



UNIVERSIDAD NACIONAL
AVENIDA DE
MEXICO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**PROGRAMA DE MAESTRÍA Y DOCTORADO EN
INGENIERÍA**

FACULTAD DE INGENIERÍA

**Base de datos histórica de localidades de la
República Mexicana para georeferenciar
colectas biológicas**

T E S I S

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN INGENIERÍA (ELÉCTRICA)

P R E S E N T A:

EDUARDO ORDUÑA JARAMILLO

TUTOR:

DR. JORGE PRADO MOLINA



México, D.F. 2010



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

Con cariño a la Mtra. Aída Rivera González por el apoyo moral que me brindó durante todo el proceso de elaboración de esta tesis, por todos sus consejos y su profesionalismo.

Especialmente al Licenciado en Ciencias de la Computación Josafat Isaí Guerrero Iñiguez por dedicarme todo su tiempo y paciencia para la realización de esta tesis, ya que sin él no hubiera sido posible terminarla.

Al Dr. Jorge Prado Molina por toda la confianza depositada en mí y por toda su dedicación y entusiasmo vertidos en esta tesis.

Como siempre y con cariño a Mary y Lalin (E2) porque han estado conmigo cuando más los he necesitado.

A Dios porque para él no hay imposibles.

Dedicatoria

Para Ángel y Sofía con mucho amor.

Eduardo Orduña Jaramillo

“No es lo que yo quiera, es lo que es”

**Base de datos histórica de localidades de la
República Mexicana para georeferenciar
colectas biológicas**

Contenido

Resumen.....	v
Prefacio.....	vi
Objetivo general	vii
Objetivos específicos.....	vii
Metas	vii
Alcances	viii
Metodología	viii
Hipótesis	ix
Mapa.....	ix
1. Introducción.....	1
1.1 Colectas biológicas	1
1.2 Georeferenciación de Localidades.....	2
1.3 Gaceteros digitales y proyectos relacionados.....	3
1.3.1. INEGI (Instituto Nacional de Estadística, Geografía e Informática)	4
1.3.2 GeoLocate.....	7
1.3.3. BioGeomancer	8
1.3.4 UNIGEO (Unidad de Informática Geoespacial)	9
1.3.5 ADL (Alexandria Digital Library)	10
1.4 Estructura de la información dentro de un gacetero digital	11
1.5 Acceso a la información en gaceteros digitales	12
1.6 Planteamiento del problema.....	13
1.6.1 Definición del problema	13
1.6.2 Justificación.....	15
2. Georeferenciación.....	16
2.1. Esquemas de códigos geoespaciales - coordenadas y Sistemas de coordenadas 18	
2.1.1 Propiedades y Elementos de la Cartografía General	18
2.1.2 Proyecciones	19
2.1.3 Escala	22
2.2 Georeferenciación de colectas biológicas.....	23
2.3 Antecedentes del desarrollo de gaceteros	25
2.4 Bibliotecas digitales	27
3. Análisis y Diseño del Sistema.....	29
3.1 Requerimientos del sistema	30
3.2 Análisis de la información	33

3.3	Diseño del sistema.....	41
3.3.1	Modelo Entidad-Relación del gacetero digital	42
3.3.2	Diseño de la base de datos.....	44
3.3.3	Diseño de la Interfaz de acceso de datos.....	46
4.	Implementación del Gacetero.....	53
4.1	Catálogos del gacetero.....	53
4.2	Implementación del gacetero con datos de localidades	56
4.2.1	Importación de datos.....	56
4.3	Consultas y reportes principales	58
4.4	Exportación de datos	61
4.5	Utilización adecuada de índices.....	62
5.	Resultados.....	65
5.1	Metodologías resultantes.....	66
5.2	Directrices	66
5.3	Programación.....	67
6.	Conclusiones y Recomendaciones.....	72
6.1	Conclusiones.....	72
6.2	Recomendaciones	73
	Bibliografía.....	75
	Glosario.....	78
	Anexo A	81
	Anexo B	83

Resumen

Los institutos de Biología, Ecología y Ciencias del Mar, entre otros, todos ellos pertenecientes a la UNAM, cuentan con acervos de información biológica que en muchos casos no está debidamente localizada geográficamente. La utilidad de esta información se ve reducida al no contar con las coordenadas del sitio de localización exacta de las colectas. El proyecto UNIGEO (Unidad de Informática Geoespacial) desarrollado por la UNAM, tiene como objetivo llevar a cabo la organización y recopilación de la información en una gran base de datos y poner la información disponible para los investigadores que así lo requieran, a través de la WWW.

En este trabajo de tesis se describe el diseño de un sistema que sirve para gestionar datos de localidades que permita en un futuro georeferenciar colectas biológicas, a través de un gacetero digital cuyo contenido son los registros históricos de localidades, adaptado para su utilización a través de medios computacionales. Este sistema de información cuenta con datos suficientes que sirven de base para poder georeferenciar las colectas biológicas para todos los proyectos futuros. Otra de sus funciones principales, es que a partir de esta fecha las colectas vayan acompañadas de la información necesaria para ser incluida en el sistema UNIGEO.

Su funcionamiento está justificado en la utilización de una base de datos que cuenta con los campos necesarios, una vez que se hizo un análisis del tipo de información más apropiada para esta aplicación, logrando además una estandarización o normalización, con respecto a otros sistemas similares desarrollados en el mundo, como lo es la ADL (Alexandria Digital Library).

Prefacio

En el presente trabajo de tesis se presentan los conceptos fundamentales que servirán de base a la propuesta de un gacetero digital que cumple con los estándares de diseño de sistemas como ADL (Alexandria Digital Library) en quien descansa la formalidad del diseño de la mayoría de los sistemas de cómputo actuales, dedicados al tratamiento de datos con fines de georeferenciación.

A lo largo del trabajo se describen algunos ejemplos de sistemas similares que existen en otros lugares del mundo y cuyos objetivos difieren un poco unos de otros. Este tipo de software está orientado al tratamiento de datos con el fin de que éstos sean utilizados a través de arquitecturas Cliente-Servidor, con miras a facilitar el acceso a colectas biológicas en cuyo caso, se pretende que estén correctamente georeferenciadas.

Esta tesis toma como marco de referencia una parte importante del proyecto ADL¹ quien desarrolla una propuesta para estandarizar el contenido de los gaceteros digitales, de tal manera que provea una referencia de trabajo flexible para almacenar la descripción de toda aquella ubicación geográfica que posea un nombre. Esta decisión no es casual, es el resultado de amplias discusiones con los propios desarrolladores del proyecto ADL y de la conclusión razonada de que una buena parte de este sistema es lo más adecuado para estandarizar la información de nuestros gaceteros digitales, por supuesto, tomando en cuenta las variantes propias de nuestra división política. Esta descripción considera para cada ubicación, diversas representaciones espaciales (punto, línea, polígono), una o más clasificaciones que tipifiquen a la ubicación acorde con algún esquema predefinido y la fuente de la cual se obtienen los datos.

Otra variante de propuesta de gaceteros digitales que se tomó como referencia para desarrollar este proyecto se llama Gazetteer Content Standard (GCS)² que contempla como prioridad documentar, tanto información actual, como aquella de carácter histórico. El GCS propone un conjunto básico de elementos para cualquier registro dentro de un gacetero, los cuales deben contar con un nombre, un tipo y una ubicación; también se sugiere documentar información temporal, relaciones entre registros, información descriptiva diversa y la fuente de la información.

En ambos casos, ADL y GCS fueron la base fundamental para la generación de un nuevo software que cumple con estándares de calidad, que es utilizado en México con todas las características de definición de lugares, tales como, Estados, Municipios, Localidades y georeferencias particulares de este país.

¹ <http://www.dlib.org/dlib/january99/hill/01hill.html> recuperado el 7/agosto/2009

² <http://www.alexandria.ucsb.edu/gazetteer/ContentStandard/version3.2/GCS3.2-guide.htm> recuperado el 7/agosto/2009

Objetivo general

Crear un sistema de información enmarcado en la arquitectura Cliente/Servidor, que sirva como solución de Bases de Datos para enfrentar la manipulación de la información de colectas biológicas en la base de datos Postgress.

Objetivos específicos

1. Diseñar un sistema de cómputo para coadyuvar con el proceso de recuperación de datos históricos necesarios para gestionar localidades para posteriormente, y con base en otras herramientas, poder georeferenciar colectas biológicas históricas.
2. Diseñar un gacetero que sea un diccionario de rasgos geográficos indexado por nombre, que utilice componentes espaciales (polígono, punto, línea) y organizado jerárquicamente.
3. Diseñar un sistema para almacenar datos generados por terceros, en este caso INEGI, y preservar el mayor número de información disponible.

Metas

El desarrollo de software profesional está basado en técnicas definidas en la teoría de Ingeniería de software, sin embargo, la meta que se persigue en este trabajo de tesis, es diseñar un sistema de información, en este caso, un gacetero, cuyo mantenimiento y desarrollo futuro quede a cargo del personal de sistemas del Instituto de Geografía de la UNAM.

Utilizando la metodología tradicional de diseño de sistemas, hacer el modelo de bases de datos compatible con estándares internacionales y diseñar la base de datos por completo.

Alcances

El Instituto de Geografía de la UNAM, actualmente tiene a su cargo el proyecto UNIGEO que tiene como propósito poner a disposición de todos los investigadores que lo necesiten, la información de colectas biológicas en forma de un gacetero digital, con el que no se cuenta en la actualidad y que es necesario desarrollar y que además sea compatible con estructuras de datos con gaceteros que están en funcionamiento en otras partes del mundo.

Utilizar los conocimientos prácticos de Ingeniería de Software para llevar a cabo un nuevo gacetero digital. Han sido varios los intentos por crear software de este tipo y en este momento se presenta una buena oportunidad de desarrollarlo.

Lograr un sistema en el que se hayan hecho propuestas adecuadas del modelo Entidad-Relación, de atributos y relaciones entre las tablas para que sea una base de datos normalizada, sin inconsistencias, sin redundancia y con una interfaz de usuario eficiente, tanto en la captura de datos, como en la explotación de los mismos.

Metodología

La metodología utilizada para el desarrollo de esta tesis fue la siguiente:

1. Se revisó el estado del arte en cuanto a desarrollo de gaceteros digitales.
2. Se seleccionaron y adecuaron los sistemas ADL y GCS para su implementación con las particularidades de nuestro país.
3. Se le dio un mayor peso a la ADL (Alexandria Digital Library) para tomarse de base ya que ésta es un estándar internacional y se propuso una adaptación en el modelo Entidad-Relación
4. De acuerdo a la Ingeniería de Software, se propuso un metodología de desarrollo de sistemas tradicional (problema, requerimientos, análisis de requerimientos, análisis de la información, diseño, desarrollo, implementación y pruebas)
5. Particularmente en la fase de diseño se realizaron las actividades siguientes:
 - Análisis detallado de los datos fuente
 - Determinación de las variables que serán necesarias al momento de desplegar y/o almacenar información.
 - Selección de equipo de cómputo y Base de Datos.
6. En la fase de desarrollo se realizaron éstas otras actividades:
 - Determinación del lenguaje de programación a utilizar, en este caso se utilizó Java
 - Determinación de la base de datos en el servidor web, en este caso se utilizó Postgress de libre distribución

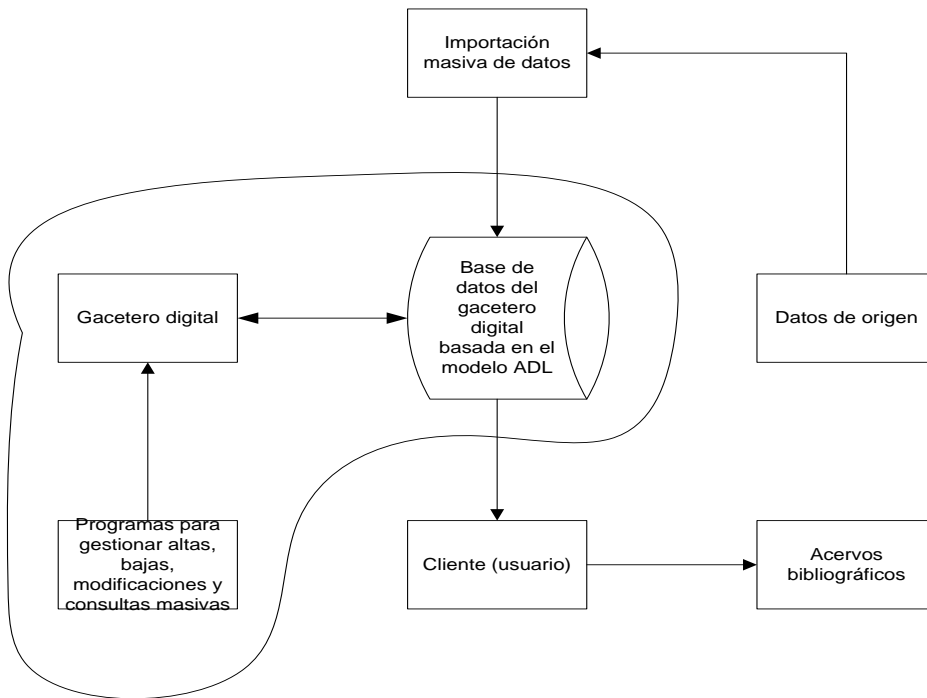
- Desarrollo de las rutinas de programación (despliegue, almacenamiento y consulta de datos), tanto en el software de captura de datos, como en la interfaz de usuario basada en la web
- Pruebas de funcionamiento con datos reales y evaluación de resultados.

Hipótesis

Es posible desarrollar un nuevo gacetero digital compatible con estándares internacionales que incluya los datos específicos necesarios para México.

Mapa

Concretamente, el gacetero diseñado incluye la base de datos y mecanismos de manejo de información tales como: entrada de datos de forma manual y masiva, y de extracción de datos de usuario, así como de exportación masiva de los mismos.



Introducción

En este capítulo se presentan los conceptos fundamentales que servirán a lo largo de la tesis, como son los de georeferenciación, colecta biológica y de gaceteros digitales. Se describe la propuesta de un gacetero que cumple con los estándares de diseño de sistemas como ADL (Alexandria Digital Library) en quien descansa la formalidad a seguir en los nuevos diseños y modelos de gaceteros actuales.

Los institutos de Biología, Química, Ciencias del Mar y Limnología, entre otras dependencias de la UNAM, cuentan con acervos y datos de colectas biológicas que son relevantes en el espacio geográfico pero que, en muchos casos, no están correctamente georeferenciados. La utilidad de estas colecciones se reduce cuando sus ejemplares no poseen una ubicación puntual sobre la tierra, lo cual se traduce como la ausencia de las coordenadas del lugar donde se realizó la colecta.

El trabajo que aquí se presenta, está enmarcado dentro del proyecto SIBA (Sistema de Informática para la Biodiversidad y el ambiente) que tiene como objetivo el crear un sistema modular de información avanzada para ordenar, sistematizar, estandarizar y analizar la vasta información primaria producida en la UNAM referente a la biodiversidad; estableciendo un vínculo con los acervos universitarios de datos espaciales y ambientales, generando una plataforma que permita producir conocimiento científico multidisciplinario sobre aspectos de la biodiversidad y desarrollar líneas de investigación que involucren diferentes escalas espaciales y temporales de nuestros recursos naturales.

