esta opción debe marcarse si el sistema de agua se encuentra en el casco urbano;
- Marginal: es cuando la población está en el límite de la zona urbana pero con característica urbano/marginal.

AccesoALaComunidad: almacena el tipo de camino que conduce a la comunidad. Éste puede ser:

- Pavimentado (asfalto, cemento, pavimento)
- Terracería (camino de tierra o balastrado)
- Vereda (en donde no puede transitar vehículo)

IdiomaPredominanteDeLaComunidad: almacena el idioma que más habla la gente de la comunidad en donde se está obteniendo la información.

EnergiaElectricaDeLaComunidad: almacena si existe o no energía eléctrica en la comunidad en donde se está obteniendo la información.

AreaDeSalud: almacena el área de salud a la que compete la inspección.

PoblacionAproximada: almacena el número aproximado de personas en la comunidad.

NumeroAproximadoDeFamilias: número aproximado de familias en la comunidad.

CodigoINE: código asignado por el Instituto Nacional de Estadística para identificar a la comunidad.

3.5.3. Dim_Boleta

En esta dimensión, se almacenan los datos relacionados con la boleta de la cual proviene la información, para poder identificar de manera apropiada la colaboración que se tiene de las instituciones, y de esta forma poder dar crédito por la misma.

Figura 7. Dimensión Dim_Boleta

Dim_Boleta	
IDFuente	
NumeroBoleta	
NombreEncuestador	
InstitucionResponsable	
FechaInicio	
HoraFinal	
Unidad	
CodigoDeFotografia	
FechaLlegadaalLaboratorio	
Cargo	
ServicioDeSalud	

Fuente: elaboración propia.

A continuación una descripción de cada uno de los campos de la tabla:

NumeroBoleta: número secuencial con el que se identifica la boleta.

NombreEncuestador: llenar con el nombre completo de la persona que está recopilando la información.

InstitucionResponsable: nombre de la institución a la que pertenece la persona que está recopilando la información.

FechaInicio: fecha y hora en la que se lleno la boleta en formato DD/MM/AA HH:MM.

HoraFinal: indica la hora en que se termino de recolectar la información.

Unidad: departamento o unidad a la que pertenece la persona dentro de la institución.

CódigoDeFotografia: si existe una fotografía, indicar el código con el cual se identifica.

FechaLlegadaAlLaboratorio: indica la fecha y hora en que la muestra llevo al laboratorio.

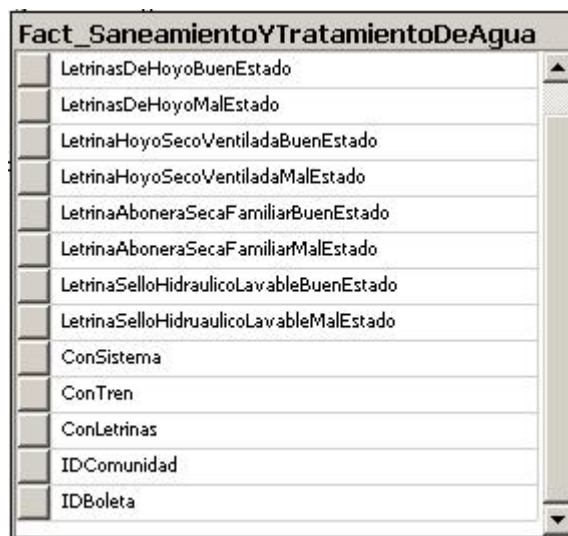
Cargo: cargo que ocupa el encuestador dentro de la institución responsable de la información.

ServicioDeSalud: servicio de salud al que pertenece la persona que realizó la inspección.

3.5.4. Fact_SaneamientoYTratamientoDeAgua

En esta tabla de hechos se almacenan todos los resultados relacionados al conteo de lugares de tratamiento de aguas, así como instalaciones de saneamiento ambiental.

Figura 8. Tabla de hechos Fact_SaneamientoYTratamientoDeAgua



Fact_SaneamientoYTratamientoDeAgua	
LetrinasDeHoyoBuenEstado	
LetrinasDeHoyoMalEstado	
LetrinaHoyoSecoVentiladaBuenEstado	
LetrinaHoyoSecoVentiladaMalEstado	
LetrinaAboneraSecaFamiliarBuenEstado	
LetrinaAboneraSecaFamiliarMalEstado	
LetrinaSelloHidraulicoLavableBuenEstado	
LetrinaSelloHidraulicoLavableMalEstado	
ConSistema	
ConTren	
ConLetrinas	
IDComunidad	
IDBoleta	

Fuente: elaboración propia.

A continuación una descripción de cada uno de los campos de la tabla:

HaySaneamientoparaDisposicionDeExcretas: almacena si las viviendas poseen o no letrinas, o servicio de drenajes, otros etc.

LetrinasDeHoyoBuenEstado: almacena el número total de letrinas de hoyo que se cuente en buen estado, poseen caseta, bote de basura e higiene adecuada.

LetrinasDeHoyoMalEstado: almacena el número total de letrinas de hoyo que se cuente en mal estado, si no posee ninguna de las características mencionadas anteriormente.

LetrinaHoyoSecoVentiladaBuenEstado: almacena el número total de letrinas de hoyo seco y ventiladas que se cuente en buen estado, poseen caseta, bote de basura e higiene adecuada.

LetrinaHoyoSecoVentiladaMalEstado: almacena el número total de letrinas de hoyo seco y ventiladas que se cuente en mal estado, si no posee ninguna de las características mencionadas anteriormente.

LetrinaAboneraSecaFamiliarBuenEstado: almacena el número total de letrinas que se cuente en buen estado, poseen caseta, bote de basura e higiene adecuada.

LetrinaAboneraSecaFamiliarMalEstado: almacena el número total de letrinas que se cuente en mal estado, si no posee ninguna de las características mencionadas anteriormente.

LetrinaSelloHidraulicoLavableBuenEstado: almacena el número total de letrinas que se cuente en buen estado, poseen caseta, bote de basura e higiene adecuada.

LetrinaSelloHidruaulicoLavableMalEstado: almacena el número total de letrinas que se cuente en mal estado, si no posee ninguna de las características mencionadas anteriormente.

3.5.5. Dim_SistemaAbastecimientoAgua

Esta dimensión almacena la información relacionada con los sistemas de abastecimiento de agua, los cuales pueden ser formales o informales, pero eso se detalla en las dimensiones Dim_SistemasFormalesDeAbastecimiento y Dim_SistemasInformalesDeAbastecimiento.

Figura 9. Dimensión Dim_SistemaAbastecimientoAgua

Dim_SistemaAbastecimientoAgua	
IDFuente	
DesinfectanElAgua	
MetodoUtilizado	
ConexionesIntradomiciliars	
ConexionesPrediales	
ConexionesLlenacantaros	
QueEntidadFueLaResponsableDeLaConstruccion	
AñoDeConstruccion	
Administrador	
FuncionaElSistema	

Fuente: elaboración propia.

A continuación una descripción de cada uno de los campos de la tabla:

DesinfectanElAgua: indica si utilizan o no algún método para desinfectar el agua.

MetodoUtilizado: método para desinfectar el agua utilizado en el sistema de abastecimiento.

ConexionesIntradomiciliares: almacena el número de conexiones intradomiciliares con el sistema de abastecimiento.

ConexionesPrediales: almacena si existen o no conexiones prediales con el sistema de abastecimiento.

ConexionesLlenacantaros: almacena si existen o no conexiones llena-cántaros con el sistema de abastecimiento.

EntidadResponsableDeLaConstruccion: nombre de la entidad responsable de la construcción del sistema de abastecimiento.

AñoDeConstruccion: año en que fue hecha la construcción del sistema de abastecimiento.

Administrador: nombre de la persona o entidad encargada de la administración del sistema.

FuncionaElSistema: almacena si en la actualidad el sistema funciona correctamente.

3.5.6. Fact_ResultadosDeLaboratorio

En esta tabla de hechos se almacenan los resultados obtenidos de ciertos análisis bacteriológicos realizados a las muestras obtenidas.

Figura 10. **Tabla de hechos Fact_ResultadosDeLaboratorio**

Fact_ResultadosDeLabor	
IDFuente	
MuestraContaminada	
LugarDeMuestreo	
MetodoDeCloracion	
SeCobraPorElServicio	
TarifaMensual	
QuienCobraPorElServicio	

Fuente: elaboración propia.

A continuación una descripción de cada uno de los campos de la tabla:

MuestraContaminada: almacena si según los resultados de laboratorio la muestra está contaminada o no.

LugarDeMuestreo: almacena el lugar donde fue realizado el muestreo en la fuente de agua.

MetodoDeCloracion: almacena el método de cloración utilizado si existiese.

SeCobraPorElServicio: almacena si se cobra por el servicio del agua o no.

TarifaMensual: almacena el monto que se cobra por el servicio del agua.

QuienCobraPorElServicio: almacena quien es la persona o institución que recibe el dinero que se cobra por el servicio del agua.

3.5.7. Fact_FuentePozoMecanico

En esta tabla de hechos se almacenan los datos obtenidos si el uso que se le da a una fuente de agua es un pozo mecánico.

Figura 11. **Tabla de hechos Fact_FuentePozoMecanico**



Fact_FuentePozoMecanico	
IDFuente	
VolumenDeExtraccion	
PotenciaDeLaBomba	
HorasDeUso	
LaBombaEs	

Fuente: elaboración propia.

A continuación una descripción de cada uno de los campos de la tabla:

VolumenDeExtraccion: almacena la capacidad que tiene el pozo mecánico para extraer agua.

PotenciaDeLaBomba: almacena la capacidad que tiene el pozo mecánico para extraer agua.

HorasDeUso: almacena el número de horas al día que se encuentra activa la bomba para extraer agua.

LaBombaEs: almacena el tipo de bomba que extrae el agua.

3.5.8. Fact_UsoHidroElectrica

En esta tabla de hechos se almacena los datos obtenidos si el uso que se le da una fuente de agua es una estación hidroeléctrica.

Figura 12. **Tabla de hechos Fact_UsoHidroelectrica**

Fact_UsoHidroElectrica	
IDFuente	
NombreDelUsuario	
PotenciaDeGeneracion	
CaudalDeGeneracion	
ExisteCurvaDeDescarga	
TipoDeOperacion	
UsaEmbalse	
VolumenDelEmbalse	
AlturaDeLaPresa	
LongitudDelEmbalse	

Fuente: elaboración propia.

A continuación una descripción de cada uno de los campos de la tabla:

NombreDelUsuario: almacena el nombre de la persona/institución que usa el agua de la fuente para generar electricidad.

PotenciaDeGeneracion: almacena la potencia la cual se puede generar con el caudal proveniente de esta fuente de agua.

CaudalDeGeneracion: almacena caudal necesario para la potencia de generación

ExisteCurvaDeDescarga: almacena si existe curva de descarga o no.

TipoDeOperación: almacena el tipo de operación de la hidroeléctrica.

UsaEmbalse almacena si se usa embalse o no para la generación de energía eléctrica.

VolumenDelEmbalse: almacena el volumen de agua que puede contener el embalse.

Altura de la presa: almacena la altura del muro frontal de la presa.

Longitud del embalse: almacena la longitud en la parte superior del muro de la presa.

3.5.9. Fact_UsoDomestico

En esta tabla de hechos se almacenan los datos obtenidos si el uso que se le da a una fuente de agua es para consumo doméstico.

Figura 13. **Tabla de hechos Fact_Usodomestico**

Fact_Usodomestico	
IDFuente	
NombreDeLaPoblacionBenefi...	
NumeroDePoblacionBenefici...	
TipoDeServicio	
SeDescargaElAguaUtilizada	
CoordenadaXDondeSeDesca...	
CoordenadaYDondeSeDesca...	
TienePlantaDeTratamiento	

Fuente: elaboración propia.

A continuación una descripción de cada uno de los campos de la tabla:

NombreDeLaPoblacionBeneficiada: almacena el nombre de la población que se beneficia del agua de la fuente.

NumeroDePoblacionBeneficiada: almacena el número de personas beneficiadas en la población.

TipoDeServicio: almacena el tipo de servicio que se presta con el agua obtenida de la fuente.

SeDescargaElAguaUtilizada almacena si se descarga el agua utilizada o no.

CoordenadaXDondeSeDescarga: almacena la coordenada X que indica el GPS en el punto donde se descarga el agua.

CoordenadaYDondeSeDescarga: almacena la coordenada Y que indica el GPS en el punto donde se descarga el agua.

TienePlantaDeTratamiento: almacena si al descargar el agua pasa por una planta de tratamiento al ser descargada.

3.5.10. Fact_UsIndustrialRecreativoPecuario

En esta tabla de hechos se almacenan los datos obtenidos si el uso que se le da a una fuente de agua es industrial, recreativo o pecuario.

Figura 14. **Tabla de hechos Fact_UsIndustrialRecreativoPecuario**

Fact_UsIndustrialRecreativoPecuario	
IDFuente	
NombreDelUsuario	
VolumenDeAguaUtilizado	
PeriodoDeTiempoQueEsUtilizada	
EpocaDelAño	
SeDescargaElAguaUtilizada	
CoordenadaXDondeSeDescarga	
CoordenadaYDondeSeDescarga	
TienePlantaDeTratamiento	
TipoDeGanado	
NumeroDeAnimales	

Fuente: elaboración propia.

A continuación una descripción de cada uno de los campos de la tabla:

NombreDelUsuario: almacena el nombre de la persona que usa el agua de la fuente.

VolumenDeAguaUtilizado: almacena el volumen de agua que se usa.

PeriodoDeTiempoQueEsUtilizada: almacena el número de horas al día en las que se hace uso del agua.

EpocaDelAño: almacena los meses del año que el agua es utilizada.

SeDescargaElAguaUtilizada: almacena si el agua utilizada es descargada o no.

CoordenadaXDondeSeDescarga: almacena la coordenada X que indica el GPS en el punto donde se descarga el agua.

CoordenadaYDondeSeDescarga: almacena la coordenada Y que indica el GPS en el punto donde se descarga el agua

TienePlantaDeTratamiento: almacena si al descargar el agua pasa por una planta de tratamiento al ser descargada.

TipoDeGanado: almacena el tipo de ganado que se beneficia del agua de la fuente.

NumeroDeAnimales: almacena el número aproximado de cabezas de ganado beneficiados por el agua de la fuente.

3.5.11. Dim_OrganizacionComunitaria

En esta dimensión se almacenan los datos de una organización comunitaria, si existiese; y la información si la organización está dedicada a temas de agua y saneamiento, o si está dispuesta a hacerlo.

Figura 15. **Dimensión Dim_OrganizacionComunitaria**

	Column Name	Data Type	Allow Null
⚡	ID	int	<input type="checkbox"/>
	ExisteOrganizacionCo...	bit	<input checked="" type="checkbox"/>
	Nombre	varchar(255)	<input checked="" type="checkbox"/>
	OrganizacionDedicada...	bit	<input checked="" type="checkbox"/>
	PodriaDedicarseAgua...	bit	<input checked="" type="checkbox"/>
	IDComunidad	int	<input checked="" type="checkbox"/>

Fuente: elaboración propia.

A continuación una descripción de cada uno de los campos de la tabla:

ExisteOrganizacionComunitaria: almacena si en la comunidad existe una organización comunitaria o no.

Nombre: almacena el nombre con el que se identifica la organización comunitaria.

OrganizacionDedicadaAguaYSaneamiento: almacena si la organización tiene actividades o trata temas de agua y saneamiento.

PodriaDedicarseAguaYSaneamiento: si en la actualidad la organización no se dedica a temas de agua y saneamiento o si tienen la disponibilidad de hacerlo.

3.5.12. Dim_MiembrosOrganizacionComunitaria

En esta dimensión se almacenan los datos de las personas que conforman la organización comunitaria.

Figura 16. Dimensión Dim_MiembrosOrganizacionComunitaria

Dim_MiembrosOrganizacionComunitaria	
ID_OrganizacionComunitaria	
Nombre	
Sexo	
Cargo	
Telefono	

Fuente: elaboración propia.

A continuación una descripción de cada uno de los campos de la tabla:

ID_OrganizacionComunitaria: identificador único de la organización comunitaria a la que pertenece el miembro.

Nombre: almacena el nombre del miembro de la organización.

Sexo: almacena el sexo del miembro de la organización.

Cargo: almacena el nombre del cargo que tiene la persona dentro de la organización.

Teléfono: almacena el número telefónico donde se puede ubicar a la persona.

3.5.13. Fact_SistemasFormalesDeAbastecimiento

En esta dimensión se almacenan los datos que identifican a un sistema formal de abastecimiento de agua.

Figura 17. **Tabla de hechos Fact_SistemasFormalesDeAbastecimiento**



Fact_SistemasFormalesD	
IDComunidad	
Nombre	
Tipo	

Fuente: elaboración propia.

A continuación una descripción de cada uno de los campos de la tabla:

IDComunidad: identificador único con el que se identifica la comunidad a la que pertenece el sistema formal de abastecimiento.

Nombre: almacena el nombre con que se identifica al sistema de abastecimiento.

Tipo: almacena el tipo del sistema.

3.5.14. Fact_SistemasInformalesDeAbastecimiento

En esta dimensión se almacenan los datos que identifican a un sistema informal de abastecimiento de agua.

Figura 18. **Tabla de hechos**
Fact_SistemasInformalesDeAbastecimiento

Fact_SistemasInformalesDeAbastecimiento	
IDComunidad	
Nombre	
UbicacionX	
UbicacionY	

Fuente: elaboración propia.

A continuación una descripción de cada uno de los campos de la tabla:

IDComunidad: identificador único con el que se identifica la comunidad a la que pertenece el sistema informal de abastecimiento.

Nombre: almacena el nombre con que se identifica al sistema de abastecimiento.

UbicacionX: almacena la coordenada X proporcionada por el GPS en la ubicación del sistema.

UbicacionY: almacena la coordenada Y proporcionada por el GPS en la ubicación del sistema.

3.5.15. Fact_FuentesDeAgua

En esta tabla de hechos se almacenan los datos de mediciones realizadas en la fuente de agua.

Figura 19. **Tabla de hechos Fact_FuentesDeAgua**

Fact_FuentesDeAgua	
IDFuente	
TipoDeLaFuente	
MetodoDeAforo	
CaudalPromedio	
TipoDeUsuario	
EstructuraDeProteccionEnLa...	
TieneCoberturaVegetal	
ActualmenteLaFuenteEstaEn...	
EstadoLegalDeLaFuente	
CondicionesDeLaFuente	
IDComunidad	

Fuente: elaboración propia.

A continuación una descripción de cada uno de los campos de la tabla:

TipoDeLaFuente: almacena el tipo de la fuente de agua

MetodoDeAforo: almacena el método de aforo que se utiliza para obtener el agua de la fuente.

IDComunidad: identificador único con el que se identifica la comunidad a la que pertenece la fuente de agua.

IDBoleta: identificador único con el que se identifica la boleta de donde proviene la información de la fuente de agua identificada.

CaudalPromedio: almacena el valor del caudal promedio en las mediciones.

TipoDeUsuario: almacena el tipo de usuario que utilizan en la fuente de agua.

EstructuraDeProteccionEnLaFuente: almacena si la fuente cuenta con algún tipo de estructura de protección.

TieneCoberturaVegetal: almacena si en la superficie de la fuente se encuentra o no cobertura vegetal.

ActualmenteLaFuenteEstaEnUso: almacena si la fuente se encuentra actualmente en uso o no.

EstadoLegalDeLaFuente: almacena el estado legal en que se encuentra la fuente de agua.

CondicionesDeLaFuente: almacena el estado actual de la fuente de agua.

4. EVALUACIÓN DE LA SOLUCIÓN PROPUESTA Y RECOMENDACIONES PARA FUTUROS DESARROLLOS

4.1. Evaluación del modelo

Para realizar la evaluación del modelo, se utilizaron datos obtenidos de las boletas PARPA, y SAS, provenientes de los departamentos de Chiquimula, Huehuetenango, Quiché y San Marcos. En total se obtuvieron datos de noventa y dos municipios. Una vez cargados los datos al modelo, se procedió a desarrollar grupos de consultas que según el ámbito de este proyecto podrían ser útiles, y de esta manera evaluar si el diseño del modelo satisface los objetivos planteados al inicio de este proyecto, de dicha evaluación se obtuvieron resultados positivos, los cuales nos permiten decir si el modelo puede implementarse y utilizarse según las recomendaciones propuestas en este capítulo, puede llegar a dar la información necesaria para dar soporte a las tomas de decisiones del Inventario Nacional de Fuentes de Agua, dando el planteamiento inicial de una solución que se le puede sacar demasiado provecho, si es bien utilizada y si llega a contar con el apoyo y la cooperación de todas las instituciones relacionadas con dicha problemática.