1.1 Colectas biológicas

Para el presente caso de estudio se hablará del concepto de colectas biológicas en su definición más simple: “Actividad que consiste en la captura, remoción o extracción temporal o definitiva de material biológico del medio silvestre, con propósitos no comerciales, para la obtención de información científica básica, integración de inventarios o para incrementar los acervos de las colecciones científicas”³ y cuya actividad está fuertemente apoyada por las Tecnologías de información en el proceso de automatización de las principales tareas que conllevan éstas colectas. En la figura 1.1 se puede observar el tipo de datos que suelen ser almacenados respecto a las colectas biológicas, normalmente estos datos están ligados a las localidades donde se hayan, por lo que la información de localidades se vuelve imprescindible en el proceso de georeferenciación.

³ Norma Oficial Mexicana Nom.126 Ecol-2000 Gaceta ecológica No. 58 Instituto Nacional de Ecología Distrito Federal, México pp-54-60

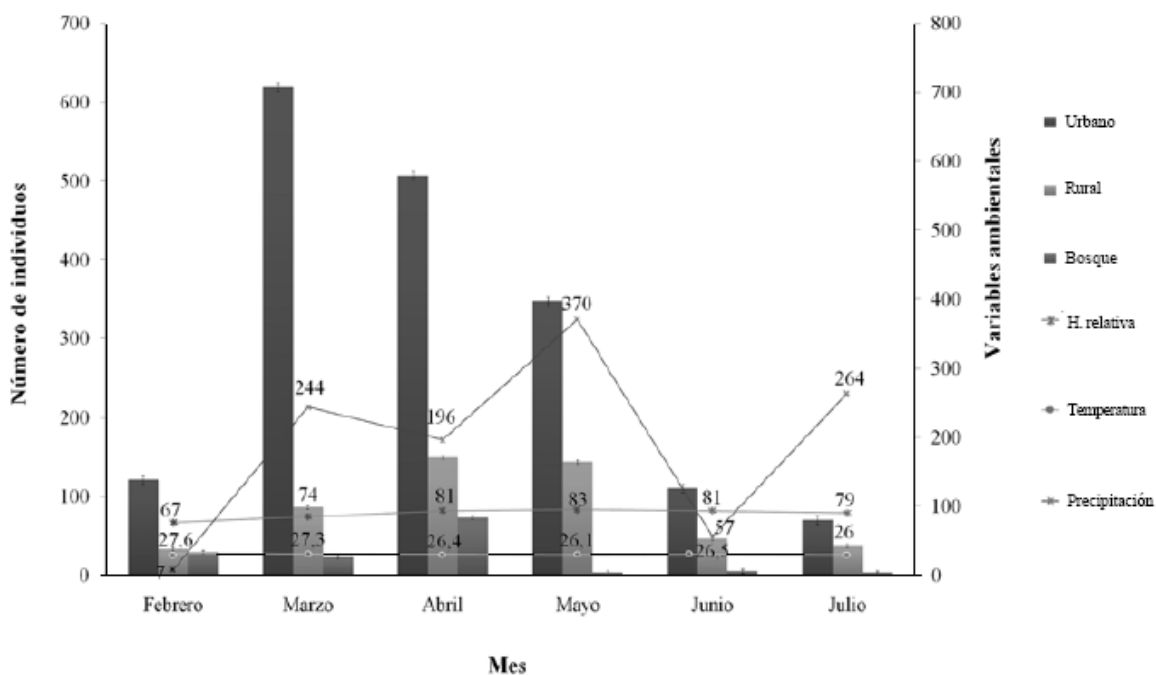


Figura 1.1 Abundancia y distribución mensual de los Califóridos entre Febrero y Julio de 2007
(Fuente: <http://www.scielo.org.co/img/revistas/rcen/v35n1/v35n1a14fig2.gif>)

La generación de grandes cantidades de datos hace necesaria la intervención de sistemas de cómputo que permitan acceder a estos mismos datos en forma eficiente y por tanto, a un estudio detallado de la generación de procesos que involucren colectas biológicas.

1.2 Georeferenciación de Localidades

La georeferenciación de localidades consiste en la asignación de coordenadas geográficas a la descripción textual de un sitio. Este permite su localización inequívoca para futuras visitas al lugar. El departamento de Análisis Geoespacial de CONABIO (Comisión Nacional para la Biodiversidad) adoptó la metodología establecida por Wiecezorek, la cual se desarrolló dentro del proyecto 'Mammal Networked Information System' (MaNIS) realizado por el museo de vertebrados de la Universidad de California en Berkeley. Esta metodología fue revisada de manera detallada por los responsables del proyecto SIBA con el propósito de adecuar los procedimientos, el material cartográfico y la nomenclatura a una serie de procesos sistematizados, los cuales se describen en el documento: Georeferenciación de localidades de colecciones biológicas: manual de procedimientos.⁴

⁴ <http://www.biodiversidad.gob.mx/region/georeferenciacion.html> recuperado el 5 de octubre de 2009

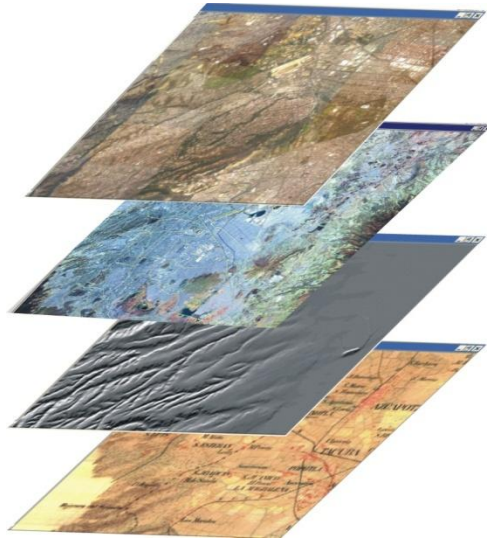


Figura 1.2. Representación gráfica de un Sistema de Información Geográfica.
(Fuente: Laboratorio de SIG y PR, Instituto de Geografía, UNAM)

La georeferenciación es el posicionamiento en el que se define la localización de un objeto espacial (representado mediante punto, vector, área, volumen) en un sistema de coordenadas y datum determinado. Este proceso de ubicación en el terreno es fundamental en los Sistemas de Información Geográfica (SIG), donde es posible encimar capas de la misma porción del territorio, pero con diferentes características y que cuenta con la capacidad de manejar bases de datos directamente asociadas. Este concepto se describe de manera sencilla en la figura 1.2

La georeferenciación posee una definición tecno-científica, aplicada a la existencia de las cosas en un espacio físico, mediante el establecimiento de relaciones entre las imágenes de *raster* o vector sobre una proyección geográfica o sistema de coordenadas. Por ello, la georeferenciación es indispensable en la utilización, el modelado y la realización de análisis espacial en un SIG. Para mayor detalle de algunos aspectos importantes de la georeferenciación, que hay que tener en cuenta durante las colectas biológicas, véase el anexo A del presente trabajo.

1.3 Gaceteros digitales y proyectos relacionados

Actualmente existe mucho trabajo desarrollado en torno a la creación de herramientas digitales con el fin de facilitar el acceso a los datos georeferenciados. Son muchos los tipos y tamaños de sistemas de cómputo basados en los datos extraídos de diferentes fuentes y sin embargo, son pocos los resultados que cumplen con ciertas normas de calidad o de protocolos de diseño. Lo importante es que si hay Instituciones serias que intentan crear herramientas que permitan el acceso a los datos de colectas biológicas. Muchas de estas actividades toman mayor importancia cuando quienes están al frente de ellas han sido reconocidas por hacer propuestas de carácter más serio y más profundo. De las actividades, se describen algunas más importantes en esta carrera por hacer mejores sistemas computarizados o automatizados en el almacenamiento y recuperación de datos de colectas biológicas en los llamados gaceteros digitales, a los cuales y a lo

largo del presente trabajo, se estará haciendo referencia cuando se trate de explicar cómo es que se manipulan los datos georeferenciados.

En la figura 1.3 se muestra una ventana de un gacetero digital, en este caso el usuario tiene diferentes opciones para la extracción de información de localidades.

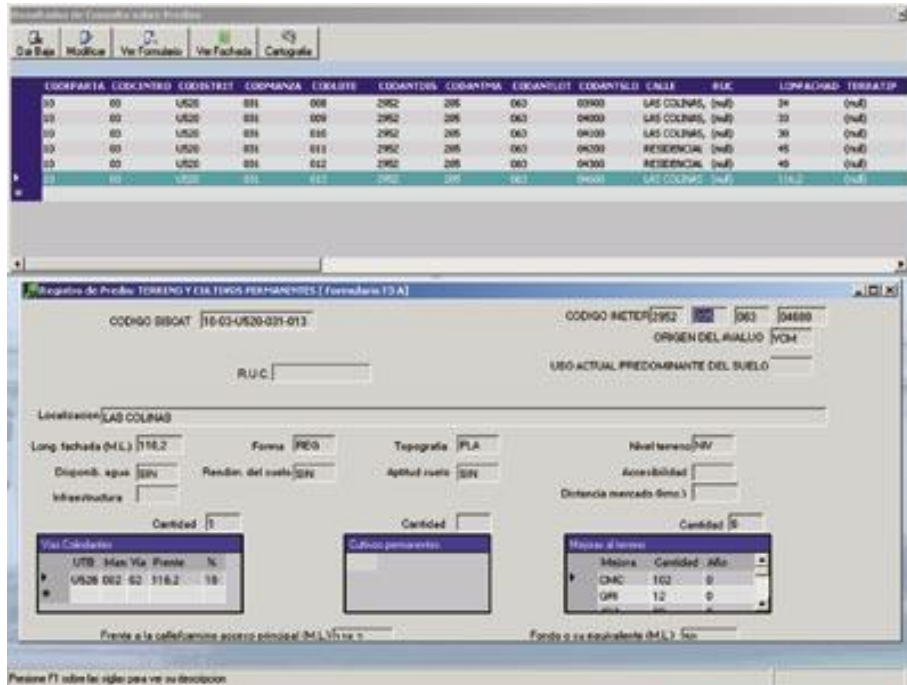


Figura 1.3 Interfaz de usuario de un gacetero digital

(Fuente: <http://www.mappinginteractivo.com/imagenes/art-03/e-octubre03/art2/fig-9.jpg> recuperado el 12 de junio de 2010)

1.3.1. INEGI (Instituto Nacional de Estadística, Geografía e Informática)

El Sistema de Consulta del Registro de Nombres Geográficos⁵ es un instrumento diseñado para la consulta de los nombres geográficos de México, mediante el cual se ofrece a los diferentes sectores de la sociedad, información toponímica fidedigna y confiable. A través de este instrumento se tiene la posibilidad de proporcionar al usuario los nombres geográficos usuales, los normalizados, los oficializados; su ubicación administrativa y geográfica, así como su referencia a la cartografía topográfica que en sus diferentes escalas produce el INEGI.

La figura 1.4 muestra una ventana en la cual los usuarios pueden realizar búsquedas. En estas consultas se muestran los rasgos que corresponden al territorio insular, el cual comprende todos o alguno de los siguientes términos genéricos: isla marítima, isla interior, cayo, roca e islote.

⁵ <http://mapserver.inegi.gob.mx/rnng/> recuperado el 30/julio/2009

El resultado de la consulta se puede filtrar para un estado en particular. De este nivel se pasa a la consulta detallada del nombre geográfico.

Archivo Histórico de Localidades
Consulta de Localidades

Estado: DISTRITO FEDERAL

Municipio: 09004 Cuajimalpa de Morelos

Criterios de Búsqueda:

- Mostrar todas las localidades
- Buscar por nombre
- Buscar por clave geoestadística

Centros de consulta y comercialización | Biblioteca digital

Condiciones de uso | Contacto | Buscador

Derechos reservados © 2010 INEGI

Para visualizar mejor este sitio se recomienda el uso de un navegador versión 4 o posterior, con resolución de 800 x 600 píxeles.

Figura .1.4 Ejemplo de una consulta en el Registro Nacional de Nombres
(Fuente: <http://www.inegi.gob.mx> recuperado el 12 de junio de 2010)

Además de la información básica del nombre geográfico, el sistema tiene la facilidad de ofrecer al usuario una vista rápida del rasgo correspondiente, esta manera de presentar la información permite referenciar el dato del elemento geográfico y su entorno. Esta vista se construye mediante el uso continuo de ortofotos (imágenes que tienen la calidad de un mapa), así como con la capa nacional de topónimos, representados por clase, para darle realce al nombre geográfico consultado.

El sistema mantiene a su vez interacción con el "Mapa Digital de México" a través del nombre geográfico consultado, de forma tal, que el usuario puede visualizarlo y tiene la posibilidad de asociarle otra capa o capas de datos fundamentales y básicos.

INEGI (registro histórico de localidades⁶)

Este archivo forma parte del proyecto del Registro Nacional de Nombres Geográficos, que tiene por objetivo integrar todas las localidades de la República Mexicana con su respectiva evolución histórica en los ámbitos administrativo, estadístico y geográfico.

⁶ http://mapserver.inegi.gob.mx/dsist/ahl2003/AHL_proyecto.cfm?c=424 recuperado el 7/agosto/2009

De acuerdo a la Ley de Información Estadística y Geográfica, el INEGI tiene la responsabilidad de difundir a todos los usuarios esta información, lo cual también se podrá hacer mediante la consulta de nuestra base de datos de manera fácil y ágil, que es de gran importancia para estudios e investigaciones sobre el desarrollo de las localidades a través del tiempo. La base de datos fue integrada por la captura de más de trescientas sesenta mil tarjetas, las cuales en versión manuscrita contenían información y datos censales, éstas fueron depuradas, validadas, actualizadas y confrontadas con documentos oficiales emitidos por el gobierno de la federación, los gobiernos estatales y las publicaciones de resultados definitivos de los diferentes eventos censales.

La figura 1.5 muestra el resultado de una búsqueda de localidades por criterios establecidos por el INEGI.