A continuación, un conjunto de recomendaciones para el proyecto, tanto para mejoras al modelo, como para la implementación, y futuras aplicaciones que pueden desarrollarse basados en el almacén de datos, etc.

4.2. Recomendaciones para el modelo

- El primer cambio que se propone es el de estandarizar el uso de coordenadas geográficas, en lugar de las actuales, que son las UTM.
- Para optimizar el modelo se pueden unificar las tablas de hechos Fact_SistemasFormalesDeAbastecimiento con Fact_SistemasInformalesDeAbastecimiento, agregando un campo de tipo booleano para decir si el sistema es formal o no, se asume que si es no, el sistema es informal.
- Estandarizar los códigos que identifican únicamente a las comunidades, de esta manera se puede mejorar drásticamente la calidad en los datos generados, y simplificando de alguna manera el diseño para el proceso de ETL. Se sugiere llegar a un consenso entre todas las instituciones, para utilizar de manera general el código que utiliza el Instituto Nacional de Estadística (INE) para identificar a las comunidades.
- Las mayor cantidad de información que se obtuvo para la colección inicial de datos proviene de las boletas SAS de cobertura, pero se encuentra de manera resumida, y en un formato poco amigable para el proceso de carga, de manera que se podría solicitar al INFOM que proporcione los resultados directos de las boletas, no los resúmenes de las mismas.
- En las boletas PARPA, se encuentra la misma situación, información parcial, que no cubre todos los aspectos que se deberían de la boleta, por lo que se debe considerar obtener una nueva colección de datos, a la hora de diseñar el proceso de ETL.

4.3. Recomendaciones ETL

El proceso de diseño para el sistema de ETL es crítico, recabar toda la información relevante, incluyendo la carga que el proceso de extracción generará en los sistemas transaccionales y probar algunas alternativas disponibles. Se debe responder a las preguntas:

- A. ¿Es factible almacenar el proceso de transformación en el sistema de origen, en el sistema de destino, o en su propia plataforma?
- B. ¿Qué herramientas hay disponibles, y que tan efectivas son?

Es muy probable que se tenga que escribir código desde el inicio, incluyendo consultas SQL y scripts de automatización. Es recomendable usar un manejador de versiones para el código.

4.3.1. Pasos para un proceso de ETL

4.3.1.1. Planificación

Pasos para planificar un proceso de ETL:

- Crear a gran nivel, un esquema del flujo del origen al destino;
- Probar, escoger e implementar una herramienta de ETL;
- Desarrollar estrategias para el manejo de dimensiones, manejo de errores, y otros procesos;
- Examinar los detalles de la tabla destino, haciendo un bosquejo gráfico de la reestructuración de la data, o las transformaciones si son necesarias, y desarrollar un conjunto de pasos preliminares necesarios.

4.3.1.2. Desarrollar una carga única de información histórica

Pasos a seguir para construir el proceso de carga histórica:

- Construir y probar la carga para la tabla de la dimensión histórica;
- Construir y probar la carga para la tabla de la dimensión histórica, incluyendo la sustitución de las llaves primarias por sustitutas, y verificar la coincidencia de las sustituciones realizadas.

4.3.1.3. Desarrollar el proceso de carga incremental

Pasos a seguir para construir el proceso de carga incremental:

- Construir y probar el proceso de carga incremental a las dimensiones;
- Construir y probar el proceso de carga incremental a las tablas de hechos;
- Construir y probar el proceso de carga incremental a las tablas adicionales, y/o del modelo OLAP;
- Diseñar, construir y probar el la automatización del proceso de ETL.

Desarrollar el sistema de ETL es uno de los más grandes esfuerzos en los proyectos de almacenes de datos, en la mayoría de los casos no dependen directamente del resultados de las entrevistas y los análisis realizados en las primeras fases del proyecto. Es necesario considerar los problemas de calidad de la data, ya que éstos pueden ser tan graves que pueden hacer fracasar el proceso de ETL, haciendo necesario rediseñar dicho proceso.

El proceso de extracción es un paso bastante complicado, es necesario trabajar de cerca con los programadores del sistema de origen para poder desarrollar un proceso de extracción que genere datos de alta calidad; que al

mismo tiempo no sature a los sistemas transaccionales, y que pueda ser incorporado dentro del flujo del sistema automatizado de ETL.

4.4. Recomendaciones para desarrollo de futuras aplicaciones informáticas

Se pueden construir los siguientes tipos de soluciones (aunque no se limita únicamente a estas) para poder divulgar la información de un almacén de datos, o para aprovechar al máximo la información del mismo:

4.4.1. Consultas realizadas directamente en la base de datos

Permite a los usuarios consultar el modelo multidimensional directamente y definir un conjunto de resultados. Estas herramientas únicamente permiten retribuir los datos de forma tabular, también existen herramientas más avanzadas que permiten la creación de reportes bastante complejos. En estos casos, dichas herramientas funcionan como el medio para desarrollar los reportes que se puedan definir para que otros usuarios puedan ejecutarlos por sí mismos.

4.4.2. Minería de datos

Normalmente un equipo de especialistas en el área (en el caso de este proyecto hidrólogos, o técnicos de las distintas instituciones) puede utilizar las estadísticas y la información en el almacén de datos, para buscar patrones útiles dentro de la data, así como correlaciones o tendencias que no son obvias a simple vista. Este tipo de tendencias o patrones pueden encontrarse gracias a la asistencia de un *software* de minería de datos y en conjunto con el entendimiento de la data por parte de los especialistas. Los modelos resultantes

de dicho proceso, son comúnmente utilizados por otras aplicaciones de inteligencia de negocios. La minería de datos usualmente nunca es vista por la mayoría de usuarios.

4.4.3. Reportes comunes

Son reportes predefinidos, con un formato constante que generalmente proveen algún tipo de interacción con el usuario, como por ejemplo la habilidad de ingresar algún parámetro, examinar los detalles a un mayor nivel de detalle, poder disponer de filtros para catalogar la información, etc.

4.4.4. Aplicaciones para análisis

Son un conjunto de reportes administrados, que usualmente tienen información inherente a la naturaleza de la organización, que requiere el conocimiento de un experto en la materia para poder ser interpretados. La mayoría de estas aplicaciones requieren una capa de interface para proporcionar acceso a los usuarios.

4.4.5. Cuadros de mandos y tarjetas de responsabilidades

Usualmente implican una combinación de reportes y gráficas que resaltan las excepciones dentro de las tendencias del análisis, y poseen capacidades de poder examinar a mayor detalle los datos desde diferentes enfoques de la organización.

4.4.6. Aplicaciones de inteligencia de negocios

Incluye el uso de aplicaciones que son más sofisticadas que los típicos reportes de operaciones. Estas aplicaciones unifican el contexto histórico contenido en el almacén de datos desde múltiples enfoques de la organización, para guiar y apoyar así al proceso de toma de decisiones. Las aplicaciones de inteligencia de negocios normalmente incluyen modelos de minería de datos, para ayudar a identificar patrones y oportunidades en el nivel operacional. Estas aplicaciones normalmente incluyen interfaces para poder ver la fuente de los datos.

4.4.7. Portal web

Regularmente, para la organización es importante tener centralizado el acceso a todas las aplicaciones que se pueden desarrollar en torno al almacén de datos, para crear una plataforma unificada, generalmente se sugiere la creación de un portal, el cual será la interface principal para acceder a las aplicaciones mencionadas. Dando así un lugar bien organizado, útil, y de fácil entendimiento para encontrar todas las herramientas y la información requerida.

Las características con las que debería contar un portal son las siguientes:

Usabilidad:

La gente tiene que poder encontrar lo que está buscando de una manera simple y ordenada.

Rico en contenido:

El portal debe incluir algo más que solamente reportes. Debería incluir información de soporte, documentación, ayuda, ejemplos y asistencia, de ser posible.

Simple:

El diseño del portal debe ser desplegado de tal forma que la gente no se confunda ni sea saturada con tanta información.

Actualizado:

Es necesario nombrar un encargado para mantener el contenido del portal actualizado, que no existan enlaces erróneos, tampoco información etiquetada como “nueva” cuando ya es de años atrás.

Interactivo:

El portal debe incluir algunas funciones para darle beneficios al usuario, y motivarlos a regresar al mismo. Una buena herramienta de búsqueda, un examinador de metadatos, o un grupo de noticias, son ejemplos con los cuales se tener la interacción entre usuarios y el portal. También puede ser útil la capacidad de personalización, que permita a los usuarios a modificar su página de reportes a su gusto, guardar reportes, o enlaces a los mismos. También ayuda tener nuevas características con cierta frecuencia. Realizar encuestas, boletines de notificaciones, para mantener un ambiente fresco y proveer retroalimentación al usuario.

Orientado a valores:

Ésta debe ser una meta de la organización, el objetivo es que todos los usuarios que visiten el portal, lo vean como algún recurso con muchos beneficios, algo que los ayuda a realizar mejor su trabajo. De cierta

manera, el portal es una de las mayores herramientas de publicidad que el equipo tiene por lo que la impresión de cada usuario cuenta.

El propósito primordial del portal es ofrecer la navegación apropiada a través de todas las aplicaciones desarrolladas, para que las personas puedan encontrar las herramientas cuando las necesiten, de una manera centralizada y estandarizada. Todo el texto que se incluya en el portal debe comunicar de manera eficiente que contiene, es decir lo que la gente encontrará dentro de la funcionalidad descrita.

Generalmente, la mejor manera de organizar el portal, es basarse en la estructura de la organización, como primera propuesta. Más allá de procesos de negocio, el portal necesita tener una estructura estándar, para que la gente pueda encontrar fácilmente lo que busca, y familiarizarse con el uso de la herramienta, con alguna herramienta que ya sea conocida dentro de la organización, o crear un nuevo formato.

Aunque algunas de las funciones principales del portal es proveer acceso a los diferentes reportes, es mejor si el portal ofrece algo más que reportes. En adición a las categorías y las listas de reportes, se pueden proveer las siguientes funciones:

Búsquedas:

Las herramientas de búsqueda pueden servir como un medio para localizar reportes, en caso de que las categorías no estén creadas apropiadamente. Una buena herramienta de búsqueda que cree índices cada documento y pagina en el sitio *web*, y todos los metadatos de los reportes, incluyendo el nombre del reporte, y la descripción del mismo puede reducir drásticamente el tiempo de respuesta de las búsquedas.

Explorador de metadatos:

Puede ser tan simple como un pequeño conjunto de páginas, o reportes que permitan al usuario recorrer las colecciones de metadatos, que contengan descripciones de las bases de datos, de los esquemas, de las tablas, las columnas, reglas del negocio, estadísticas del proceso de carga, uso de los reportes, y el contenido de los reportes. En poco tiempo los usuarios interesados pueden aprender suficiente del sistema del almacén de datos, a través de esta herramienta.

Foro de usuarios:

Ésta puede ser una buena manera en que los usuarios puedan encontrar ayuda cuando la necesiten. También se puede crear un historial de los problemas y de las soluciones para referencias posteriores. Requiere una cantidad considerable de usuarios para poder generar el contenido crítico del foro y poder hacer un grupo de noticias exitoso. También requiere mucho tiempo y esfuerzo moderar el foro de usuarios.

Personalización:

Los usuarios deberían de tener la disponibilidad de guardar reportes, o enlaces hacia los reportes en sus páginas personales, así como el conjunto de parámetros que ellos han escogido. Esta personalización puede ser un gran incentivo para que las personas regresen al portal con mayor frecuencia.

Centro de información:

Para ayudar a mantener el sitio interesante, es necesario que aparezcan nuevas características con cierta regularidad. Realizar encuestas, ofrecer entrenamientos a las personas para utilizar la herramienta, o publicar noticias de próximas reuniones para los usuarios.

4.4.8. Calendarización de los reportes

Usualmente hay un grupo de tareas que se necesitan ejecutar automáticamente en con una frecuencia programada, ya sea por desempeño o por otras circunstancias. Supongamos que existen un grupo de reportes que toma unos cuantos minutos para cargar, y hay cientos de usuarios que necesitan revisarlos, diariamente a la misma hora, es recomendable que los reportes se totalicen tan pronto como el proceso de ETL ha finalizado, y se guarde la información ya calculada. De esta manera cuando los usuarios consulten este reporte, el sistema únicamente seleccionará los datos ya totalizados, en lugar de calcularlo cada vez que el reporte sea solicitado.

La mayoría de las herramientas proveen un servicio de reportes que permiten programar la generación de los reportes con ciertos parámetros predefinidos, o pueden ser ejecutados como el resultado de un evento, como la finalización del proceso de carga de datos. Los resultados de los reportes pueden ser centralizados para su almacenamiento, y ser retribuidos conforme vayan siendo solicitados por los usuarios o ser distribuidos automáticamente a ciertos usuarios predefinidos, por medio de email, algún medio físico, o incluso impresos.

4.5. Recomendaciones de implementación

Para implementar exitosamente el nuevo almacén de datos con todas sus aplicaciones, o si son cambios de una iteración posterior, hay que planear con antelación y probar repetidamente hasta que se ha verificado que:

- Los procedimientos de pruebas del sistema son lo suficientemente rigurosos, que es posible comprobar a los usuarios o a auditores que el sistema ha sido probado exhaustivamente;
- Los datos son precisos y están completos, tanto como para la carga histórica de información, como para los procesos incrementales de carga;
- Las bases de datos se cargan y son procesadas correctamente;
- El desempeño del sistema al momento de la carga, al realizar consultas, y reportes es el deseado;
- Los usuarios del sistema pueden encontrar lo que necesitan y pueden realizar sus tareas requeridas;
- El plan de implementación es robusto, y que ha sido ensayado;
- Las computadoras de los usuarios cumplen con los requisitos mínimos para poder utilizar las nuevas herramientas, y que las herramientas se encuentran instaladas.

4.6. Recomendaciones post-implementación

4.6.1. Mantener el portal actualizado

El portal descrito con anterioridad es el mejor lugar para publicar información sobre el sistema completo. A continuación un conjunto de operaciones y mantenimiento de la información que debería de colocarse en dicho portal:

- Estado del almacén de datos: cuál es la más reciente carga de datos para cada uno de los procesos del negocio;
- Horarios en que el sistema no estará disponible, por mantenimiento o cualquier motivo;

- Información y estado completo en caso de que el sistema no se encuentre en línea por algún incidente;
- Advertencias claras a los usuarios acerca de problemas en el sistema, tales como problemas en la calidad de los datos. Incluyendo un aproximado de cuando los problemas serán solucionados;
- Estado actual del estado operacional del sistema, incluyendo:
 - Cuantas consultas han sido respondidas en la última hora
 - Cuantos reportes se han generado
 - Número de usuarios activos
 - Si es requerido, reporte de consultas activa, incluyendo el tiempo que llevan transcurrido en su ejecución, y el usuario que la está ejecutando.
- Estadísticas de la capacidad del sistema, para crear confianza en los usuarios para usar el sistema:
 - Cuánto tiempo tardo la ultima carga de datos
 - Qué tan grande eran en conjunto los datos cargados, y cuanto tardó
 - Máximo número de usuarios simultáneos.

4.6.2. Copias de respaldo y recuperación

Sin importar los requerimientos de disponibilidad del sistema, es necesario tener un plan de recuperación. E igualmente importante probar estos procedimientos para generar las copias de respaldo así como la recuperación de las mismas. Es recomendable simular una emergencia aunque sea una vez,

para tener la tranquilidad que todo funcionará como se espera al momento en que se necesite. Documentar todo lo necesario para realizar un proceso de recuperación exitoso, y dónde se pueden encontrar todas las herramientas necesarias. Es recomendable tener un ambiente de desarrollo, para poder realizar en el mismo todas las pruebas necesarias.

Un almacén de datos y las aplicaciones desarrolladas pueden sufrir de las mismas emergencias que un sistema transaccional, por lo que no hay que subestimar realizar copias de seguridad de la información. Así mismo, hay que considerar planes de contingencia por si los sistemas de origen de datos, los que alimentan el almacén de datos no se encuentran disponibles, hay que generar notificaciones si esto sucede durante el proceso de ETL haciendo que el proceso falle.

También es necesario considerar un proceso de almacenaje para información que ya es demasiado antigua, y que no se quiere desechar, pero se quiere tener disponible en caso de que sea necesaria. Información que puede ser sacada del periodo de interés actual, para que no impacte en el desempeño del sistema.

Una implementación exitosa de un almacén de datos requiere una buena planeación y coordinación, antes de completar los esfuerzos de desarrollo. La primera área en la que hay que enfocarse es en el enfoque técnico, probar la aplicación y la migración de elementos desde el ambiente de desarrollo al ambiente de producción.

Un proceso sólido de pruebas, incluye diferentes tipos de prueba, incluyendo verificación de la calidad de los datos, del proceso de operaciones, desempeño, confiabilidad, usabilidad y de implementación. Una vez se han

finalizado las pruebas, se puede dar visto bueno para la implementación del sistema. Usualmente la implementación es un proceso muy simple para un sistema nuevo, pero puede ser una operación bastante complicada para mejoras y modificaciones al sistema de producción.

También hay que considerar las instalaciones en las computadoras de los usuarios, entrenamiento, y soporte. Una vez el sistema este en producción, es necesario considerar formar un equipo de soporte. Para que se encargue de responder dudas de los usuarios, resolver conflictos, dar capacitaciones, mantener el portal, etc.

Una vez se tiene implementada una primera iteración del almacén de datos y todos los sistemas, esto no necesariamente termina aquí. Tan pronto como sea posible, se deben empezar proyectos para añadir nuevo contenido al almacén de datos, y así alcanzar nuevos segmentos de la organización para que encuentren de utilidad el sistema. Es necesario encontrar un balance entre estos nuevos esfuerzos de desarrollo, así como el mantenimiento del sistema actual y el soporte, para mantener a todos los usuarios motivados a utilizar el sistema.

CONCLUSIONES

1. Se pudo observar que el inventario nacional de fuentes de agua es un esfuerzo descentralizado que se lleva a cabo por varias instituciones cada una con distintos enfoques de manera descentralizada, los directivos de estas instituciones se basan muchas veces en la misma información para la toma de decisiones, esto causa muchas coincidencias y redundancia de trabajo de campo para recabar los datos por una mala comunicación.
2. La realización de este proyecto, así como la emisión de la boleta SAS de cobertura, pueden ser tomadas como la base para definir un proceso estandarizado para la recolección, almacenamiento y publicación de la información centralizada del inventario nacional de fuentes de agua.
3. La unificación del inventario nacional de fuentes de agua es de beneficio no únicamente para las instituciones involucradas, si no que será de beneficio para el país en general, ya que muchas de estas instituciones tienen incidencia en las políticas o decisiones de los altos mandos a nivel nacional.
4. Con un sistema centralizado de información para el inventario se puede llegar a coordinar a las distintas instituciones para no tomar mediciones repetidas, evitar duplicar esfuerzos, priorizar zonas, prevenir a la población sobre desastres, generación de indicadores, etc. y de esta manera utilizar mejor los recursos presupuestarios, lo que finalmente permitirá aumentar la cobertura del inventario.

5. La realización de este proyecto fortaleció la iniciativa que la Secretaría Técnica ya tenía, es decir, que las organizaciones involucradas en el inventario nacional de fuentes de agua converjan en una iniciativa colectiva que permitirá centralizar un sistema integral y unificado.

RECOMENDACIONES

1. Trabajar en los aspectos actitudinales: buscar el bien común, innovación, fomentar la creatividad, entrega, y mejora continua; de los integrantes de las distintas instituciones, para poder desarrollar un proceso funcional que pueda alcanzar madurez.
2. Darle seguimiento a este proyecto o utilizarlo como base para la implementación y puesta en marcha de un sistema nacional de inventario de fuentes de agua.
3. Una vez el sistema esté implementado sea de acceso público, evitando la burocracia para que la población en general pueda informarse y tomar las decisiones oportunas para el cuidado de los recursos hídricos con los que cuenta el país.
4. Que el sistema a implementar contemple indicadores los cuales permitan tomar decisiones a distintos niveles: comunitarios, instituciones públicas y privadas, y organismos internacionales.