Clave Geoestadística	Entidad	Municipio	Localidad
090040001	DISTRITO FEDERAL	Cuajimalpa	Cuajimalpa (Cab.) Cuajimalpa de Morelos (Cab.)
090040004	DISTRITO FEDERAL	Cuajimalpa	Azoyapan
090040005	DISTRITO FEDERAL	Cuajimalpa	Buenavista
090040006	DISTRITO FEDERAL	Cuajimalpa	Buenavista
090040007	DISTRITO FEDERAL	Cuajimalpa de Morelos	La Cañada
090040008	DISTRITO FEDERAL	Cuajimalpa	La Cañada
090040009	DISTRITO FEDERAL	Cuajimalpa	Cola de Pato
090040010	DISTRITO FEDERAL	Cuajimalpa de Morelos	Cola de Pato
090040011	DISTRITO FEDERAL	Cuajimalpa	Contadero
090040012	DISTRITO FEDERAL	Cuajimalpa	Contadero
090040013	DISTRITO FEDERAL	Cuajimalpa	El Contadero
090040014	DISTRITO FEDERAL	Cuajimalpa de Morelos	Las Cruces
090040015	DISTRITO FEDERAL	Cuajimalpa	Las Cruces
			Cruz Blanca
			Cruz Blanca
			Cruz Manca
			Cruz Manca
			Chimalpa
			Chimalpa
			Hueyatla
			Hueyatla
			Jesús del Monte
			Jesús del Monte
			Locaxco
			Locaxco

Figura 1.5 Consulta de localidades del INEGI
(Fuente: <http://mapserver.inegi.org.mx/AHL/realizaBusqueda.do> recuperado el 12 de junio de 2010)

En esta misma figura 1.5 se muestran las localidades correspondientes al estado y municipio seleccionado. De igual forma que en los niveles estatal y municipal, se presenta el nombre de uso común y nombre oficial de cada localidad. Cada registro incluye una liga para ver el detalle de la localidad.

Por otra parte INEGI posee un registro del conjunto de cambios en los nombres de las localidades a través del tiempo desde 1900.

1.3.2 GeoLocate⁷

Las colecciones de los museos de historia natural han servido tradicionalmente como un recurso para la investigación sistemática sobre la taxonomía de los organismos. Cada vez más, las muestras y sus datos están siendo utilizados en los estudios de la ecología, la conservación y la biología del medio ambiente. Un obstáculo importante en la aplicación de dichos estudios, con estos especímenes del museo, es la falta de referencias geográficas del lugar de la colecta (coordenadas en el mapa, como latitud y longitud). Los métodos tradicionales de captura de datos de georeferenciación en las descripciones de texto son engorrosos y poco claros, caracterizados por la búsqueda de la localidad, ya sea en mapas impresos o digitales, trazándola y determinando sus coordenadas. GEOLocate es una completa solución de georeferenciación electrónica, financiada por la Fundación Nacional de Ciencia (EEUU) y desarrollada por la Universidad de Tulane (Museo de Historia Natural). Esta herramienta fue diseñada para facilitar la tarea de asignar coordenadas geográficas a los datos de localización en las colecciones de historia natural. Los principales objetivos del desarrollo de GEOLocate fueron:

- Desarrollar un algoritmo para convertir información de texto de historia natural, en datos de la latitud y longitud de América del Norte.
- Proporcionar una interfaz para la visualización y ajuste al generar las coordenadas.
- Dar una solución simple para los usuarios al importar y georeferenciar sus datos.
- Incorporar la característica de auto-actualización.
- Apoyar el proyecto GEOLocate a través del sitio web, correo electrónico y teléfono.

La figura 1.6 muestra las características anteriores en forma visual a través de una interfaz genérica desarrollada para estos fines.

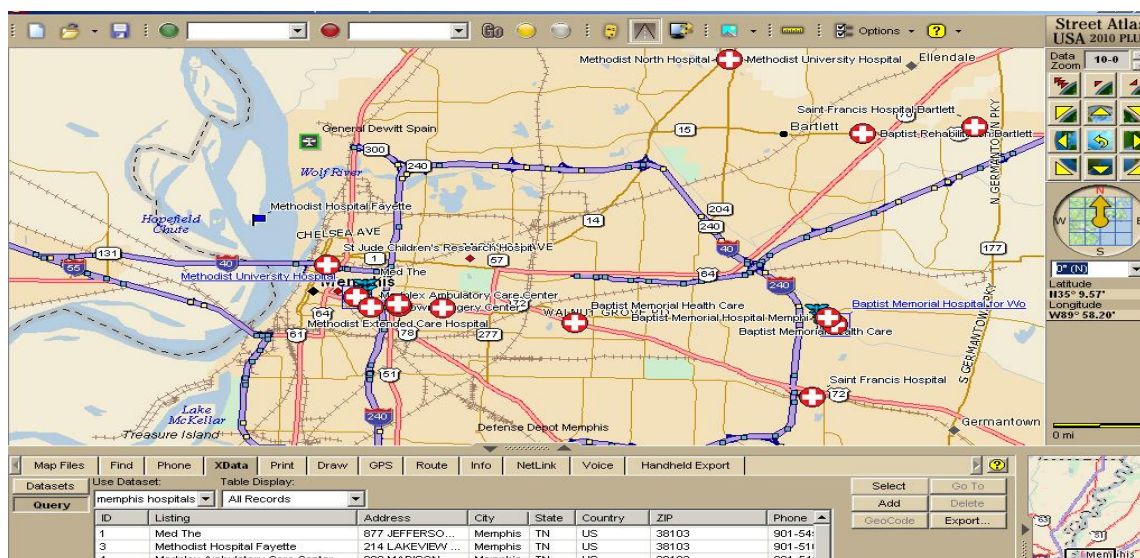


Figura 1.6 Interfaz para la visualización y ajuste al generar las coordenadas
(Fuente: http://www.software-maps.com/assets/DeLorme/SA10+_GeoLocateContacts.jpg recuperado el 12 de junio de 2010)

⁷ <http://www.museum.tulane.edu/geolocate/default.aspx> recuperado el 2 de octubre de 2009

Los resultados de este proyecto fueron presentados en el 2003 las colecciones de Ciencias Naturales Reunión de la Alianza, la Sociedad Americana para 2003 ictiólogos y herpetólogos de la reunión, georeferenciación talleres sobre georeferencing en la Universidad de Kansas y la Universidad de Yale y en la reunión de 2004 TDWG. Desde su lanzamiento, más de 800 copias de GEOLocate han sido distribuidas a varios museos, colecciones y a diferentes investigadores.

1.3.3. BioGeomancer⁸

El proyecto Biogeomancer es una colaboración internacional entre expertos en el manejo de historia natural y de especialistas en datos Geoespaciales. El consorcio Biogemancer está desarrollando una serie de herramientas en línea, servicios web y aplicaciones de gabinete que coadyuvarán a la georeferenciación de colectas a los curadores y usuarios de especímenes de historia natural. Estas herramientas se apoyan en programas de computadora para el procesamiento del lenguaje natural de registros históricos que provienen de diversas fuentes.

Para el caso de los trabajos realizados por Biogeomancer, se puede apreciar en la figura 1.7 un avista de una de sus interfaces.



Figura 1.7 Interfaz de Biogeomancer

(Fuente: http://pondpond.blogspot.com/2009_03_01_archive.html recuperado el 12 de junio de 2010)

⁸ <http://74.6.146.127/search/cache?ei=UTF->

[8&p=Guide+to+Best+Practices+for+Georeferencing&rd=r1&fr=yfp&u=arobis.cenpat.gov.ar/recursos-tecnicos/georeferenciacion/material-taller/BestPractices.pdf&w=guide+best+practices+georeferencing&d=EDw9OBIMTG6O&icp=1&.intl=mx](http://74.6.146.127/search/cache?ei=UTF-8&p=Guide+to+Best+Practices+for+Georeferencing&rd=r1&fr=yfp&u=arobis.cenpat.gov.ar/recursos-tecnicos/georeferenciacion/material-taller/BestPractices.pdf&w=guide+best+practices+georeferencing&d=EDw9OBIMTG6O&icp=1&.intl=mx)

recuperado el 7/agosto/2009

El consorcio de Biogemancer⁹ está coordinado por la Universidad de los Ángeles, Berkeley, donde se desarrolla un sistema universal de georeferenciación de este tipo de datos, el cual permite reducir el costo involucrado en la georeferenciación de colectas.¹⁰

1.3.4 UNIGEO (Unidad de Informática Geoespacial)

El proyecto UNIGEO (Unidad de Informática Geoespacial) perteneciente a SIBA está actualmente en desarrollo en el Instituto de Geografía, UNAM, y tiene dentro de sus objetivos el organizar y recopilar la información geoespacial disponible en el instituto. Esta información se procesa, cataloga y almacena dentro de repositorios digitales y bases de datos geoespaciales, con el objetivo de su posterior aprovechamiento por los investigadores que así lo requieran a través de medios electrónicos.

Los objetivos particulares del proyecto¹¹ UNIGEO son los siguientes:

- Habilitar el acceso a la información espacial desde cualquier sistema de cómputo con accesos a Internet.
- Permitir mecanismos de distribución de información.
- Integrar diversas fuentes de información espacial.
- Posibilitar el encadenamiento de aplicaciones, datos y servicios.
- Habilitar la interoperabilidad al adoptar estándares y especificaciones abiertas.
- Crear oportunidades de participación entre instituciones educativas, dentro y fuera de la UNAM.
- Plantear la posibilidad de integración de sectores públicos y privados a nivel nacional e internacional.
- Contar con un diseño distribuido que produzca un sistema robusto que genere redundancia contra fallas.
- Proveer mecanismos de seguridad y control de acceso.

Dentro de las recomendaciones que recibe UNIGEO por parte de una comisión internacional que se encarga de supervisar su desarrollo, en la realización de sus actividades para el año 2010, se contempla lo siguiente:

“UNIGEO debe proveer herramientas, procesos y servicios para una georeferenciación automatizada de primer paso utilizando las mejores prácticas existentes, incluyendo herramientas para extracción de rasgos particulares de conjuntos de descripciones textuales y creación visual de rasgos para aumentar los gaceteros actuales. Esto no debería ser visto como una apropiación de la tarea de georeferenciación, sino más bien como servicios centrales para hacer más eficiente el componente puramente geográfico de esta complicada labor.

⁹ (<http://www.biogemancer.org/about.html>) recuperado el 7/agosto/2009

¹⁰ <http://www.alexandria.ucsb.edu/gazetteer/> recuperado el 7/agosto/2009

¹¹ <http://www.unigeo.igeograf.unam.mx/unigeo/index.php/UNIGEO.html> recuperado el 6/agosto/2009

Adicionalmente, UNIGEO debería desarrollar herramientas que ayuden a las unidades en la segunda etapa de georeferenciación, en la cual los resultados automatizados pueden ser seleccionados, examinados, editados y descargados para su incorporación en bases de datos primarias.¹²

El presente trabajo atiende parte de las recomendaciones establecidas para el proyecto UNIGEO, al colaborar con el desarrollo de herramientas para la georeferenciación. En este trabajo se describe el diseño estructural de un repositorio de datos, necesario para la consolidación de un gacetero digital con información histórica y pertinente de México. El repositorio está conceptualizado para almacenar diversos rasgos geográficos, poniendo especial atención en la validez de sus registros a través del tiempo. El conjunto de datos que se considera para el desarrollo de la propuesta se centra en información disponible de localidades, así como la serie de cambios en los nombres geográficos a través de diferentes eventos históricos.

1.3.5 ADL (Alexandria Digital Library)

El proyecto ADL¹³ desarrolla una propuesta para estandarizar el contenido de los gaceteros digitales, de tal manera que provea un marco de trabajo flexible para almacenar la descripción de toda aquella ubicación geográfica que posea un nombre. Esta descripción considera para cada ubicación, diversas representaciones espaciales (punto, línea, polígono), una o más clasificaciones que tipifiquen a la ubicación acorde con algún esquema predefinido y la fuente de la cual se obtienen los datos. La figura 1.8 muestra la interfaz típica de los trabajos de ADL.

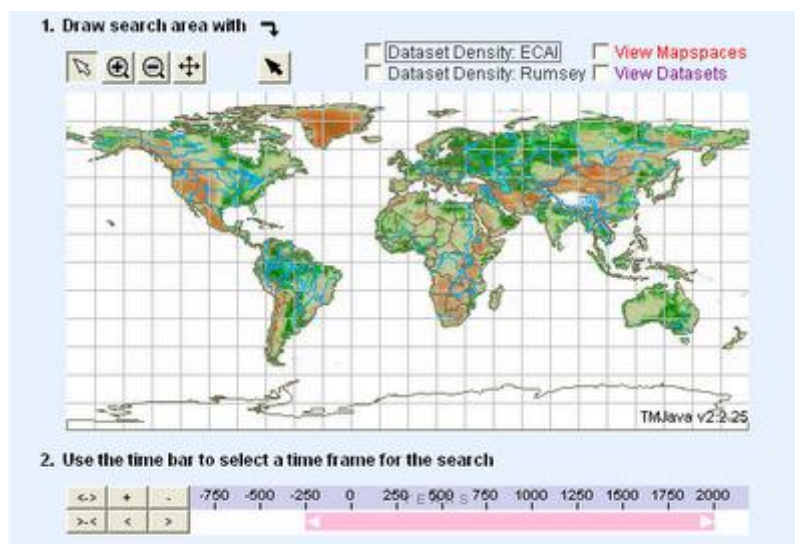


Figura 1.8 Interfaz de ADL
(<http://2.bp.blogspot.com/> recuperado el 12 de junio de 2010)

¹² Evaluación 2009.pdf (Informe de la evaluación anual de SIBA durante el periodo 2008 - 2009, realizada del 13 al 16 de abril.); http://siba.ibiologia.unam.mx/index.php?option=com_remository&Itemid=2&func=startdown&id=42 recuperado el 29/julio/2009

¹³ <http://www.dlib.org/dlib/january99/hill/01hill.html> recuperado el 7/agosto/2009

Esta propuesta se llama Gazetteer Content Standard (GCS)¹⁴ y contempla como prioridad documentar, tanto información actual, como aquella de carácter histórico. El GCS propone un conjunto básico de elementos para cualquier registro dentro de un gacetero, los cuales deben contar con un nombre, un tipo y una ubicación; de manera recomendada se sugiere documentar información temporal, relaciones entre registros, información descriptiva diversa y la fuente de la información.

El proyecto ADL integra información procedente del proyecto GNPS (Geographic Names Processing System) para todas aquellas ubicaciones fuera de los Estados Unidos de América junto con información del GNIS (Geographic Names Information System), la cual es la información de nombres dentro del país mencionado. En total se integran 5.9 millones de registros.

1.4 Estructura de la información dentro de un gacetero digital

Los gaceteros son una herramienta básica para toda actividad de georeferenciación de colectas y su versión digital es de suma utilidad debido a la posibilidad que ofrecen para sobreponer esta información con otros elementos cartográficos, auxiliares a la georeferenciación. La generación y adecuación de datos espaciales para su uso dentro de gaceteros digitales es una actividad laboriosa y propensa a errores, y la experiencia de trabajos previos facilita la creación de los mismos y evita incurrir en ellos.

El conjunto de datos que conforma el diccionario geográfico tiene mucha importancia para la creación de los gaceteros digitales, pero la generación de herramientas necesarias para acceder y aprovechar estos datos, es tan importante como los datos mismos del gacetero, ya que estas serán las que permitan acceder y explotar el contenido del acervo. En este trabajo se resume la propuesta de creación de un gacetero de localidades y topónimos, usando como base la información puntual proveniente del INEGI para generar datos relacionados a estas localidades, sus datos de georeferencia y sus características más destacadas con posibilidades de poderse organizar con base a los criterios de las personas que usen el gacetero digital.

Partiendo de que un gacetero se define como una colección o base de datos de ubicaciones geográficas, la cual sirve para traducir entre distintas representaciones de referencias geoespaciales, como son nombres de lugar y coordenadas geográficas; se denota que una buena estructura de orden dentro de un gacetero debe poder proveer los datos eficientemente. El gacetero también puede verse como una lista de nombres geográficos, junto con su ubicación espacial e información descriptiva anexa.

Es posible clasificar diversos gaceteros a partir de la información que contienen y las relaciones entre sus registros:

1. El tipo más simple o básico solamente contiene el nombre y la localización del rasgo (coordenada). Comúnmente contiene información descriptiva del rasgo.

¹⁴ <http://www.alexandria.ucsb.edu/gazetteer/ContentStandard/version3.2/GCS3.2-guide.htm> recuperado el 7/agosto/2009

2. La siguiente clasificación agrega un atributo a cada entrada (dato) dentro del gacetero. Esta entrada es el tipo rasgo, el cual usualmente se encuentra dentro de una jerarquía de tipos, es decir, cada entrada se compone de un nombre, su localización y un tipo jerárquico. La inclusión de un tipo sugiere la creación de vocabularios controlados para su designación.
3. El siguiente tipo de gaceteros define relaciones entre sus entradas. Estas relaciones normalmente siguen alguna estructura jerárquica definida, como las divisiones o niveles administrativos. Las entradas se conforman de un nombre, su localización, el tipo de rasgo, una o más relaciones con otros rasgos.

1.5 Acceso a la información en gaceteros digitales

El gacetero en cuestión, que se ha desarrollado en términos bien establecidos, corresponde al tercer tipo del apartado anterior. Partiendo de la definición de un gacetero como un conjunto de entradas, las cuales se forman por:

- Un nombre geográfico
- Uno o más componentes georeferenciados
- Un tipo de dato o clasificación dentro de un vocabulario controlado
- Información extra relacionada al rasgo
- Cero o más relaciones con otros rasgos y
- Una serie de esquemas para catalogar la información contenida

Es evidente que la creación de un software basado en la arquitectura cliente-servidor en la que confluyen una serie de aplicaciones que cumplen funciones diferentes (una requiere servicios y la otra los brinda), son justamente las aplicaciones para la consulta (desde un cliente) de los datos albergados en el gacetero (desde el servidor), esto permite explotar la información eficientemente, permitiendo cruzarla en el orden que el usuario lo disponga y por lo tanto, el esquema empleado para la creación del gacetero, permite responder preguntas del tipo

- ¿Existe la localidad x?
- ¿Qué información hay de x?
- ¿Dónde está x?
- ¿A qué categoría pertenece x?

La generación de un gacetero digital adecuado a la georeferenciación de colectas biológicas requiere considerar el tipo de información contenida en la ficha, la cual es una descripción detallada acerca del lugar de colecta, así como los métodos empleados al realizarla.

El gacetero puede dar acceso a la información de las colectas biológicas pero los elementos en esta descripción normalmente corresponden a la división administrativa de los países, que en el caso de México son estados, municipios y localidades.

Adicionalmente se emplean elementos presentes en las cartas topográficas, tanto a escala 1:250,000, como 1:50,000. De los elementos contenidos, aquellos con más importancia corresponden a las vías de comunicación, los rasgos hidrográficos, cuerpos de agua, toponimia, etc. Dentro de la cartografía existente en el Instituto de Geografía de la UNAM, la cual proviene en su mayoría de INEGI, se pueden distinguir las siguientes capas como información clave para la generación de gaceteros:

- División Estatal
- División Municipal
- Localidades
- Cartas topográficas
- Toponimia
- Vías de comunicación

Este conjunto de cartografía compone gran parte de los elementos del gacetero, como son:

- El nombre geográfico y
- Componentes geométricos; ya sean puntuales, de línea o de área.

1.6 Planteamiento del problema

Después de revisar algunos de los esfuerzos más importantes por mantener un orden y una metodología clara en el acceso a los datos georeferenciados de las colectas biológicas, es preciso centrarse en la generación de un software con arquitectura cliente-servidor que entregue un producto y un método, particularmente un gacetero digital que se pueda utilizar de manera inmediata para consultar datos georeferenciados.

1.6.1 Definición del problema

La descripción del sitio de colecta de un espécimen es parte esencial de la información asociada a cada colecta, pero su utilización en el análisis espacial a través de los sistemas de información geográfica sólo es factible si dentro de estos datos se cuenta con las coordenadas geográficas.

Actualmente, el uso de los sistemas de posicionamiento global (GPS) ha permitido agregar de manera más frecuente y con mayor precisión el dato de las coordenadas geográficas de un sitio de colecta; sin embargo, el volumen de registros sin coordenadas geográficas es considerablemente elevado. En términos prácticos, esto significa que una enorme cantidad de información permanece desaprovechada, de ahí la importancia de la

georeferenciación de localidades. La descripción de las localidades donde son recolectados los organismos es realizada de diversas maneras ya que involucra subjetividad, pues el colector describe un sitio de acuerdo a los medios que tiene a su disposición y a su propia percepción del lugar, lo cual, a su vez, determina la precisión de la localidad.

El proceso de georeferenciación¹⁵ tiene por requisito contar con acceso a gaceteros pertinentes a las colectas, idealmente en formato digital. Estos gaceteros requieren contar con un conjunto de registros organizado de manera congruente, con información temporal y referencia geográfica, así como detalles sobre su creación. Estas características son necesarias para dar validez a los resultados obtenidos y comprobarlos geográficamente. Cuando la información que se pretende emplear para alimentar un gacetero cumple con estos requisitos, se considera que estará lista para incorporarse dentro de dicho acervo.

Debido al limitado acceso a la información disponible en los gaceteros digitales desarrollados por los proyectos mencionados anteriormente, al incipiente contenido de tales acervos para México, así como al bajo nivel de control en su contenido, resulta necesario sentar las bases para estructurar un gacetero que sea flexible, que permita incorporar información histórica de diversas fuentes y almacenar diversas representaciones espaciales de sus registros. Sentar adecuadamente estas bases para el desarrollo del gacetero se vuelve el principal problema a resolver, ya que se debe partir de lo que ya se tiene desarrollado en sistemas de cómputo parecidos, pero con el requisito indispensable de poder cumplir con los lineamientos que se han establecido por el mismo Instituto de Geografía y que permitan almacenar los datos obtenidos de las colectas biológicas para su explotación eficiente. Pero el trabajo realizado en otros sistemas deja claro que se deben distinguir perfectamente las diferencias respecto del gacetero propuesto y los ya existentes; la manera más clara de resaltar estas diferencias es poder dar respuesta a las siguientes preguntas:

- ¿Qué bases teóricas vamos a utilizar para estructurar el contenido del gacetero propuesto?
- ¿Cómo se almacena, representa y preserva la información histórica en dicho gacetero digital?
- ¿Cómo se preservan las distintas representaciones espaciales en un rasgo geográfico?
- ¿Cómo se puede garantizar la inclusión de información proveniente de fuentes diversas dentro de la estructura del gacetero digital propuesto?

Para poder contestar las preguntas anteriores han de tomarse en cuenta principalmente los modelos propuestos por la Alexandria Digital Library (ADL) y más especialmente Gazetter Content Standard (GCS), ya que estos dos modelos abstractos, actualmente se consideran un estándar capaces de cumplir con las necesidades del Instituto de Geografía. A pesar de que proponen ciertas estructuras y normas para el almacenamiento

¹⁵ Georeferenciación de localidades de colecciones biológicas, manual de procedimientos, CONABIO, México 2008, pp. 34

y manejo de gaceteros digitales, es importante una nueva propuesta que acote las necesidades particulares a cubrir, ya que de esta manera se garantiza la originalidad del nuevo gacetero pero a su vez que pueda ser adaptable o compatible con otros diseños propuestos.

1.6.2 Justificación

El presente trabajo tiene como objetivo atender una necesidad latente del proyecto UNIGEO. La idea es desarrollar un sistema de cómputo capaz de albergar el contenido de un gacetero digital de México, ya que hasta el momento no existe nada semejante en términos de datos, tanto por la calidad histórica de los mismos, como por su extensión. Los esfuerzos internacionales en este sentido no incluyen a México de manera extensiva, por lo que es necesario que se lleve a cabo un esfuerzo para generar tal sistema. Los pocos avances en esta línea están basados en una fuerza organizacional burda, es decir, solo abarcan algunos de los requisitos planteados anteriormente, y en su mayoría a nivel municipal. La presente propuesta abarca un nivel más detallado, en este caso, a nivel localidad. Un sistema automatizado con las características propuestas forma parte del conjunto de herramientas necesarias para realizar la gestión de localidades que posteriormente sirvan para la georeferenciación de colectas biológicas. Este proyecto es de importancia institucional y nacional debido a la necesidad que existe por conservar y administrar la biodiversidad en México.

Como ya se ha mencionado, ADL y GCS, dos de los modelos ya existentes, brindan algunas ideas en torno a la creación de nuevos gaceteros, sin embargo, habrá que tomarlas en cuenta sin dejar de hacer una propuesta propia para asegurar un buen uso de un estándar pero con características adicionales que se adecuen a México, todo con el fin de crear un gacetero funcional.

Georeferenciación.

Es un proceso en el cual los objetos desprenden información cuyo contenido puede ser de dos tipos, de acuerdo al método que se haya utilizado para otorgarles esta característica. Pueden ser objetos con información geoespacial y que hacen referencia a la tierra, tales como mapas e imágenes capturadas desde el espacio. También pueden ser objetos de información georeferenciada de los topónimos, codificación geográfica, y direcciones, pero sin hacer referencia geoespacial explícita, tales como listados, documentos de texto (por ejemplo, declaraciones de impacto ambiental y artículos de revistas), las historias orales, y la catalogación de colecciones de libros, especímenes biológicos, y otros. El tipo de topónimo de georeferenciación a veces se le conoce como de georeferencia indirecta, lo que implica que se requieren pasos adicionales para identificar los lugares en un mapa. Todos los tipos de objetos de información tienen el potencial de ser georeferenciados a través de enlaces a elementos geográficos con nombre (es decir, el nombre de lugares geográficos), que a su vez puede ser vinculada a lugares sobre la tierra.

Se debe contar con cierta familiaridad de las características, aplicaciones y ventajas de los objetos geoespaciales porque son los principales componentes de los sistemas de información para las aplicaciones del usuario. Su disponibilidad ha revolucionado nuestra comprensión de la Tierra y de nuestro entorno cultural local y mundial. Este tipo de sistemas se utilizan a menudo en la gestión y la atención del personal de las mapotecas (bibliotecas de mapas), los centros de datos geoespaciales, y dependencias gubernamentales. Son creados para la producción de mapas, los programas de detección remota (es decir, los datos y las imágenes recogidas por naves espaciales y la cartografía del fondo marino) y de fotografías aéreas.

Otras fuentes de datos geoespaciales incluyen a la tierra y el mar, basados en los instrumentos que captan ondas sísmicas, las temperaturas oceánicas, el flujo de corriente, etc. Uno de los aspectos más importantes en la Georeferenciación es poder representar localidades, es decir, determinar geoespacialmente lugares en la tierra y describir todas sus características. Esto implica que se deben conocer algunos aspectos importantes sobre el planeta.

La forma de la Tierra.

El planeta Tierra es un elipsoide abultado. Si fuera una esfera perfecta, el desarrollo de métodos de geocodificación de lugares en su superficie sería menos complejo. Pero con la forma que tiene, no hay un modelo matemático simple que describa su forma y su tamaño, sino varios modelos para diferentes propósitos y adaptados a diferentes zonas

del planeta. La ciencia involucrada en la determinación del tamaño y la forma de la Tierra se conoce como geodesia.

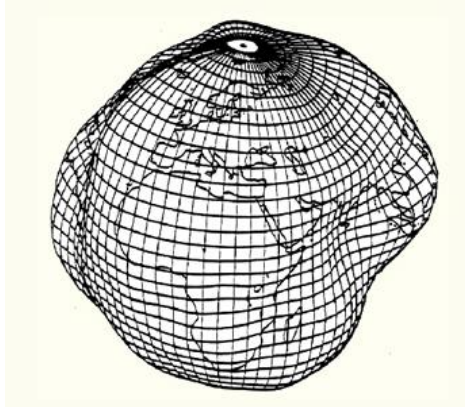


Figura 2.1 Forma de la Tierra.

(Fuente: <http://kartoweb.itc.nl/geometrics/Reference%20surfaces/body.htm>)

Existen diferentes modelos que determinan la forma de la Tierra, los más comunes son: geode, esferoide y elipsoide. Este último es una de las formas más comunes de representar a la Tierra y para describirla matemáticamente se utilizan diversas versiones; que difieren según la aplicación y la zona geográfica que se esté utilizando. Algunos ejemplos de modelos de elipsoide son: el de Clarke 1866, el GRS80 y el WGS84.

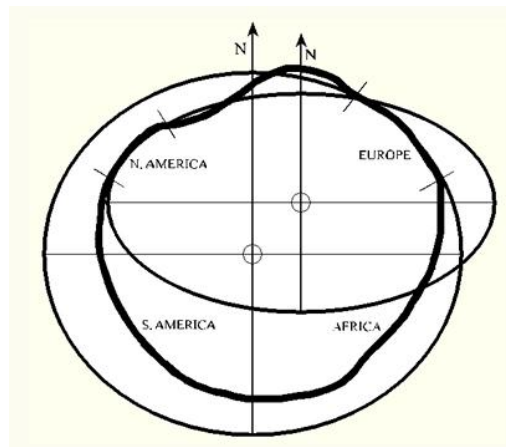


Figura 2.2. Un elipsoide se adapta muy bien al contorno de Norteamérica, sin embargo, para Europa un esferoide describe de mejor manera su forma.

(Fuente: <http://kartoweb.itc.nl/geometrics/Reference%20surfaces/body.htm>)

Los modelos que describen la forma de la Tierra no son exactos y para corregir las imprecisiones existen una serie de redes de puntos de referencia. A éstas redes se les conoce como datum y podemos mencionar, por ejemplo, el NAD27 (*North American Datum*) 1927, el NAD83 (*North American Datum* 1983), el WGS84 (*World Geodetic System* 1984) y el ITRF92 (*International Terrestrial Reference Framework* 1992). A lo largo de los años cada datum se ha venido desarrollando, corrigiendo y actualizando, en respuesta a los movimientos de la corteza terrestre, pero también en buena medida a que

se ha contado con equipos de medición más precisos. Sistemas como el HARN (*high accuracy reference network*) red de referencia de alta precisión, basada en receptores GPS, que funcionó en estados unidos entre 1987 y 1997, sirvió para mejorar la precisión del datum NAD83, pasando de 1 m a 10 cm. En la actualidad, el sistema CORS (*Continuously operating Reference Stations*) estaciones de referencia de operación continua, sustituye las funciones del HARN, logrando que la precisión de NAD83 sea del orden de 0.02 metros en horizontal y de 0.04 metros de altura con respecto al elipsoide de referencia.