REFERENCIAS

- [1]. Wikipedia, La enciclopedia libre. *Metodología Top Down*, [en línea]. 2001.
<http://en.wikipedia.org/wiki/Data_warehouse#Top-down_design>
[Consulta: febrero 2009]
- [2]. — *Definición de Ralph Kimball*. [en Línea]. 2005.
<http://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos#Definici.C3.B3n_de_Ralph_Kimball> [Consulta: febrero 2009]
- [3]. — *Metodología Bottom Up*. [en línea]. 2005.
<http://en.wikipedia.org/wiki/Data_warehouse#Bottom-up_design>
[Consulta: febrero 2009]
- [4]. — *DSS Decision Support System*. [en línea]. 2005.
<http://es.wikipedia.org/wiki/Sistemas_de_soporte_a_decisiones>
[Consulta: febrero 2009]
- [5]. — *EIS Executive Information System*. [en línea]. 2005.
<<http://es.wikipedia.org/wiki/EIS>> [Consulta: febrero 2009]
- [6]. — *ERP Enterprise Resource Planning*. [en línea]. 2005.
http://en.wikipedia.org/wiki/Enterprise_resource_planning [Consulta: febrero 2009]

- [7]. ——— *OLTP OnLine Transaccion Processing*. [en línea]. 2005
<<http://es.wikipedia.org/wiki/OLTP>> [Consulta: febrero 2009]
- [8]. *Sinnexus business intelligence. Datamining*. [en línea]. 2007.
<http://www.sinnexus.com/business_intelligence/datamining.aspx>
[Consulta: febrero 2009]
- [9]. Wikipedia, La enciclopedia libre. *Libro blanco (WhitePaper)*, [en línea].
2005. <http://en.wikipedia.org/wiki/White_paper> [Consulta: febrero
2009]
- [10]. *The CBS interactive Business network, Hyperion Software and Arbor
Software finalize merger*, [en línea]. 1998.
http://findarticles.com/p/articles/mi_m0FOX/is_16_3/ai_50345495/
[Consulta: febrero 2009]
- [11]. Oracle. *Oracle Essbase*. [en línea]. 2002.
<[http://www.oracle.com/technetwork/middleware/essbase/overview/
index.html](http://www.oracle.com/technetwork/middleware/essbase/overview/index.html)> [Consulta: febrero 2009]
- [12]. *Microsoft Development Network. MDX Multi Dimensional Expressions*. [en
línea]. 2005. <[http://msdn.microsoft.com/es-
es/library/ms144785.aspx](http://msdn.microsoft.com/es-es/library/ms144785.aspx)> [Consulta: mayo 2009]
- [13]. Hispamedios. URL y URI. [en línea]. 2003.
<<http://www.hispamedios.com/articles/id34-url-y-uri>> [Consulta:
febrero 2009]

- [14]. W3C. *World Wide Web*. [en línea]. 1995. <www.w3.org> [Consulta: febrero 2009]
- [15]. Contratos informáticos. *Acuerdo de Nivel de Servicio en contratación*. [en línea]. 2004. <<http://www.contratosinformaticos.com/sla/>> [Consulta: febrero 2009]
- [16]. *Microsoft Development Network. Conceptos Básicos de la captura de datos modificados*. [en línea]. 2004. <<http://msdn.microsoft.com/es-es/library/cc645937.aspx>> [Consulta: febrero 2009]
- [17]. Wikipedia, la enciclopedia libre. *Middleware*. [en línea]. 2005. <<http://es.wikipedia.org/wiki/Middleware>> [Consulta: febrero 2009]
- [18]. Mapping interactivo, *La gestión de la información espacial mediante sistemas gestores de bases de datos relacionales*. [en línea]. 2003. <http://www.mappinginteractivo.com/plantilla-ante.asp?id_articulo=742> [Consulta: febrero 2009]
- [19]. Programa de las Naciones Unidas. *Objetivos del milenio*. [en línea]. 2000. <<http://www.undp.org/spanish/mdg/goallist.shtml>> [Consulta: abril 2009]
- [20]. *Tercer informe de avances en el cumplimiento de los Objetivos de desarrollo del milenio. Objetivo 7: Garantizar la sostenibilidad del medio ambiente*. Secretaria de Planificación y Programación de la Presidencia de la República SEGEPLAN. Guatemala: Serviprensa, 2009. 80p.

- [21]. *International Lake Environment Committee, Data Summary: Lago de Amatitlán.* [en línea]. 2004. <<http://www.ilec.or.jp/database/nam/dnam41.html>> [Consulta: abril 2009]
- [22]. IARNA, *Cantidad de agua en Guatemala.* [en línea]. 2004. <<http://www.infoiarina.org.gt/guateagua/>> [Consulta: abril 2009]
- [23]. Cuerpo de Ingenieros de los Estados Unidos de América. *Evaluación de los recursos de agua de Guatemala.* Guatemala, 2000. 106p.
- [24]. Instituto Nacional de Estadística, *Indicadores Medio Ambiente.* [en línea]. 2006. <<http://www.ine.gob.gt/index.php/ambiente>> [Consulta: abril 2009]
- [25]. Oficina regional de la FAO para América Latina y el Caribe. *Serie forestal N° 7.* Santiago, Chile. 1996. 10p.
- [26]. INSIVUMEH. *Estadísticas Históricas.* [en línea]. 2007. <<http://www.insivumeh.gob.gt/meteorologia/ESTADISTICAS.htm>> [Consulta: abril 2009]
- [27]. Wikipedia, la enciclopedia libre. *Georreferenciación.* [en línea]. 2005. <<http://es.wikipedia.org/wiki/Georreferenciación>> [Consulta: mayo 2009]

BIBLIOGRAFÍA

1. *Cross Industry Standard Process for Data Mining. CRISP-DM 1.0.* [en línea].1998.<<http://www.crisp-dm.org/>> [Consulta: mayo 2009]
2. *Estrategia Nacional de Gestión Integrada de los Recursos Hídricos.* Secretaria de Planificación y Programación de la Presidencia de la República SEGEPLAN, 2007. 31p.
3. *ETL tools. ETL process and concepts.* [en línea]. 2006.
<http://etl-tools.info/en/bi/etl_process.htm> [Consulta: febrero 2009]
4. *Global Water Partnership Technical Comité. Catalyzing Change. Global Water Partnership.* Estocolmo, Suecia, 2004. 48 p. ISBN: 91-974559-9.
5. *Global Water Parnership Technical Advisory Comitee. Water as a Social and Economic Good: How to Put the Principle into Practice.* Estocolmo Suecia 1998. 38 p. ISBN: 91-586-7620.
6. Guatemala. Acuerdo Gubernativo Número 204-2008. *Creación del Gabinete Específico del Agua*, 2008, 6p.
7. Instituto Nacional de Estadística y Geografía. *Metadatos.* [en línea]. 2001.
<<http://antares.inegi.org.mx/metadatos/metadat1.htm>> [Consulta: febrero 2009]

8. KIMBALL, Ralph. *The Data Warehouse Lifecycle Toolkit*. Segunda Edición. New York: *John Wiley & Sons*, 2002. 800 p. ISBN 0-471-25547-5.
9. MORALES-DE LA CRUZ, M.; PÉREZ, M.; CALVO, R.; DEVERS, D. “Sistema integrado de información y conocimiento del agua de Guatemala (SIAGua): aplicación del subsistema para tomar decisiones de inversión pública en agua potable y saneamiento”. En: *Memorias técnicas del XXVI Congreso Centroamericano de Ingeniería Sanitaria y Ambiental*. Guatemala 2009. 8 p.
10. *Política Nacional de Gestión integrada de los Recursos Hídricos. Guatemala*. Secretaria de Planificación y Programación de la Presidencia de la República SEGEPLAN, 2007. 31p.
11. PYLE, Dorian. *Business Modeling and Data Mining*. Primera Edición. Estados Unidos: *Morgan Kaufmann*, 2003. 650 p. ISBN 1-55860653X
12. Wikipedia, La enciclopedia libre. *Almacenes de datos*. [en línea]. 2003.<http://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos> [Consulta: febrero 2009]
13. — *Data Mart*. [en línea]. 2005. <http://es.wikipedia.org/wiki/Data_mart> [Consulta: febrero 2009]
14. — *Definición Técnica de un cubo OLAP*. [en línea]. 2005. <http://es.wikipedia.org/wiki/Cubo_OLAP> [Consulta: febrero 2009]

15. — *Manipulación de metadatos.* [en línea]. 2005.
<<http://es.wikipedia.org/wiki/Metadato>> [Consulta: febrero 2009]
16. — *Tabla de hechos.* [en línea]. 2005.
<http://es.wikipedia.org/wiki/Tabla_de_hechos#Las_medidas_del_negocio_.28hechos.29> [Consulta: febrero 2009]

ANEXO A

Acuerdo Gubernativo 204-2008 Creación Gabinete Del Agua

| Fundado en 1880 |

Diario de Centro América

DECANO DE LA PRENSA CENTROAMERICANA | ÓRGANO OFICIAL DE LA REPÚBLICA DE GUATEMALA, C. A.

www.dca.gob.gt Directora General: Ana María Rodas

Sumario

ORGANISMO EJECUTIVO

PRESIDENCIA DE LA REPÚBLICA

Acuérdase CREAR EL GABINETE ESPECÍFICO DEL AGUA.

MINISTERIO DE GOBERNACIÓN

Acuérdase emitir el siguiente REGLAMENTO DE LA LEY DE LA DIRECCIÓN GENERAL DE INTELIGENCIA CIVIL.

MINISTERIO DE EDUCACIÓN

Acuérdase crear el programa de formación musical denominado "Pentagrama".

MINISTERIO DE ENERGÍA Y MINAS

Acuérdase otorgar Autorización Definitiva a la entidad Central Hidroeléctrica Sulín, Sociedad Anónima, para utilizar bienes de dominio público para el proyecto hidroeléctrico denominado "Hidroeléctrica Sulín", ubicado en el municipio de Purulhá del departamento de Baja Verapaz.

PUBLICACIONES VARIAS

CORTE DE CONSTITUCIONALIDAD

EXPEDIENTE 541-2006

EXPEDIENTE No. 2426-2008

MUNICIPALIDAD DE RETALHULEU

LEY TEMPORAL Y ESPECIAL DE REPOSICIÓN DE CÉDULAS DE VECINDAD DEL MUNICIPIO DE RETALHULEU, SEGÚN EL DECRETO 30-2006 DE FECHA 21 DE SEPTIEMBRE DE 2006.

ANUNCIOS VARIOS

Matrimonios • Líneas de Transporte • Constituciones de Sociedad • Modificaciones de Sociedad • Disolución de Sociedad • Patentes de Invención • Registro de Marcas • Títulos Supletorios • Edictos • Remates •

Diario de Centro América

Las publicaciones que se realizan en el Diario de Centro América, se publican de conformidad con el original presentado por el solicitante, en consecuencia cualquier error que se cometa en ese original, el Diario de Centro América no asume ninguna responsabilidad.

Por lo antes descrito se les solicita cumplir con los siguientes requisitos:

1. Tamaño de letra según Acuerdo Gubernativo No. 163-2001, no menor de 6.5 (Letra Tipográfica).
2. Letra clara e impresión firme.
3. Legibilidad en los números.
4. No correcciones, tachones, marcas de lápiz o lapicero.
5. No se aceptan fotocopias.
6. Que la firma de la persona responsable y sello correspondiente se encuentren fuera del texto del documento.
7. Documento con el nombre completo del Abogado, Sello y Número de Colegiado.
8. Nombre y Número de teléfono de la persona responsable de la publicación, para cualquier consulta posterior.

Dirección

ORGANISMO EJECUTIVO



PRESIDENCIA DE LA REPÚBLICA

Acuérdase crear el GABINETE ESPECÍFICO DEL AGUA

ACUERDO GUBERNATIVO NÚMERO 204-2008

Guatemala, 1 de agosto de 2008

EL VICEPRESIDENTE DE LA REPÚBLICA EN FUNCIONES DE PRESIDENTE

CONSIDERANDO:

Que conforme la Constitución Política de la República, el aprovechamiento de las aguas es un asunto de interés social. Los recursos hídricos, como bien natural, cumplen funciones multidimensionales que requieren acciones intersectoriales; por lo tanto, debe asegurarse que los diferentes usos y medidas de conservación del agua contribuyan a mejorar las condiciones de desarrollo humano intergeneracional, al crecimiento económico y al enriquecimiento ambiental.

CONSIDERANDO:

Que es necesaria la acción gubernamental asociada con la gestión integral de las aguas territoriales, dentro de un conjunto de elementos físicos, hidrológicos, económicos, sociales, ambientales, culturales y políticos, respecto a los cuales se prevean medidas para su aprovechamiento eficiente y eficaz, acceso equitativo, protección y conservación, precisamente para garantizar en el tiempo el abasto para generaciones futuras en un contexto de armonía social.

CONSIDERANDO:

Que de conformidad con la Ley del Organismo Ejecutivo, los gabinetes específicos son creados para la formulación de propuestas y la coordinación del diseño, gestión e implementación de políticas, programas y acciones en determinada materia o sector.

POR TANTO:

En ejercicio de las funciones contenidas en los incisos e) y m) del artículo 183 de la Constitución Política de la República, y 5 y 18 del Decreto número 114-97 del Congreso de la República, Ley del Organismo Ejecutivo.

EN CONSEJO DE MINISTROS

ACUERDA

Artículo 1. CREACIÓN. Se crea el GABINETE ESPECÍFICO DEL AGUA, con el propósito de coordinar los esfuerzos gubernamentales de diseño y gestión de políticas, planes y presupuestos del agua, para contribuir al logro de metas y objetivos de desarrollo nacional.

Artículo 2. INTEGRACIÓN: El Gabinete Específico del Agua se integra por los siguientes funcionarios:

- Vicepresidente de la República, quien lo preside y coordina. En caso de ausencia del Vicepresidente de la República, éste designará al Ministro que en su representación, coordinará el Gabinete Específico;
- Ministro de Agricultura, Ganadería y Alimentación;
- Ministro de Ambiente y Recursos Naturales;
- Ministro de Comunicaciones, Infraestructura y Vivienda;
- Ministro de Cultura y Deportes;
- Ministro de Economía;
- Ministro de Educación;
- Ministro de Energía y Minas;
- Ministro de Finanzas Públicas;
- Ministro de Relaciones Exteriores;
- Ministro de Salud Pública y Asistencia Social;
- Secretario de Seguridad Alimentaria y Nutricional de la Presidencia de la República;
- Secretario de Coordinación Ejecutiva de la Presidencia de la República;
- Secretario de Asuntos Agrarios;
- Secretario de Comunicación Social;
- Secretario Ejecutivo de la Coordinadora Nacional para la Reducción de Desastres;
- Secretario Ejecutivo del Consejo Nacional de Áreas Protegidas; y,
- Secretario de Planificación y Programación de la Presidencia de la República, quien fungirá como Secretario Técnico.

El Vicepresidente de la República podrá asimismo, invitar a integrarse al Gabinete Específico del Agua, en forma temporal o para algún asunto específico, a cualquier otro funcionario gubernamental, de entidades autónomas y descentralizadas, o representantes de entidades de carácter privado o social, cuya naturaleza o ámbito de acción sea de interés a la materia.

Artículo 3. OBJETIVOS. El Gabinete Específico del Agua tiene como objetivos:

- Promover la adopción e implementación de criterios de gestión integrada de los recursos hídricos en todas las entidades públicas, centralizadas y descentralizadas;
- Asegurar la contribución del agua para el logro de las metas y objetivos del desarrollo nacional, y para ello, priorizar las acciones que favorezcan a su cumplimiento;
- Propiciar la coordinación de las acciones de las entidades gubernamentales, autónomas y de carácter civil o privado, para lograr un manejo sostenible de los recursos hídricos del país;
- Velar por la ejecución armónica de los recursos financieros y humanos que se destinen a la gestión hídrica;
- Promover el fortalecimiento institucional y la participación ciudadana en el sector hídrico del país, a fin de favorecer la gobernabilidad en la materia.

Artículo 4. FUNCIONES. El Gabinete Específico del Agua deberá desarrollar todas las funciones que sean necesarias para la consecución de los objetivos relacionados en el artículo anterior, realizando las actividades que así lo requieran. De esa cuenta, sus funciones principales son:

- Emitir instructivos y lineamientos para la efectiva gestión hídrica, de carácter general o para temas específicos, a fin de establecer protocolos y directrices para las dependencias gubernamentales y entidades descentralizadas. A fin de propiciar su adopción en las entidades autónomas y las de carácter privado o social, para la elaboración de los instructivos y lineamientos deberá invitarse a participar en las reuniones del Gabinete Específico a personeros de las mismas;
- Establecer los mecanismos de control y seguimiento necesarios para asegurar la adecuada utilización del recurso hídrico del país en función del desarrollo sostenible y ecológicamente equilibrado;
- Diseñar o identificar las herramientas de política, planificación y presupuesto que permitan la eficaz y eficiente coordinación de acciones y evaluación anual del desempeño público en la gestión hídrica;
- Diseñar una estrategia de movilización y captación de fondos nacionales, bilaterales y multilaterales para implementar las medidas y acciones adoptadas en el Gabinete Específico del Agua;
- Promover el diálogo para favorecer la gobernabilidad del agua, sobre la base del fortalecimiento institucional del sector;
- Diseñar y velar por la implementación de una estrategia de participación ciudadana en los niveles pertinentes de la gestión hídrica, principalmente en el propositivo y en el de auditoría y fiscalización. La estrategia también deberá contemplar la participación del gobierno local en la gestión hídrica;
- Revisar, actualizar y velar por la implementación de las acciones contenidas en la Política Nacional de Gestión Integrada de los Recursos Hídricos (PNGIRH) y la Estrategia Nacional de Gestión Integrada de los Recursos Hídricos (ENGRH);
- Formular las propuestas respectivas a efecto de lograr una adecuada y eficaz regulación del tema del Agua, en forma acorde a los principios del desarrollo sostenible y ecológicamente equilibrado.

Artículo 5. REUNIONES. El Gabinete Específico del Agua se reunirá ordinariamente una vez al mes conforme la programación que para el efecto elabore el Vicepresidente de la República, quien se encargará de la convocatoria respectiva. Extraordinariamente, se reunirá cuantas veces sea necesario, por convocatoria del Presidente de la República o del Vicepresidente de la República.

Artículo 6. COLABORACIÓN GUBERNAMENTAL. Dada la importancia y trascendencia que para el desarrollo sostenible y ecológicamente equilibrado tiene la gestión hídrica, todos los ministerios y entidades descentralizadas del Estado deberán prestar la máxima colaboración al Gabinete Específico del Agua, en lo que fueren requeridos.

Artículo 7. SECRETARÍA TÉCNICA. La Secretaría General de Planificación y Programación de la Presidencia –SEGEPLAN-, por medio de su Dirección de Recursos Hídricos y como parte de sus actividades, tendrá a su cargo la Secretaría Técnica del Gabinete del Agua; coordinará el trabajo de las diferentes comisiones que se integren dentro de este gabinete; y se encargará de la presentación de los proyectos respectivos, para su aprobación.

Cualquier funcionario o empleado público de la Dirección de Recursos Hídricos de la Secretaría General de Planificación y Programación de la Presidencia –SEGEPLAN- que colabore con la Secretaría Técnica del Gabinete Específico del Agua realizará tales funciones en forma *ad honorem*. Los costos y gastos que el funcionamiento de la Secretaría Técnica del Gabinete del Agua requiera serán cubiertos por la Secretaría General de Planificación y Programación de la Presidencia –SEGEPLAN- con los recursos correspondientes al funcionamiento de su Dirección de Recursos Hídricos.

Artículo 8. INFORMES. El Gabinete Específico del Agua deberá rendir informe bimensual al Presidente de la República sobre el avance o estado que guardan las políticas que se implementan para la consecución de sus objetivos.

Artículo 9. PLAZO. El gabinete específico que por medio de este Acuerdo Gubernativo se crea funcionará por un plazo de cuatro años.

Artículo 10. VIGENCIA. El presente Acuerdo Gubernativo entrará en vigencia el día siguiente al de su publicación en el Diario de Centro América.

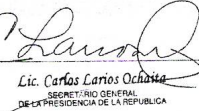
COMUNIQUESE



RAFAEL ESPADA



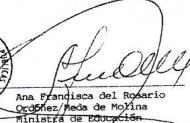

Lic. Arnaldo Villagrán
Encargado del Despacho
Ministerio de Gobernación

Lic. Carlos Larín Ochaita
SECRETARIO GENERAL
DE LA PRESIDENCIA DE LA REPUBLICA



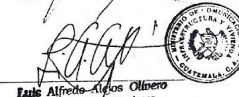
Juan Alberto Fuentes K.
MINISTRO DE FINANZAS PUBLICAS

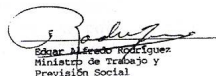
Ana Francisca del Rosario
Encargada de la Oficina
Ministerio de Turismo



Rosalío Mauricio Illescas García
Jefe del Estado Mayor de la
Defensa Nacional
Encargado del Despacho

Elio Alfredo-Angel Olivero
Ministro de Comunicaciones,
Infraestructura y Vivienda

Edgar Alfredo Rodríguez
Ministerio de Trabajo y
Previsión Social



Romulo Alfredo Caballero Otero
Ministerio de Economía



MINISTERIO DE GOBERNACIÓN
 Acuérdose emitir el siguiente REGLAMENTO DE LA LEY DE LA DIRECCIÓN GENERAL DE INTELIGENCIA CIVIL.