Al obtener e integrar los datos de GPS en un mapa o en un SIG, es importante saber que entre las diferentes combinaciones de modelos para describir la forma de la Tierra y cada datum puede haber diferencias de varios cientos de metros. El sistema GPS utiliza el datum WGS-84 y el centro del elipsoide WGS-84 como el origen de un marco de referencia centrado en la Tierra y fijo a la Tierra (Earth-centered Earth-fixed, ECEF). Esta información a veces es pasada por alto por algunos usuarios, con las consecuencias que esto implica. Esto es extremadamente importante al llevar a cabo las colectas y es indispensable corroborar esta información al momento de llenar las bases de datos.

2.1. Esquemas de códigos geospaciales - coordenadas y Sistemas de coordenadas

Al existir diferentes modelos para representar a la tierra y diferentes redes de referencia locales, es de capital importancia saber con que sistema estamos tratando para evitar confusiones.

2.1.1 Propiedades y Elementos de la Cartografía General

Las propiedades que presentan los mapas sirven para diferenciarlos de cualquier otra representación de la superficie terrestre, como puede ser una fotografía o una pintura. De ahí que sean tres las propiedades esenciales de todo mapa geográfico:

- Ley matemática espacial de la estructura del mapa: consiste en el paso de la imagen física de la Tierra a un plano, apoyado en una superficie matemática, para lo cual se utilizan las proyecciones cartográficas que son las que establecen una relación entre las coordenadas de la superficie de la Tierra y las correspondientes con el plano. Esta ley matemática nos permite obtener datos sobre la posición y la forma de los objetos representados, así como también las dimensiones del plano.
- Método de representación cartográfica: este método se emplea en la representación de objetos y eventos que se muestran en el mapa, los signos cartográficos juegan un papel esencial, ya que en ellos se vacía la información que debido a su importancia debe ser representada tomando en cuenta la temática del mapa.
- Generalización: significa incrementar la legibilidad de un mapa disminuyendo su contenido, de tal forma que mantenga en lo posible la precisión de su información. Es la propiedad más importante del mapa, destaca los aspectos más sobresalientes de la superficie de la Tierra representada, y está relacionada con la asignación del tema y la escala.

Para poder emplear en forma correcta los mapas es necesario conocer todos los elementos que los constituyen, con el propósito de entender las correlaciones que existen entre ellos. Los elementos de los mapas son cuatro:

- **Representación cartográfica:** es la parte principal de cualquier mapa, está dividida en distintas ramas de acuerdo al objetivo, topografía, hidrología, asentamientos humanos, vegetación, vías de comunicación, así como también los aspectos de cultura y división política, etc.
- **Base matemática:** es necesaria para pasar una parte de la superficie de la Tierra a un plano tratando de que no pierdan sus verdaderas dimensiones, para lo cual se apoya en la proyección, la escala y la base geodésica.
- **Elementos auxiliares:** estos ayudan a entender de una mejor manera el mapa y son el título, tema, leyenda, información cartométrica, datos informativos, como son la fecha, edición y fuentes de consulta que se emplearon para realizar el mapa.
- **Información complementaria:** entre estos se encuentran los cuadros de referencia que se ubican en el mapa con respecto a otros que se han realizado en el país a nivel general, también se localizan los mapas adyacentes que son de ayuda en la ubicación del mapa pero a nivel local con respecto a otros, y por último las gráficas y los perfiles que aclaran, complementan y enriquecen la propia imagen cartográfica.

2.1.2 Proyecciones

Por definición, la cartografía trata sobre la elaboración de mapas e incluye todas las operaciones comprendidas desde los reconocimientos de campo hasta la impresión final. No es posible llevar directamente la superficie esférica o elipsoidal a una superficie plana sin modificar las relaciones geométricas, por lo que se ha estudiado una transformación que pueda guardar en la carta varias características que tiene la superficie esférica. Esta forma de la Tierra ha obligado a que se utilice una figura geométrica auxiliar desarrollable (plano, cono o cilindro), a la cual se le proyecten los paralelos y meridianos desde un punto de vista o centro de perspectiva, el que también puede ser variable en su situación.

La figura geométrica empleada: plano, cono o cilindro; la posición que se les aplique a la mencionada figura: normal, paralela o inclinada al ecuador y la situación del punto de vista dan origen a casi todas las proyecciones cartográficas conocidas. Los puntos se localizan por sus valores coordenados de latitud y longitud "red de paralelos y meridianos", a esta red se le denomina gradícula o canevá y a la construcción en una superficie plana se le nombra proyección cartográfica. Cada una de las proyecciones tiene sus propias características y con el conocimiento de éstas se elige la proyección a utilizarse según las necesidades y objetivos requeridos. En relación con las propiedades de forma, área, distancia y dirección se han ideado las proyecciones: conformes, equivalentes,

equidistantes y azimutales; además, también se clasifican por su construcción en: cilíndricas, cónicas, azimutales o convencionales.

En México son dos las proyecciones más usadas de acuerdo a los propósitos y objetivos que se quieren alcanzar. Son la proyección la Cónica Conforme de Lambert (CCL) y la Universal Transversa de Mercator (UTM).

Proyección Cónica Conforme de Lambert. Esta proyección es la más empleada en las cartas aeronáuticas, por su pequeña anamorfosis y sus acimuts relativamente rectilíneos para una región de varios cientos de kilómetros cuadrados. En esta proyección los paralelos concéntricos están espaciados de tal modo que cada cuadrilátero del canevá tienen las mismas proporciones que en el globo. Los meridianos radiados desde el centro de los paralelos cortan a los dos paralelos principales en partes de verdadera magnitud.

Existen varias modificaciones de esta proyección, aplicadas recientemente a necesidades aeronáuticas. También se empleó esta proyección en los mapas de la zona occidental en la primera guerra mundial, por no exceder el error de un 0.5 por 100 del área representada (Figura 2.3).

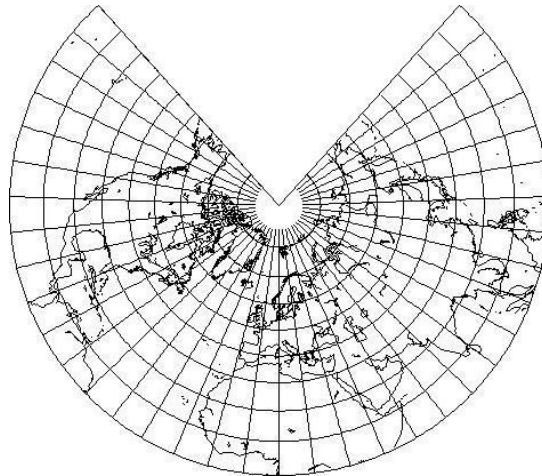


Figura 2.3. Proyección Cónica Conforme de Lambert.

Esta proyección fue ideada por el matemático alemán J.H. Lambert (1728-1777), el primero que dio carácter realmente matemático al estudio de las proyecciones cartográficas y que introdujo antes que nadie la idea de las proyecciones conformes y equivalentes.

La proyección Cónica Conforme de Lambert es la más apropiada para representar a todo el país y se utiliza para mostrar aspectos generales, de acuerdo a su forma, extensión y ubicación, se elaboran mapas con esta proyección a escala 1:1,000,000 y más chicas.

Proyección Mercator. En 1569, construyó Mercator su mapamundi en esta proyección, reseñando sobre el mismo mapa su fundamento y características. La Proyección Mercator consta de paralelos horizontales y meridianos verticales. Los meridianos equidistantes entre sí están colocados de tal modo que, en el ecuador, esta equidistancia está representada en verdadera magnitud a la escala correspondiente. Los paralelos están

dispuestos de tal manera que, en una zona de dimensiones relativamente pequeñas, la relación entre dos distancias tomadas respectivamente sobre meridianos y paralelos es igual a la relación entre las longitudes homólogas en el globo terráqueo. Por ejemplo, a los 60° de latitud, la distancia entre dos paralelos consecutivos es doble que en el ecuador; y como los meridianos guardan entre sí la misma separación en todas las latitudes, resulta que las dimensiones del mapa están exageradas en un ciento por ciento en la latitud de 60°. A los 80° de latitud, esta amplificación de dimensiones es de seis veces. Es evidente que en esta proyección no puede estar representado el polo, ya que los meridianos son paralelos entre sí y por lo tanto no se cortan. La proyección Mercator no es la proyección de una esfera sobre un cilindro, sino una modificación de esta clase de representaciones (Figura 2.4).

En la esfera los paralelos van siendo más cortos a medida que se acercan a los polos, y a su longitud es proporcional al coseno de la latitud. En la proyección Mercator, los paralelos tienen todos la misma longitud, lo cual significa que cada paralelo está aumentado en $1: \cos \delta = \sec \delta$, donde δ es la latitud expresada en grados. Para que sea una misma la escala para meridianos y paralelos, cada grado de latitud debe aumentarse en la secante de la latitud.

La proyección UTM es la que más aplicación tiene en el país, muchos de los trabajos cartográficos se han realizado en dicha proyección. Siendo el INEGI el principal en elaborar su cartografía en la proyección UTM, se utilizan en aspectos regionales y locales, así mismo sirven para elaborar proyectos de campo, entre otras aplicaciones. Para el caso de México, los parámetros utilizados en la proyección UTM son:

Proyección Transversa de Mercator
Esferoide de Clarke 1866
Meridiano Central: Para la zona 11 es -117°, zona 12 -111°, zona 13 -105°, zona 14 -99°, zona 15 -93°, zona 16 -87°
Latitud de Referencia: 0
Factor de Escala: 0.09996
Falso Este: 500,000
Falso Norte: 0

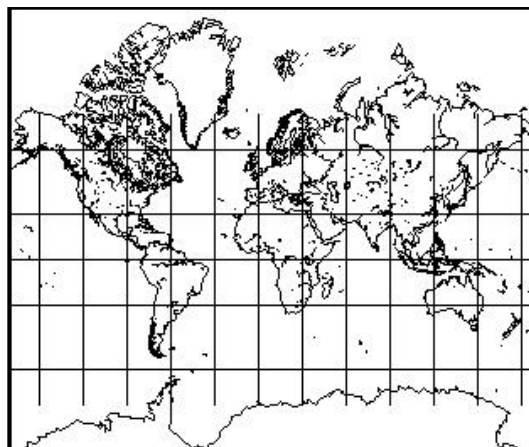


Figura 2.4. Proyección Universal Transversa de Mercator.

2.1.3 Escala

Se ha definido el mapa como una representación convencional de la configuración superficial de la Tierra. Toda representación, como toda imagen, está en cierta relación de tamaño (proporción) con el objeto representado. A esta proporción se le llama escala. Los cuadros y retratos corrientes no son ni mucho más grandes ni mucho más pequeños que el objeto representado y no requieren escala. Pero un mapa es millares y hasta millones de veces más pequeño que la superficie de la Tierra representada y es preciso poner la escala porque a simple vista no se aprecia la proporción a que está dibujado.

La escala, está directamente relacionada con el contenido, el propósito, objetivos, dimensiones y precisión del mapa. Ella caracteriza al mapa, y su correcta elección es determinante para representar mediante la expresión numérica, y/o la expresión gráfica.

La escala se representa generalmente, por los modos siguientes:

- Escala numérica o fracción representativa que da la relación entre la longitud de una línea en el mapa y la correspondiente en el terreno en forma de quebrado con la unidad por numerador.

$$\text{Esc} = 1:50\,000 \quad \text{ó} \quad \text{Esc} = 1/50\,000$$

- Escala gráfica, que representa las distancias en el terreno sobre una línea recta graduada. Esta escala tiene la ventaja de que sirve siempre que el mapa se reproduce por métodos fotográficos (Figura 2.5).

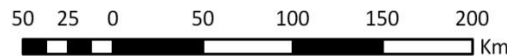


Figura 2.5. Escala gráfica.

La expresión gráfica de la escala la constituye un segmento de recta graduado a intervalos regulares que corresponden a las distancias reales del terreno. La porción izquierda, representada por el segmento 0 al 50, se denomina extensión debido a que es donde la escala gráfica tiene divisiones con menor espaciamiento y al resto, que es la mayor parte con espaciamientos más amplios, se le da el nombre de escala principal. La escala gráfica se usa como escalímetro para medir las distancias en el mapa y la extensión se utiliza para dar mayor exactitud. Esta escala, a diferencia de la numérica, tiene la particularidad de permanecer constante en forma relativa al mapa, a pesar de las reducciones o ampliaciones a que éste se someta.

Es muy importante mencionar que independientemente de los varios métodos para la georeferenciación, existen también servicios automatizados para facilitar el acceso a la información. Estos servicios se basan normalmente en sistemas de cómputo y están disponibles en varios formatos y métodos. Para el gacetero que se ha desarrollado en el presente trabajo se ha dispuesto de un contexto teórico-práctico de acuerdo al protocolo ADL, el cual basa su estrategia de diseño de gaceteros en una estructura que se menciona en el capítulo tres del presente trabajo.

2.2 Georeferenciación de colectas biológicas

La georeferenciación por topónimo (nombres de rasgos) es la forma más común de referenciar una ubicación geográfica y es un medio informal de georeferenciación. Normalmente entre las personas usamos nombres de lugares en la conversación diaria, correspondencia, informes y documentación. Hemos recurrido a ellas a través del tiempo para referirse a los lugares y los nombres de esos lugares se han ajustado a los tiempos cambiantes. Normalmente reutilizamos algunos nombres de lugares, por ser evidente que son los mismos nombres que han tenido con el paso del tiempo, o están basados en nombres que son familiares a los residentes que emigraron de otros lugares y épocas. La gente nombra todo tipo de características (por ejemplo, continentes, países, ciudades, montañas, ríos, humedales, edificios, parques, vías férreas, torres de radio y aeropuertos). Un estudio de la etimología de los nombres podría concluir que los nombres de los lugares son el resultado de nuestra historia y de nuestra cultura.

Existen algunos Diccionarios de nombres de lugares llamados también gaceteros, que contienen información descriptiva, que pueden incluir su ubicación geográfica, los tipos (categorías), y otra información. La Enciclopedia Electrónica Columbio (sexta edición, 2003) proporciona alguna información interesante acerca de los orígenes de los gaceteros. Como puede apreciarse, en el gran número de nombres de lugares y ubicaciones geográficas pueden o no existir otros datos importantes que se relacionan con esos lugares, tales como especies animales, de plantas y otras.

La necesidad principal consiste en establecer un punto sobre la superficie de la tierra en donde se encuentran las colectas biológicas de interés, esto es georeferenciar. El impacto del cambio global en la biodiversidad puede ser observado en su forma más crítica a través del efecto que produce en el comportamiento de los rangos de distribución de las especies, este puede generar tanto restricciones como ampliaciones en los rangos, lo que conlleva efectos locales y regionales que definen la estructura y funcionamiento de los ecosistemas. Se hace urgente el generar sistemas computacionales que cuantifiquen estos efectos y que permitan proyectar los resultados dentro de un escenario de cambio global, con el objetivo de servir como herramienta para los científicos mediante la evaluación de escenarios, como para los tomadores de decisiones del sector público y privado para la planificación de acciones en el territorio. Una de estas herramientas computacionales es el gacetero digital y como puede observarse, es de gran ayuda en lo que se refiere a almacenar y gestionar localidades para poder georeferenciar datos de colectas biológicas posteriormente.

En México, la georeferenciación espacial de la flora y fauna nativa ha sido desarrollada en muy pocas investigaciones. Sólo se ha cartografiado la distribución de las principales especies arbóreas y de animales. El principal problema que ha existido es el escaso número de colectas que tienen una relación con su ubicación espacial (georeferenciación), lo que no ha permitido llevar a cabo aplicaciones de modelamiento espacial confiables. Como consecuencia de esto el tratamiento de las especies nativas, ha sido en términos de localidades individuales más que de entidades biológicas dependientes a un área espacial en particular.

En la actualidad se están usando distintas técnicas de georeferir con el fin de determinar el grado de asociación entre conjuntos de variables bióticas y abióticas y las distribuciones de las especies. Estas asociaciones pueden ser usadas para predecir

respuestas dentro de un contexto de cambio global. El estado de arte en la modelación de los rangos de distribución de especies se basa en complejos análisis estadísticos y el aprovechamiento de los nuevas capacidades computacionales, las cuales facilitan la determinación y el manejo de conjuntos de variables que se correlacionan con los rangos de distribución de una especie dada.

En esta insistencia para la modelación de los nichos ecológicos y de la distribución geográfica potencial de las especies nativas, se utilizará normalmente un sistema de cómputo basado en las reglas de los llamados gaceteros. Especialmente con el objetivo de dar un tratamiento a los datos de localidades para poder georeferir colectas biológicas posteriormente.

El modelo de gacetero ha cambiado significativamente desde sus primeras manifestaciones como las listas alfabéticas o jerárquicas impresas de nombres de lugares con la información asociada a las coordenadas geoespaciales. Ahora los gaceteros se consideran un tipo de sistema de conocimiento que organiza la información sobre el nombre de lugares geográficos. Este tipo de sistemas relacionan etiquetas de texto para los lugares (es decir, los topónimos) a los sitios geoespaciales (por ejemplo, latitud y longitud) y de un sistema de categorización (por ejemplo, las áreas administrativas, las montañas, edificios). Estos sistemas llamados gaceteros son básicos en la relación lugar-dato.

Los gaceteros desempeñan diversas funciones en la gestión y recuperación de la información. Proporcionan servicios de traducción entre los medios formales e informales de la georeferenciación, para ello, se vinculan con las huellas geoespaciales (topónimos). Una aplicación que utiliza esta capacidad de traducción es conocida como *geoparsing*, donde las referencias en el texto se buscan en un diccionario para que las coordenadas geoespaciales puedan ser asociadas con el texto. También es útil para la recuperación de información, clasificar los lugares geográficos de acuerdo a un esquema, lo que permite respuestas a preguntas como "¿qué escuelas están en o cerca de Ciudad Universitaria, México?" Como fuentes de referencia básica, los gaceteros responden a las preguntas "donde está" y también describen los lugares en términos de su historia, sus orígenes, y otra información descriptiva.

Para la investigación académica, los gaceteros registran los hechos básicos conocidos acerca de los lugares históricos o enlaces a sitios de investigación y los mapas correspondientes a esos lugares. Para la gestión administrativa de los topónimos, la localidad o el país, estos gaceteros son el conjunto de datos primarios y cuya información sobre los lugares con nombre es esencialmente de naturaleza local. Es decir, el mejor, el más detallado, y con la información más actualizada acerca de los lugares con nombre que se conoce a nivel local. Los gaceteros se suelen crear ya sea para fines locales (por ejemplo, nombrando a los parques, barrios, distritos administrativos, edificios, instalaciones) o están relacionados con alguna actividad, como los viajes, la investigación genealógica, la investigación histórica, de ciencia-investigación, actividades empresariales, etc. La naturaleza local de los datos del gacetero lleva a dispersos, diccionarios independientes que guardan datos sobre los lugares en cuestión. Añadir a esto la naturaleza siempre cambiante de los datos del gacetero (por ejemplo, los nombres de los lugares cambian con el tiempo), y está claro

que hay una necesidad de acuerdo sobre las normas de acceso y el intercambio de datos diseñados específicamente para distribuidos.

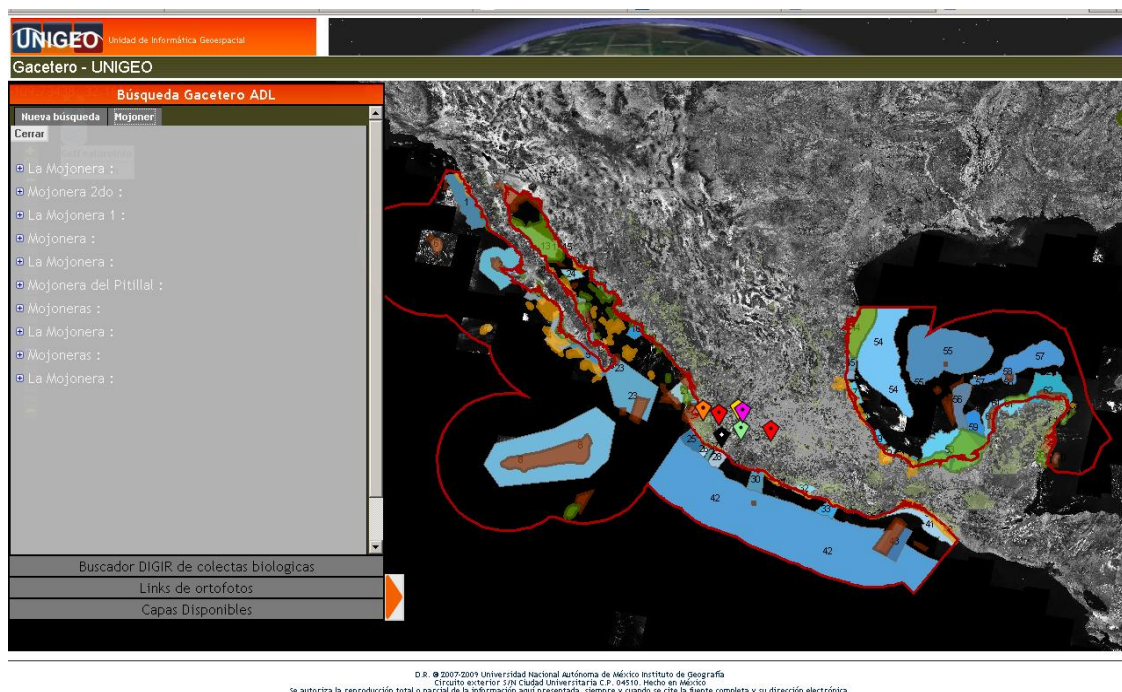


Figura 2.6. Ejemplo de un gacetero de localidades.
(Fuente: <http://132.248.26.13/trac/unigeo/wiki/ClienteGWTGacetero>)

El estudio minucioso de la teoría en la que está basada la creación de gaceteros conlleva a conocer la estructura en la que está basado el gacetero propuesto en este trabajo; por lo que cabe destacar que el Protocolo de ADL Gazetteer, en el que está sustentado el gacetero propuesto, está siendo desarrollado por el Open Geospatial Consortium (OGC) y cuyo detalle se verá más adelante en este mismo documento, especialmente el capítulo siguiente.

2.3 Antecedentes del desarrollo de gaceteros

Existe gran diversidad en los enfoques para la descripción de los lugares geográficos y, hasta hace poco, no más allá de la normalización de nombres propios. Hay muchos tipos de gaceteros impresos. Algunos diccionarios geográficos, tales como el de Columbia Gazetteer of the World, proporcionan información descriptiva sobre los lugares, incluido el historial de la localización, datos de población, altitud, o la pronunciación de su nombre, entre otros tópicos.

Algunos conjuntos de topónimos se usan para la recuperación de la información, los cuales son utilizados para describir los documentos, objetos de museo, o de otros recursos de información. El gacetero Getty de Nombres Geográficos se utiliza

ampliamente como un vocabulario controlado de los nombres históricos y actuales para describir el arte y la literatura de la arquitectura y las tenencias del lugar en cuestión.

Con la finalidad de establecer la importancia que tiene la estructura de todo gacetero digital, especialmente con el que se propone en este trabajo, se hablará un poco de la historia de ADL¹⁶. La Biblioteca Digital de Alejandría (ADL) se estableció en la década de 1990 como una respuesta a varios problemas de percepción de las bibliotecas de mapa tradicional, en particular el acceso y la organización. En 1999 se había convertido en una biblioteca digital funcional, ofreciendo un conjunto bien definido de servicios a una amplia comunidad de usuarios, basado en una extensa colección de objetos de información georeferenciada. La visión de ADL sigue evolucionando, ya que la tecnología hace posible nuevos servicios, al mismo tiempo los usuarios se vuelven más sofisticados y exigentes, lo que ha llevado a que el campo de la ciencia de información geográfica (GIScience) determine nuevas posibilidades de investigación y aplicación

Los conceptos propuestos por primera vez por ADL, se tornaron más tarde en otro gran número de proyectos, y muchos sitios tienen ahora las características de un geolibrary. La más prominente de estas características es, por supuesto, la posibilidad de buscar por ubicación geográfica, representada por la latitud y longitud, un concepto que es totalmente nuevo para la comunidad bibliotecaria. Un sistema digital del tipo geolibrary sólo puede existir en un mundo digital, por lo tanto, sigue siendo uno de los conceptos más poderosos que han salido de la investigación de la biblioteca digital.

En un principio, el equipo de investigación de ADL combinó los dos principales diccionarios geográficos del gobierno federal de los EE.UU. en uno de esos gaceteros contiene información de más de cuatro millones de entradas. Como resultado de esta experiencia, ADL ha desarrollado la ADL Gazetteer de contenido estándar y un diccionario solo para características y luego reconstruido en este nuevo modelo. Este gacetero está integrado de las coordenadas de las consultas de las colecciones de ADL. También hace uso de búsquedas a través de la ADL Gazetteer y mediante el Protocolo de ADL Gazetteer.

Durante el período de tiempo que ADL utilizó en la investigación y desarrollo de su modelo de gacetero, la Organización Internacional de Normalización (ISO), con responsabilidad para el ámbito de la información en el campo de la información geográfica digital desarrolló un estándar para gaceteros, y el Open Geospatial Consortium (OGC) también desarrolló un protocolo de gaceteros basado en la norma ISO.

A partir de 1998, varios talleres se han celebrado y se han centrado en los gaceteros, dada su importancia. Los informes y las presentaciones de cada uno de estos talleres están disponibles en línea. El taller distribuidos Geolibraries se celebró en 1998, incluyó en sus conclusiones la siguiente declaración: "Un gacetero integral, vinculando el nombre de lugares y ubicaciones geográficas, sería un componente esencial de una geolibrary distribuida".

¹⁶ <http://www.dlib.org/dlib/may04/goodchild/05goodchild.html> recuperado el 17 de octubre de 2009

En 2002, un taller sobre gaceteros digitales se realizó en el marco de la Conferencia Conjunta de Bibliotecas Digitales (JC DL) en Portland, Oregon, patrocinado por la Organización de Sistemas de Conocimiento en red / Servicios (NKOS) del grupo. El objetivo de este taller fue sobre gaceteros digitales como componentes de las bibliotecas digitales, protocolos y herramientas para el gacetero.

2.4 Bibliotecas digitales

Son una colección de bases de datos que ofrecen documentos en formato digital, disponibles en línea para el usuario. Estas bases de datos normalmente contienen información general, la cual puede ser accedida por temas de interés, solo basta especificar ya sea una o varias palabras, un tema, un nombre o cualquier palabra que sirva de comparación con el contenido de la base de datos que se está consultando. Normalmente, incluyen títulos de libros, revistas y fuentes de información mexicanas e internacionales. Muchos de estos documentos digitales están en texto completo (textos originales), ya sea de revistas, periódicos, fuentes noticiosas, libros; así como recursos de Internet clasificados. Además las fuentes están organizadas en bases de datos, agrupadas en colecciones, según el tema o el tipo de información que contienen.

Los gaceteros digitales a los que se ha hecho referencia a lo largo del presente trabajo, son un ejemplo concreto de una biblioteca digital, más aún, el contenido de dicha biblioteca está especializada en almacenar datos georeferenciados para fines muy específicos. Esta georeferenciación está dedicada a las colectas biológicas que son importantes para aquellas personas cuyo principal interés es la investigación de especies biológicas.

Ya anteriormente se ha mencionado en este mismo trabajo que la ADL hace siempre la descripción de lugares geográficos con sus respectivos nombres y esto puede llegar a ser bastante complejo si se quiere diseñar para documentar todos los aspectos de topónimos, de lugares, de referencias geoespaciales, o las relaciones entre los lugares, los cambios que ha sufrido a lo largo del tiempo, y otra información descriptiva. Por otra parte, se sabe que un subconjunto de esta información es todo lo que se necesita para apoyar la documentación básica de los lugares con nombre para la gestión y recuperación de información; característica imprescindible de cualquier biblioteca digital.

Un gacetero digital intentará normalmente extraer información de ciertas bases de datos, habiendo tres elementos básicos de una entrada o consulta en el gacetero:

- Un nombre (podría haber variantes de nombres también)
- Una clase o tipo (seleccionado de un sistema de tipificación de categorías de lugares o funciones)
- Un lugar (una huella; coordenadas que representa un punto)

Estos tres elementos tienen al menos un nombre, un tipo, y una huella geoespaciales en una búsqueda tradicional. Estos elementos básicos permiten la traducción de un elemento de un lugar a otro, y esta capacidad es compatible con funciones de gran alcance en los servicios de información, tales como permitir a los usuarios buscar un catálogo empezando con un topónimo y ampliar la búsqueda a la ubicación geoespacial, con lo que la recuperación de elementos tales como imágenes de teledetección o mapas de los lugares, se hace también fácil de encontrar.

Una Biblioteca digital¹⁷ es una colección de información que se guarda y accede electrónicamente y donde toda la información contenida debe referirse a un tema común. Es decir, una Biblioteca digital puede diseñarse y utilizarse para innumerables temas como gráficas por computadora, sistemas operativos, redes o cualquier otro tópico; todas estas librerías separadas pueden combinarse bajo una misma interfaz, pero es esencial que la información contenida dentro de cada una se mantenga separada.

El propósito de estas bibliotecas es mantener una locación central para el acceso a información sobre un tema en particular. Lo que un usuario quiere, cuando busca información sobre un tema específico, es encontrar lo que realmente le interesa. Una biblioteca digital debe guardar los temas separados y ordenados con cierta estructura, de otra manera sería totalmente inútil. También debe tener una interfaz de usuario que sea amigable; estos son los aspectos que se han considerado en el gacetero propuesto en el presente trabajo.

¹⁷ <http://www.acm.org/crossroads/espanol/xrds5-2/diglib.html> recuperado el 20 de octubre de 2009

Análisis y Diseño del Sistema.