ACUERDO GUBERNATIVO NÚMERO 203-2008

Guatemala, 30 de julio del 2008

EL PRESIDENTE DE LA REPÚBLICA,

CONSIDERANDO:

Que mediante Decreto Número 71-2005 del Congreso de la República, se emitió la Ley de la Dirección General de Inteligencia Civil, en la cual se creó la Dirección General de Inteligencia Civil y se establecieron sus bases jurídicas, orgánicas y funcionales.

CONSIDERANDO:

Que en el artículo 29 de la ley citada, se dispuso que el Organismo Ejecutivo debe emitir el Reglamento de la misma, por lo que en cumplimiento de esa estipulación se emite la disposición legal correspondiente.

POR TANTO:

En ejercicio de las funciones que le confiere el Artículo 183 literal e) de la Constitución Política de la República de Guatemala y con fundamento en el Artículo 29 del Decreto Número 71-2005 del Congreso de la República, Ley de la Dirección General de Inteligencia Civil,

ACUERDA:

Emitir el siguiente,

REGLAMENTO DE LA LEY DE LA DIRECCIÓN GENERAL DE INTELIGENCIA CIVIL

TÍTULO I

Disposiciones Generales

CAPÍTULO ÚNICO

Artículo 1. **OBJETO.** El presente Reglamento tiene como objeto desarrollar la Ley de la Dirección General de Inteligencia Civil en cuanto a determinar la estructura orgánica, funciones y mecanismos de coordinación, así mismo las disposiciones necesarias para lograr el eficiente y eficaz cumplimiento de las funciones que legalmente le corresponde desarrollar.

Artículo 2. **NATURALEZA.** La Dirección General de Inteligencia Civil, en adelante también denominada DIGICI, es el ente de carácter civil, constituido por personal civil, para asesorar en materia de inteligencia civil al Ministerio de Gobernación en la toma de decisiones y la formulación de las políticas y planeamientos orientados a la prevención, control, y combate del crimen organizado y de la delincuencia común.

Artículo 3. **FUNCIONES.** A la Dirección General de Inteligencia Civil le corresponde las funciones establecidas en la Ley de la Dirección General de Inteligencia Civil y otras leyes que le sean aplicables. Las actividades y procedimientos de la DIGICI son considerados como asuntos de seguridad nacional, por lo que se deberá cumplir con la confidencialidad que ordena la ley.

Artículo 4. **CONTROL INTERNO.** Independientemente de las instancias de control interno que la Dirección General de Inteligencia Civil cree, corresponde al Ministerio de Gobernación, a través del Primer Viceministro, la supervisión permanente del cumplimiento de las actividades de inteligencia de la DIGICI.

TÍTULO II

ORGANIZACIÓN

CAPÍTULO I

ESTRUCTURA ORGÁNICA

Artículo 5. **ESTRUCTURA ADMINISTRATIVA.** Para cumplir sus funciones, la Dirección General de Inteligencia Civil, se estructura y organiza en la forma siguiente:

1. Dirección Superior. Se integra por:

- 1.1 Dirección General
- 1.2 Sub-Dirección General.

2. Órganos Sustantivos. Se integran por:

- 2.1. División de Inteligencia.
 - 2.1.1. Departamento de Obtención de Información.
 - 2.1.2. Departamento de Apoyo Técnico.
 - 2.1.3. Departamento de Procesamiento de la Información.
- 2.2. División de Contrainteligencia.
 - 2.2.1. Departamento de Riesgos Externos.
 - 2.2.2. Departamento de Riesgos Internos.
- 2.3 División de Planes.
 - 2.3.1. Departamento de Planes de Inteligencia.
 - 2.3.2. Departamento de Planes de Contrainteligencia.
 - 2.3.3. Departamento de Planes para Apoyo Técnico.

3. Órganos Administrativos. Se integran por:

- 3.1. Secretaría General.
- 3.2. División Financiera.
 - 3.2.1. Departamento de Tesorería.
 - 3.2.2. Departamento de Presupuesto.
 - 3.2.3. Departamento de Contabilidad.
- 3.3. División de Administración de Personal.
 - 3.3.1. Departamento de Reclutamiento, Selección y Gestión de Personal.
 - 3.3.2. Departamento de Formación y Capacitación.
- 3.4. División de Logística.
 - 3.4.1. Departamento de Armería.
 - 3.4.2. Departamento de Mantenimiento.
 - 3.4.3. Departamento de Transporte.
 - 3.4.4. Departamento de Inventarios.
 - 3.4.5. Departamento de Compras.
 - 3.4.6. Departamento de Almacén.
- 3.5 División de Seguridad.
 - 3.5.1. Departamento de Seguridad de la Información.
 - 3.5.2. Departamento de Seguridad de las Instalaciones.
 - 3.5.3. Departamento de Seguridad de los Funcionarios y Empleados.

4. Órganos de Apoyo Técnico. Se integran por:

- 4.1. División de Asesorías Legales y Técnicas.
 - 4.1.1. Departamento de Asesoría Legal.
 - 4.1.2. Departamento de Asesoría Técnica.
- 4.2. División de Informática.
 - 4.2.1. Departamento Técnico y Mantenimiento de Informática.
 - 4.2.2. Departamento de Desarrollo, Automatización y Sistema de Información Georeferencial.
 - 4.2.3. Departamento de Seguridad Informática y Redes.

5. Órganos de Control. Se integran por:

- 5.1. División de Asuntos Internos.
 - 5.1.1. Departamento de Control Disciplinario de Personal.
 - 5.1.2. Departamento de Control del Procedimiento de la Información.
- 5.2. Unidad de Auditoría Interna.
- 5.3. Unidad de Control de Eficiencia.

CAPÍTULO II

DIRECCIÓN SUPERIOR

Artículo 6. **Despacho del Director General.** La Dirección General de Inteligencia Civil estará a cargo de un Director General, quien como autoridad administrativa superior, es el responsable de la dirección y administración de la institución, a quien le corresponde proponer a la autoridad nominadora, el nombramiento y remoción del personal de la DIGICI, así como dirigir sus unidades administrativas asegurando la adecuación de sus actividades a las funciones, atribuciones y cumplimiento de sus objetivos, establecidos en la Ley de la Dirección General de Inteligencia Civil y el presente Reglamento.

Artículo 7. **Responsabilidades del Director General.** El Director General de Inteligencia Civil al proporcionar inteligencia a otros entes del Estado para la prevención, control y combate de las acciones de la delincuencia común y organizada, de conformidad con la Ley de la Dirección General de Inteligencia Civil, determinará a su juicio la relevancia de la información que debe ser trasladada y dará cuenta inmediata al Primer Viceministro de Gobernación.

Es responsabilidad del Director General de Inteligencia Civil determinar cuándo una información es considerada reservada y establecer los niveles de acceso que de ella se tenga a lo interno de la DIGICI.

Artículo 8. **Despacho del Subdirector General.** La Subdirección General estará a cargo del Subdirector General de Inteligencia Civil, a quien le corresponde las atribuciones que determina la Ley de la Dirección General de Inteligencia Civil y sustituirá al Director General de Inteligencia Civil en casos de ausencia temporal.

ANEXO B

Boleta SAS de cobertura

DIRECCIÓN GENERAL DE REGULACIÓN, VIGILANCIA Y CONTROL DE LA SALUD
DEPARTAMENTO DE REGULACIÓN DE LOS PROGRAMAS DE LA SALUD Y AMBIENTE

BOLETA- SAS- DE COBERTURA (Agua y Saneamiento)

Instrucciones: Llenar la boleta claramente con LETRA DE MOLDE, de acuerdo al instructivo adjunto.

Nombre(s) de encuestador(es): _____ Fecha (dd/mm/aa): _____

Institución(es) responsable(s): _____ Departamento/Unidad: _____

1. DATOS GENERALES DE LA COMUNIDAD:

Comunidad: _____ Municipio: _____

Departamento: _____ Código INE (ver manual): _____

Coordenadas geográficas: LAT _____ LONG _____

Modelo del GPS: _____ Dátum: Asegurar que es el WGS84

Altitud en metros sobre el nivel del mar (m snm): _____

1.1 Tipo de comunidad

Rural Urbana Marginal Aldea Caserío Otro _____

1.2 Acceso a la comunidad

Pavimento Terracería Vereda Kilómetros a la cabecera municipal _____

1.3 Demografía de la comunidad

Número de habitantes Número Total de Viviendas

1.4 Idioma Predominante de la comunidad

Español K'iche' Q'eqchi' Kaqchikel Mam Otro _____

1.5 Energía eléctrica de la comunidad

¿Existe energía eléctrica en la comunidad? Sí No

2. SISTEMA DE ABASTECIMIENTO DE AGUA

¿Hay sistema de abastecimiento de agua? Sí No En caso NO ¿cómo se abastece? _____

Si hay Sistema ¿quién lo construyó? _____ Año de construcción: _____

2.1 Tipo de Sistema

Gravedad Bombeo Mixto Otro _____



MINISTERIO DE SALUD PÚBLICA Y ASISTENCIA SOCIAL

2.2 Cantidad y Tipo de Conexiones de Agua

	Número según "Tipo de Conexión"			
	Domiciliar	Predial	Llena cántaro	Otro
Número de viviendas CON conexión y agua				
Número de viviendas CON conexión pero SIN agua				

NOTA: Colocar el número y entre paréntesis la letra del tipo de sistema

Número de viviendas SIN conexión

Tipo de sistema
G: Gravedad
B: Bombeo
M: Mixto

2.3 Funcionamiento del Sistema de Abastecimiento de Agua

¿Cómo es la continuidad del servicio? Intermittente Continuo

¿Se le da mantenimiento al servicio? Si No

¿Existe fontanero a cargo del servicio? Si No

Responsable del funcionamiento del sistema: Comunidad Municipalidad Otro _____

Nombre de la persona responsable del funcionamiento del sistema: _____

¿Cuánta agua se clora en el sistema de abastecimiento?