En el presente trabajo se realizó el análisis y diseño de un sistema para manejar bases de datos de colectas biológicas y administrar registros geográficos, que recibe el nombre de gacetero digital. Este se basa en los mismos lineamientos que cualquier sistema de cómputo convencional requiere, es decir, se abordaron las fases de análisis, diseño, desarrollo, pruebas e implantación, con todas estas fases se produjo el gacetero digital que almacena las localidades que permitan la georeferenciación de colectas biológicas. La principal base teórica empleada en este sistema se extrae de la propuesta de la ADL, que está basada en protocolos que consideramos adecuados en la implementación de este tipo de gaceteros digitales.

La disponibilidad actual de las computadoras ha generado un incremento y una diversificación de la información, tanto para la sociedad en general, como para los negocios en particular. La administración de la información que se genera por computadora, difiere en diversas formas de aquella que se obtiene manualmente. A menudo, se tiene una mayor cantidad de información si ésta se genera utilizando sistemas computacionales; los costos para crear y mantener la información computarizada, son aparentemente mayores; la información que genera la computadora puede llegar a multiplicarse a velocidades impresionantes y con frecuencia, se trata con menos escepticismo que la obtenida por otros medios.

Existen diferentes tipos de sistemas de información, particularmente el gacetero digital para la gestión de localidades para permitir georeferir colectas biológicas es un ejemplo del tipo de sistemas administrativos. La metodología empleada en el desarrollo fue la tradicional, que consta de definir un problema, analizar los requerimientos del usuario y la información que ya se tiene, diseñar, desarrollar e implementar el sistema. Para fines prácticos en el presente trabajo solo se hará énfasis en la propuesta de ADL, respecto a la estructura que propone para los gaceteros digitales, pero siguiendo el desarrollo metodológico tradicional mencionado.

En otras palabras, en este documento se pretende dar un marco de referencia para el diseño de sistemas de información, que sea adecuado para mantener el contenido de un gacetero digital, basado en el protocolo ADL. Éste propone estándares para desarrollar sistemas del tipo gacetero. El estándar propuesto se llama GCS (Gazetter Content Standard) y contempla un conjunto de estructuras de datos llamadas tablas que contienen los topónimos que se encuentran en los Sistemas de Información Geográfica comunes. El análisis y diseño del gacetero digital, motivo del presente trabajo, contempla estructuras totalmente válidas del protocolo ADL.

3.1 Requerimientos del sistema

De acuerdo a las necesidades que surgieron dentro del Instituto de Geografía de la UNAM, se procedió a verificar un modelo conocido como ADL desarrollado por la Biblioteca Digital de Alejandría, de la Universidad de California, en Santa Bárbara y se pudo observar que este modelo abstracto contiene una serie de características que permiten apegarse a un estándar. Puesto que éste ha sido aceptado por la mayoría de los gaceteros para almacenamiento de datos con fines de georeferenciación, se procedió al estudio de su estructura, de la cual no debió perderse de vista que todavía es un modelo abstracto semiformal de un gacetero, pero su esencia fue fundamental para poder partir hacia una nueva propuesta, la cual pudo adaptarse a los requerimientos del propio Instituto de Geografía.

En realidad, el estudio de ADL junto con los requisitos del Instituto arrojó una serie de características dignas de aprovecharse para generar un gacetero que fuera lo suficientemente adaptable a lo que realmente se esperaba resolver. Estas características pudieron dar respuesta parcial a la serie de preguntas que se han generado en la sección 1.6.1. La parte rescatable de este modelo posee las siguientes características:

1. Un gacetero es un conjunto de entradas a un sistema del tipo catálogo de localidades georeferenciadas.
2. Un registro o entrada del gacetero describe un lugar geográfico único de un sitio y varios atributos o características de dicho lugar: uno o más nombres, una o más referencias, y cero o más clases. No hay una estructura en particular o específica de un gacetero en este modelo más allá de la contención simple de los registros o entradas del gacetero, aunque las relaciones entre estas entradas o registros pueden ser representadas de manera explícita por el gacetero.
3. Un identificador es una cadena que reconoce de forma inequívoca la entrada o registro en el gacetero. El identificador no tiene que ser universal o único.
4. Cada lugar se debe identificar por un código específico y único, uno o más nombres y poseer datos de tiempo y espacios georeferenciados.
5. El estatus del lugar en cuestión debe ser temporal, que pueda cambiar de acuerdo a un acontecimiento, ya sea geográfico, histórico o simplemente aleatorio.
6. El nombre no debe prestarse a modificaciones, en caso de ser así, debe guardarse un histórico de nombres del lugar. Una entrada o registro puede tener más de un nombre, en cuyo caso se usarán nombres alternativos para la descripción del lugar.
7. Una huella es una aproximación, expresada en coordenadas de longitud / latitud y puede consistir en una serie de áreas contiguas, es decir, lugares contiguos y diferentes con características específicas.
8. Una clase clasifica el lugar con respecto a un conjunto de términos predefinidos. Más específicamente, una clase es la asociación del lugar, con un catálogo

establecido de un vocabulario designado. Una entrada o registro del gacetero puede pertenecer a varias clases e incluso ser del mismo catálogo.

9. Algunos valores de los atributos o campos de una entrada o registro del gacetero (es decir, de cada nombre, cada huella, y cada clase) están más calificados con dos tipos de clasificación. El calificador principal, indica si el atributo es el valor preferido. Por ejemplo, un lugar dado puede ser una Colonia, o bien una Delegación.
10. Para cada entrada o registro del gacetero, las condiciones siguientes son indispensables:
 - a. Solamente un nombre debe ser marcado como principal.
 - b. Solamente una huella debe ser marcada como principal.
 - c. Si la entrada o registro ha sido clasificada, al menos una clase debe ser marcada como principal.

En el diagrama de la figura 3.1 se muestra gráficamente estas características, de las cuales se han tomado todas ellas y otras tantas de otra variante del mismo Gazetteer Content Standard (GCS), el cual se describirá con más detalle en la siguiente sección.

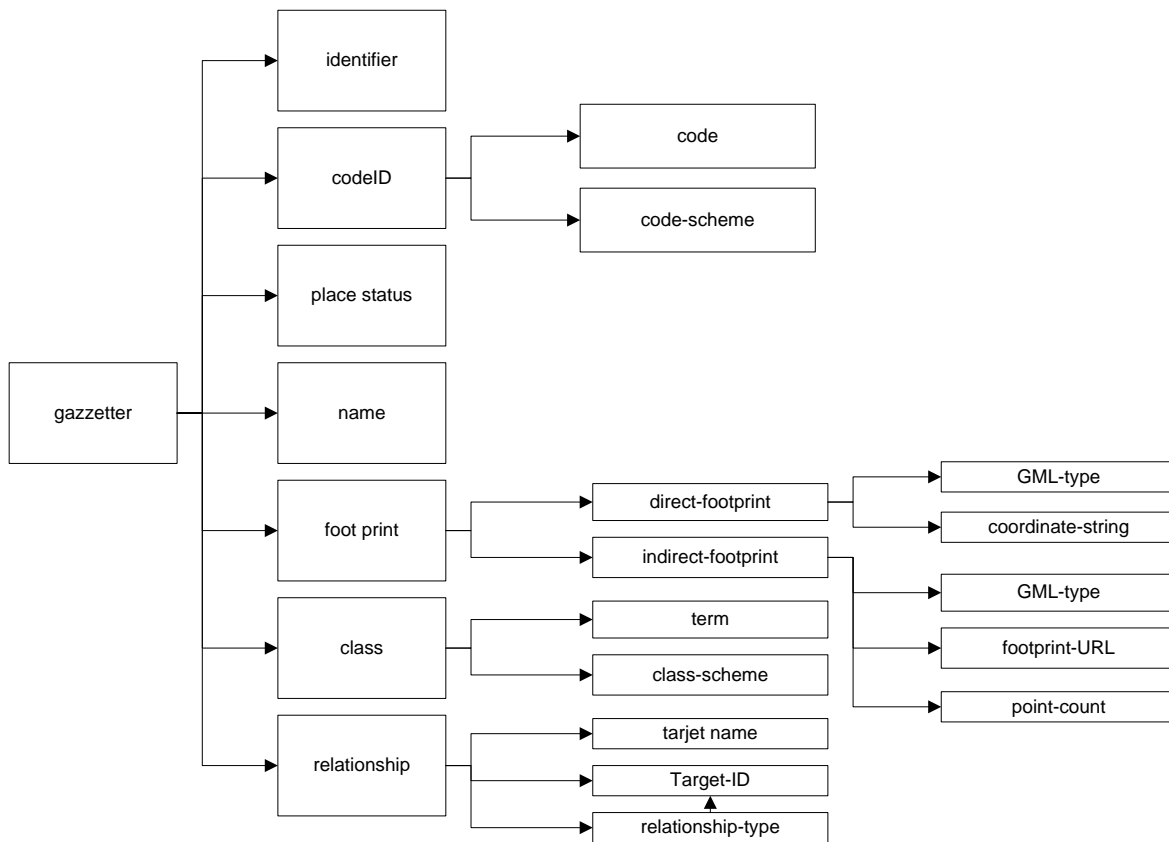


Figura 3.1 Modelo abstracto de un gacetero especificado por el protocolo ADL

El modelo GCS provee otra serie de características de diseño complementarias que en conjunto dieron las peculiaridades del nuevo diseño del gacetero digital. Éste pudo lograr una ventaja competitiva gracias a su potencial y llegar a establecer un estándar respecto a gaceteros digitales para manipular datos de localidades en el uso y explotación de colectas biológicas. El Instituto de Geografía de la UNAM está siempre a la vanguardia en la innovación de métodos apropiados en la explotación de la información mencionada ya que constantemente se revisan las necesidades de los usuarios de esta información.

En la etapa de requerimientos del sistema es importante tomar en cuenta las capacidades del Instituto de Geografía, así como las necesidades de las Instituciones que utilizan el servicio de gacetero digital; es decir, hubo que conocer la necesidad de biólogos, botánicos y personas que requieren contar con la descripción del sitio donde ha sido realizada cierta colecta de un espécimen, pero siempre y cuando esta información fuera susceptible de guardarse junto con sus coordenadas geográficas y otros datos de interés.

A continuación se mencionan los requerimientos que debió cumplir el nuevo gacetero digital, los cuales fueron enlistados tomando en cuenta las opiniones de los usuarios:

- Generar un sistema de cómputo flexible y fácil de utilizar que permita incorporar la información existente dentro del Instituto de Geografía de la UNAM, a este nuevo desarrollo llamado ahora gacetero digital.
- El gacetero digital debe contar con un diseño moderno y que sea capaz de almacenar los datos ágilmente, de preferencia basándose en los estándares internacionales, tales como la ADL (Alexandria Digital Library), ya que de contar con un diseño de este estilo facilitaría el intercambio de información entre las diferentes versiones existentes, puesto que ADL captura una gran variedad de atributos que son necesarios para cumplir correctamente con la tarea de georeferenciar los especímenes.
- Garantizar la utilidad del nuevo gacetero digital para la manipulación de datos de localidades y que en un futuro permita georeferenciar colectas biológicas, para lo cual es necesario que el diseño y formato de los datos empleados sean compatibles con la georeferenciación propuesta.

Este nuevo gacetero digital intentó ir más allá de los métodos hasta ahora utilizados, ya que los sistemas existentes tradicionalmente se basan en sistemas de posicionamiento global (GPS), los cuales han permitido acceder de manera más frecuente y con mayor precisión el dato de las coordenadas geográficas de un sitio de colecta. Debido a que el volumen de registros sin coordenadas es considerablemente elevado, existe una enorme cantidad de información que permanece desaprovechada, de ahí la importancia de guardar espacio para que esta información pueda ser completada en el futuro, sobre todo que el nuevo gacetero agilice las búsquedas de los datos de espacio, tiempo y localización de las colectas biológicas.

Si después de aseverar que la descripción de las localidades donde son recolectados los organismos, es realizada de diversas maneras, entre ellas una donde se involucra subjetividad, es decir, el colector describe un sitio de acuerdo a los medios que tiene a su disposición y a su propia percepción del lugar, entonces se tiene un problema importante y necesario de resolver: determinar la precisión de la localidad. En este caso el uso de la

tecnología, particularmente de los sistemas de información, será la mejor solución a este problema.

Ya se ha mencionado también que es preferible adoptar un protocolo debidamente probado como lo es ADL. Esto ha permitido finalmente contrarrestar el limitado acceso a la información disponible en los gaceteros digitales actuales.

El proceso de georeferenciación¹⁸ tiene por requisito contar con acceso a gaceteros pertinentes a las colectas, idealmente en formato digital. Estos gaceteros requieren contar con un conjunto de registros organizado de manera congruente, con información temporal y referencia geográfica, así como detalles sobre su creación. Estas características son necesarias para dar validez a los resultados obtenidos y comprobarlos geográficamente. Cuando la información que se pretende emplear para alimentar un gacetero cumple con estos requisitos, se considera que estará lista para incorporarse dentro del mismo.

3.2 Análisis de la información

Para esta etapa se ha realizado un planteamiento basado en estándares de la *ADL*, específicamente en el *Gazetteer Content Standard (GCS)*, lo cual resuelve de manera sencilla y ordenada la disposición de los campos y su agrupamiento, para que de llevarse a cabo este orden, se cumpla con el requerimiento original del sistema: crear un gacetero digital basado en un estándar.

El modelo *GCS* cubre otros aspectos importantes respecto a la estructura que debe guardar cualquier gacetero que pretenda intercambiar datos en forma eficiente, esto se debe al grado de detalle que es capaz de soportar, por ejemplo, datos históricos, fechas, información de periodos de tiempo, etcétera. Con esta capacidad fue muy fácil adaptar este modelo a las necesidades del nuevo gacetero digital y cubrir muy bien las respuestas a las preguntas que fueron planteadas anteriormente. Estas preguntas concretamente son:

- ¿Qué bases teóricas vamos a utilizar para estructurar el contenido del gacetero propuesto?
- ¿Cómo se almacena, representa y preserva la información histórica en dicho gacetero digital?
- ¿Cómo se preservan las distintas representaciones espaciales en un rasgo geográfico?
- ¿Cómo se puede garantizar la inclusión de información proveniente de fuentes diversas dentro de la estructura del gacetero digital propuesto?

¹⁸ Georeferenciación de localidades de colecciones biológicas, manual de procedimientos, CONABIO, México 2008, pp. 34

A continuación se describen algunos aspectos importantes que cubre GCS y como puede observarse, estos aspectos responden a las preguntas anteriores; aún así estará mas claro en la siguiente sección ya que es ahí donde se propone directamente el nuevo gacetero y su estructura interna.

ADL Gazetteer Content Standard (GCS)

ADL Gazetteer Content Standard (GCS) y toda su compatibilidad se debe a que también ha sido desarrollado por la Biblioteca Digital de Alejandría de la Universidad de California, en Santa Bárbara. El GCS está construido como un esquema XML.