La de todo el sistema La de una parte del sistema No se clora el agua

¿Se paga alguna tarifa por el consumo de agua? Si No

En caso afirmativo, ¿cuál es la cuota mensual? _____

3. SANEAMIENTO

3.1 Saneamiento existente

¿Hay saneamiento para la disposición de excretas en la comunidad? Si No

Si la comunidad cuenta con saneamiento básico indique lo siguiente:

Tipo de letrina	Número total de letrinas en la comunidad	Número de letrinas y estado en que se encuentran	
		Bueno	Malo
Letrina de Hoyo (LT)			
Letrina de hoyo seco Ventilada (LHSV)			
Letrina abonera seca Familiar (LASF)			
Letrina de sello hidráulico Lavable (LSHL)			



MINISTERIO DE SALUD PÚBLICA Y ASISTENCIA SOCIAL

4. MUESTREO DE LA CALIDAD DEL AGUA EN LA COMUNIDAD

Fecha muestreo: _____ Realizado por: _____

Punto de muestreo	Turbiedad [UNT]	Cloro residual [mg/L]	pH	Coliformes Fecales / 100 mL			
				No. De muestra	Volumen	Contaje	Cantidad
Tanque de almacenamiento (T1)							
Tanque de almacenamiento (T2)							
Vivienda (V1)							
Vivienda (V2)							
Tipo de fuente (F1): _____							
Tipo de fuente (F2): _____							

5. FUENTES DE RECURSOS HIDRICOS

5.1 Ubicación de la fuente

Coordenadas geográficas: LAT _____ LONG _____

Modelo del GPS: _____ Dátum: Asegurar que es el WGS84

Altitud (m snm): _____ Nombre de la fuente: _____

5.2 Datos del Tipo de la fuente

Manantial/nacimiento Río Lago Laguna Embalse

Pozo perforado con máquina Pozo excavado

5.3 Método de Aforo

Flotador Molinete Otro

Cubeta Volumen _____

Distancia [m]	Ancho [m]	Tiempo [seg]	Profundidad [m]
		1	1
		2	2
		3	3
			4
			5

Tiempo _____

Tiempo	Volumen
1	1
2	2
3	3

ml seg litros galones

Caudal promedio de la fuente: _____ m³/s L/s gpm

5.4 Si la fuente es un pozo mecánico o excavado (artesiano)

Caudal extraído del pozo mecánico: _____ m³/s L/s gpm

Si hay pozo artesiano ¿a qué profundidad promedio está el agua?

En época lluviosa [m] _____ En época seca [m] _____



MINISTERIO DE SALUD PÚBLICA Y ASISTENCIA SOCIAL

5.5 Otra información sobre la fuente

- Actualmente, ¿Está en uso la fuente? Si No
- Actualmente ¿Tiene cobertura vegetal? Si No
- Condiciones de la fuente Mala Regular Buena
- Estatus legal Con escritura pública Con escritura registrada Acta municipal
Convenios verbales

6. OBSERVACIONES GENERALES

7. VISTO BUENO DE LA ORGANIZACIÓN COMUNITARIA

- ¿Existe organización comunitaria? Si No Nombre (en caso afirmativo): _____
- ¿Se dedica la organización a actividades de agua y saneamiento? Si No
- Si la organización no se dedica a "agua y saneamiento" ¿podría hacerlo? Si No
- Miembros del comité u organización comunitaria

Nombre	Sexo	Cargo	Teléfono o forma de contactarlo

f. _____
Firma del encuestador

Firma y sello del COCODE, u autoridad local



Fuente: Secretaria Técnica del Gabinete Específico del Agua. <http://www.segeplan.gob.gt/2.0/>, febrero de 2009.

ANEXO C

Boleta PROVIAGUA



BOLETA POR COMUNIDAD "PROVIAGUA-COM-001"

Fecha de la Inspección: _____

RESPONSABLE

Nombre: _____ Cargo: _____ Servicio de Salud: _____

COMUNIDAD

Nombre: _____ Tipo (Aldea, Caserío, etc.): _____

Área de Salud: _____ Municipio: _____

DEMOGRAFÍA

Población Aproximada: _____ Número Aproximado de Familias: _____

ABASTECIMIENTO DE AGUA

Existen Sistemas Formales (Sí/No): _____ Existen Sistemas Informales (Sí/No): _____

SISTEMAS FORMALES DE ABASTECIMIENTO

Nombre: _____ Tipo (Gravedad/Bombeo/Mixto): _____

Nombre: _____ Tipo (Gravedad/Bombeo/Mixto): _____

Nombre: _____ Tipo (Gravedad/Bombeo/Mixto): _____

SISTEMAS INFORMALES DE ABASTECIMIENTO

Fuente: _____ Ubicación: _____

Fuente: _____ Ubicación: _____

Fuente: _____ Ubicación: _____

Fuente: _____ Ubicación: _____

Fuente: _____ Ubicación: _____

Fuente: _____ Ubicación: _____





BOLETA POR SISTEMA "PROVIAGUA-SIS-001"

Fecha de la Inspección: _____

RESPONSABLE

Nombre: _____ Cargo: _____ Servicio de Salud: _____

COMUNIDAD

Nombre: _____ Tipo (Aldea, Caserío, etc.): _____
Área de Salud: _____ Municipio: _____

DEMOGRAFIA

Población Aproximada: _____ Número Aproximado de Familias: _____

SISTEMA DE ABASTECIMIENTO

Nombre o Identificación: _____ Tipo (Formal/Informal): _____

Fuente: _____ Ubicación: _____

Fuente: _____ Ubicación: _____

Fuente: _____ Ubicación: _____

DESINFECCIÓN

¿Desinfectan el agua? (Sí/No/Parcial): _____ Indique el método utilizado: _____

Si la respuesta es "No" y/o "Parcial", indique la razón: _____

CONEXIONES

Intradomiciliares (Sí/No): _____ Prediales (Sí/No): _____ Lienacántaros (Sí/No): _____

CONSTRUCCIÓN

¿Qué entidad fue la responsable? _____ Año de Construcción: _____

ADMINISTRACIÓN

Administrado por: _____ ¿Funciona el sistema? (Sí/No/Irregular): _____

Firma del Responsable



Vo. Bo. Jefe Inmediato



BOLETA DE CALIDAD DEL AGUA "PROVIAGUA-CAL-001"

Fecha de la Inspección: _____

RESPONSABLE

Nombre: _____ Cargo: _____ Servicio de Salud: _____

COMUNIDAD

Nombre: _____ Tipo (Aldea, Caserío, etc.): _____

Área de Salud: _____ Municipio: _____

DEMOGRAFÍA

Población Aproximada: _____ Número Aproximado de Familias: _____

SISTEMA DE ABASTECIMIENTO

Nombre o Identificación: _____

CALIDAD DEL AGUA EN LA(S) FUENTE(S)

Fuente: _____	Coliformes Fecales (Colonias por 100 mL): _____
Fuente: _____	Coliformes Fecales (Colonias por 100 mL): _____
Fuente: _____	Coliformes Fecales (Colonias por 100 mL): _____

CALIDAD DEL AGUA EN EL SISTEMA

Tanque 1	Cloro Residual (mg/L): _____	Coliformes Fecales (Colonias por 100 mL): _____
Tanque 2	Cloro Residual (mg/L): _____	Coliformes Fecales (Colonias por 100 mL): _____
Tanque 3	Cloro Residual (mg/L): _____	Coliformes Fecales (Colonias por 100 mL): _____
Vivienda 1	Cloro Residual (mg/L): _____	Coliformes Fecales (Colonias por 100 mL): _____
Vivienda 2	Cloro Residual (mg/L): _____	Coliformes Fecales (Colonias por 100 mL): _____
Vivienda 3	Cloro Residual (mg/L): _____	Coliformes Fecales (Colonias por 100 mL): _____

CALIDAD DEL AGUA EN OTROS PUNTOS

Camión Cisterna	Cloro Residual (mg/L): _____	Coliformes Fecales (Colonias por 100 mL): _____
Fábrica de Hielo	Cloro Residual (mg/L): _____	Coliformes Fecales (Colonias por 100 mL): _____
Envasadora de Agua	Cloro Residual (mg/L): _____	Coliformes Fecales (Colonias por 100 mL): _____



Firma del Responsable

PROGRAMA NACIONAL DE VIGILANCIA DE LA CALIDAD DEL AGUA

Vo. Bo. Jefe Inmediato

ANEXO D

Boleta PARPA

BOLETA

Nombre del Encuestador	Institución Responsable	
Fecha (dd/mm/aa)	Hora de Inicio	Departamento/Unidad

SECCIÓN A

Parte I: Datos Generales

1. Departamento: _____ 2. Municipio: _____

3. Comunidad/Fincas: _____

SECCIÓN B

Parte I: Ubicación de la fuente

4. Coordenadas UTM x: _____ y: _____

5. Modelo de GPS: _____

6. Altitud: _____

7. Nombre de la Fuente: _____

Parte II: Datos del tipo de fuente

8. Tipo de la fuente: _____

9. Método de afloramiento: _____

Distancia (mts)	Ancho (mts)	Tiempo	Profundidad (mts)

Cubeta

Tiempo	Volumen

10. Observaciones: _____

11. Caudal promedio: _____

12. Tipo de usuario: _____

13. Estructura de protección en la fuente: _____

14. ¿Tiene cobertura vegetal? _____

15. ¿Actualmente la fuente está en uso? _____

16. Estatus legal de la fuente: _____

17. Condiciones de la fuente: _____

Parte III: Si la fuente es un pozo mecánico

18. Volumen de extracción: _____ 19. Potencia de la bomba: _____ HP

20. Horas de uso: _____ 21. La bomba es: _____

SECCIÓN C: Usos de la Fuente (No continúe si la fuente no está en uso)

Tomas y caudales

22. Tipo de captación: _____

23. Cuadro No. 1

Toma #	Toma #	Toma #
Coordenadas de salida: X Y	Coordenadas de salida: X Y	Coordenadas de salida: X Y
Elevación de la salida:	Elevación de la salida:	Elevación de la salida:
Tipo de sección:	Tipo de sección:	Tipo de sección:
Materiales:	Materiales:	Materiales:
Observaciones:	Observaciones:	Observaciones:

! Tipo de sección: T = Tubería CS = Canal semicircular CR = Canal rectangular CT = Canal trapezoidal, en caso triangular base 0.

() Si su uso es doméstico, llene únicamente sección D. Si la pregunta no aplica a su situación coloque "NA".
 () Si su uso es para irrigación, llene únicamente la sección E. Si la pregunta no aplica a su situación coloque "NA".
 () Si su uso es para hidroeléctrica, llene únicamente la sección F. Si la pregunta no aplica a su situación coloque "NA".
 () Si su uso es industrial, recreativo, pecuario u otro, llene únicamente la sección G. Si la pregunta no aplica coloque "NA".

SECCIÓN D: Uso doméstico

24. Nombre de la población beneficiada: _____

25. Número de población beneficiada: _____