“XML es una tecnología en realidad muy sencilla, que tiene a su alrededor otras que la complementan y la hacen mucho más grande y con mayores posibilidades. XML, con todas las tecnologías relacionadas, representa una manera distinta de hacer las cosas, más avanzada, cuya principal novedad consiste en permitir compartir los datos con los que se trabaja a todos los niveles, por todas las aplicaciones y soportes. Así pues, el XML juega un papel importantísimo en este mundo actual, que tiende a la globalización y la compatibilidad entre los sistemas, ya que es la tecnología que permitirá compartir la información de una manera segura, fiable, fácil. Además, XML permite al programador y al equipo de soporte dedicar esfuerzos a las tareas importantes, ya que algunas de ellas son tediosas. Por ejemplo, la validación de los datos o el recorrido de las estructuras va por cuenta del lenguaje y está especificado por el estándar, de modo que el programador no tiene que preocuparse por ello.”¹⁹

Se publica tanto en línea como un esquema y una aplicación de base de datos relacional. Es compatible con el modelo de gacetero abstracto de la ADL Gazetteer Protocol, con la excepción de que en este esquema, el elemento de clasificación es necesario. El GCS está diseñado para ser la estructura de datos para los gaceteros y, por consiguiente incluye muchos más detalles, así como directrices para rellenar los datos del gacetero digital, producto de este estándar. En contraste, el modelo abstracto está diseñado para apoyar la operación de los gaceteros sin importar qué estructura de datos locales se utilicen. Formalmente, las bases del GCS para su diseño de gaceteros tienen las siguientes premisas:

1. Una entrada en uno de los registros del gacetero es un simple rasgo con características. Además, una característica se diferencia de otra por su nombre, tipo (clase) y la huella. Es posible que para una localidad en particular existan dos rasgos con el mismo nombre. Esto ocurre cuando dos rasgos independientes de diferentes tipos ocupan el mismo espacio, por ejemplo, un Estado y su capital con nombres iguales.
2. Una localidad es un conjunto de entradas al gacetero y cada una de ellas tiene una identificación única en el gacetero.
3. Una entrada de gacetero completo consta de al menos una ocurrencia de cada elemento necesario y de cada atributo, cero o más elementos y atributos opcionales, y la atribución de la las fuentes de los datos. Las fuentes incluyen la identificación de los sistemas de clasificación, los contribuyentes, y los documentos que justifiquen la nueva entrada.
4. Cuando hay múltiples ocurrencias de un elemento necesario, uno de ellos es marcado como la llave primaria, esta ocurrencia se utilizará para mostrar en la pantalla principal, muestra en línea y los informes que se requieran.

¹⁹ <http://www.desarrolloweb.com/articulos/449.php> recuperado el 11 de noviembre de 2009

Estructura de nivel superior de GCS

La estructura de nivel superior del esquema que propone GCS se muestra en la figura 3.2, los primeros seis de los trece elementos de nivel son necesarios, lo que significa que se requiere al menos un subelemento de cada uno de estos. En resumen, los seis elementos necesarios y subelementos son:

- Una característica única de identificación para la entrada del gacetero.
- Una indicación de si el rasgo está en curso, activo o es una clasificación temporal de este rasgo.
- Al menos se tiene un nombre para la función y una clasificación temporal para el nombre dado como localidad.
- Al menos un término de tipo o número de clasificación, la identificación del sistema de tipificación utilizado, y una clasificación temporal para el tipo (como arriba).
- Al menos una huella espacial expresada tanto en la geometría y como delimitador, con una clasificación temporal y detalles adicionales que se puedan explicar posteriormente con más detalle.
- Un par de datos administrativamente importantes: la fecha en que la entrada al gacetero se realizó y cuando ha sido modificada o hubo adición de datos.

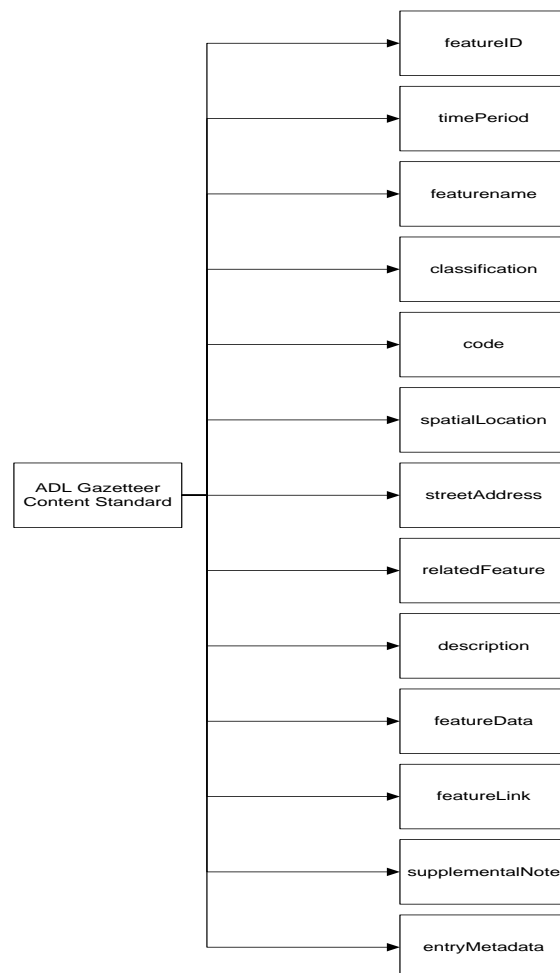


Figura 3.2 Elementos en el ADL GCS. Estos elementos muestran los tipos de información que pueden ser documentados para cada uno de los rasgos en un gacetero digital.

Elementos de periodo de tiempo de GCS

GCS también posee elementos de periodo de tiempo, los cuales se utilizan para añadir un estado temporal en general y las fechas de inicio y final del rasgo en sí y sus nombres, huellas, las relaciones, y las clases (tipos). Todos estos aspectos de los rasgos pueden cambiar a través del tiempo. Por ejemplo, el uso de un edificio puede cambiar de un uso religioso a un uso educativo, y por lo tanto, el estado debe cambiar las condiciones administrativas con la construcción y la disolución de leyes que avalen este tipo de cambios. En este esquema para la documentación de los periodos de tiempo, sólo el estado general conocido como *período de tiempo* es necesario; los demás elementos son opcionales. El mismo conjunto de elementos de periodo de tiempo es utilizado a través de GCS donde es necesario establecer fechas y periodos para los nombres, huellas, relaciones y datos asociados con el lugar, por ejemplo, la población.

El periodo de tiempo está documentado por cinco elementos y solamente el primero es requerido. Véase la figura 3.3

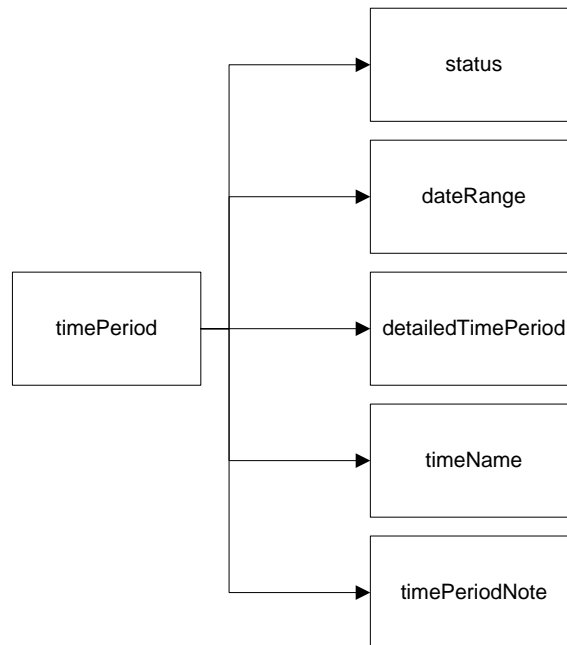


Figura 3.3 Partes de la documentación de periodos de tiempo en el ADL Gazetteer Content Standard

Con estas características se describen los periodos de tiempo:

1. Un estatus del periodo de tiempo de un lugar, rasgo, nombre, país, etc. Por ejemplo, el estatus de un nuevo país como Yugoslavia podría ser “en formación”.

2. Un rango de fechas con inicio y fin expresado de acuerdo a la norma estándar ISO 8601²⁰ y con una nota opcional para explicar su intervalo.
3. Un período de tiempo detallado expresado como uno de los siguientes:
 - a. Las fechas de inicio y fin, expresadas de acuerdo a la norma citada en el punto anterior, con la posibilidad de registrar los valores de confianza y notas acerca de cada una de ellas.
 - b. Cadena de fecha codificada en la que distintos sistemas de calendario pueden ser registrados de acuerdo con la norma citada en el punto anterior.
4. Un período de tiempo de vigencia del nombre del rasgo, (por ejemplo, la época napoleónica, 1799-1815) y, opcionalmente, citas que provean la forma del nombre y las fechas de inicio y final.
5. Un periodo de tiempo para agregar las explicaciones de las circunstancias inusuales.

Elementos de nombres de lugar de GCS

También existen los elementos de Nombres de Lugar para GCS en donde cada entrada al gacetero digital tiene un nombre para cada rasgo, esto es un topónimo inmodificable por el administrador del gacetero. Por ejemplo, el nombre de la ciudad de México es "México" no "México, D.F."

Puede haber varios nombres para un rasgo. Por ejemplo, la ciudad conocida como "Köln" es también conocida como "Cologne" y el país "Tailandia" era conocida como "Siam". Para fines de presentación, uno de los nombres es designado como el nombre principal de pantalla. Esta designación no implica corrección, sólo significa que en este registro del gacetero es el nombre que se mostrará en primer lugar, o el que se mostrará si se tiene un solo nombre. Esta distinción es importante porque el nombre oficial de un lugar no es universal. Por ejemplo, "Alemania" es el nombre autorizado para el país por la Junta de Nombres Geográficos de EE.UU., pero es conocido localmente como "Deutschland". Un gacetero que elige "Alemania", como el nombre principal y "Deutschland" como una variante de nombre no implica que un nombre sea más correcto que el otro, pero es sólo indicativo de que uno de ellos se usará para fines de llave principal.

Además de con el nombre principal de pantalla, no puede haber cualquier número de nombres adicionales. Las características que a continuación se listan pueden describir a cada nombre, tal y como se muestra en la figura 3.4.

²⁰ ISO 8601 es estándar internacional para fecha y tiempo, las representaciones se publicaron por International Organization for Standardization (ISO).

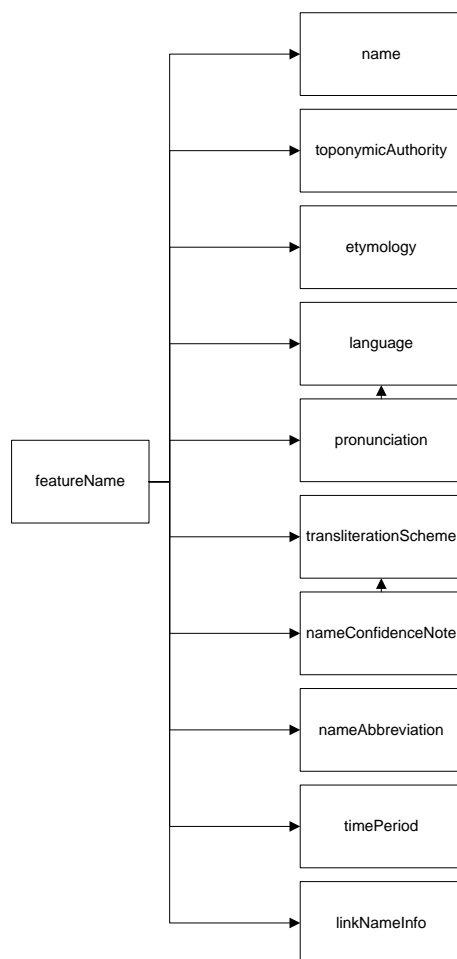


Figura 3.4 Partes de la documentación de nombres en el ADL Gazetteer Content Standard

En cada caso, sólo el elemento de período de tiempo es necesario, el cual puede ser indicado por el uso de nombres antiguos, actuales o propuestos.

1. La identificación de un funcionario o agencia que autoriza el uso de este nombre para este rasgo.
2. Etimología (derivación) del nombre.
3. Lenguaje de la denominación en forma de un código de lenguaje y la cita del código de esquema. Cuando el nombre es el mismo en varios idiomas, el idioma nativo es el representado aquí para evitar una lista de todos los idiomas (por ejemplo, "París" es utilizado universalmente para la capital de Francia, el idioma de "Paris" se cita como "FRA", el Código ISO 3166²¹ para el francés).

²¹ ISO 3166 es un estándar que codifica los nombres de países y áreas dependientes y sus principales subdivisiones.

4. Pronunciación del nombre en la forma de una cadena de texto o enlace a un archivo de audio y de una nota acerca de cómo se hace la pronunciación, por ejemplo, "en Francia" o "en el sur de Estados Unidos “.
5. Sistema de transliteración, si un lenguaje se utilizó para convertir el nombre a otro idioma diferente.
6. Expresión de la confidencialidad asociada a esta forma de nombre (por ejemplo, citando que el nombre fue tomado de una fuente arqueológica que podría solamente ser parcialmente leída).
7. Forma abreviada del nombre, por ejemplo, "DF" y "D.F." para "Distrito Federal".
8. Periodo de tiempo para el uso del nombre (con la estructura descrita anteriormente).
9. Vínculos o referencias a más información sobre el nombre, tales como enlaces a los documentos académicos que describen el origen del nombre.

Elementos de clasificación de rasgos en GCS

Los Elementos de Clasificación como son: clase, categoría o tipo, en rasgos tales como "ríos" o "edificios", son indispensables. Los términos de clasificación son seleccionados de un sistema de tipificación primaria para todos los registros en el gacetero y, opcionalmente, de otro esquema de tipificación de rasgos. En algunos programas de cómputo especiales se pueden utilizar, por ejemplo, sistemas de clasificación para los humedales o bosques. El tipo de esquema no se limita en modo alguno, pero los esquemas son generalmente un vocabulario simple de términos jerárquicos (por ejemplo, vocabularios controlados, estructurado con las relaciones entre términos). Tómese en cuenta que si un gacetero consta de una sola clase de lugares, como por ejemplo cementerios, cada uno de sus registros, necesariamente debe llevar el registro de su clasificación para apoyar la interoperabilidad y la búsqueda de información a través de los gaceteros distribuidos. Un periodo de tiempo puede ser utilizado para mostrar que el tipo de rasgo, como un edificio, ha cambiado con el tiempo (por ejemplo, de una iglesia a una escuela).

Elementos de código de rasgo en GCS

Un rasgo puede ser conocido como un código de un sistema de numeración o por un estándar del Federal Information Processing System (FIPS) que se utiliza en los Estados Unidos para las unidades administrativas (por ejemplo, "39" es el código para el estado de Ohio y "39-18000" es el código para Columbus, Ohio). El código en sí y el régimen asociado a él puede ser documentado.

Elementos de localización espacial de GCS

Los elementos de localización espacial en el GCS se muestran en la figura 3.5. Hay dos maneras de representar las huellas espaciales para el lugar: por una o más geometrías y por un cuadro delimitador, es necesario uno de cada representación. El cuadro

delimitador es una generalización de la geometría, está incluido en cada rasgo como una huella de nivel básico que puede ser utilizado por todos los sistemas de información. Es especialmente importante para los sistemas que no cuentan con un software sofisticado de procesamiento espacial.

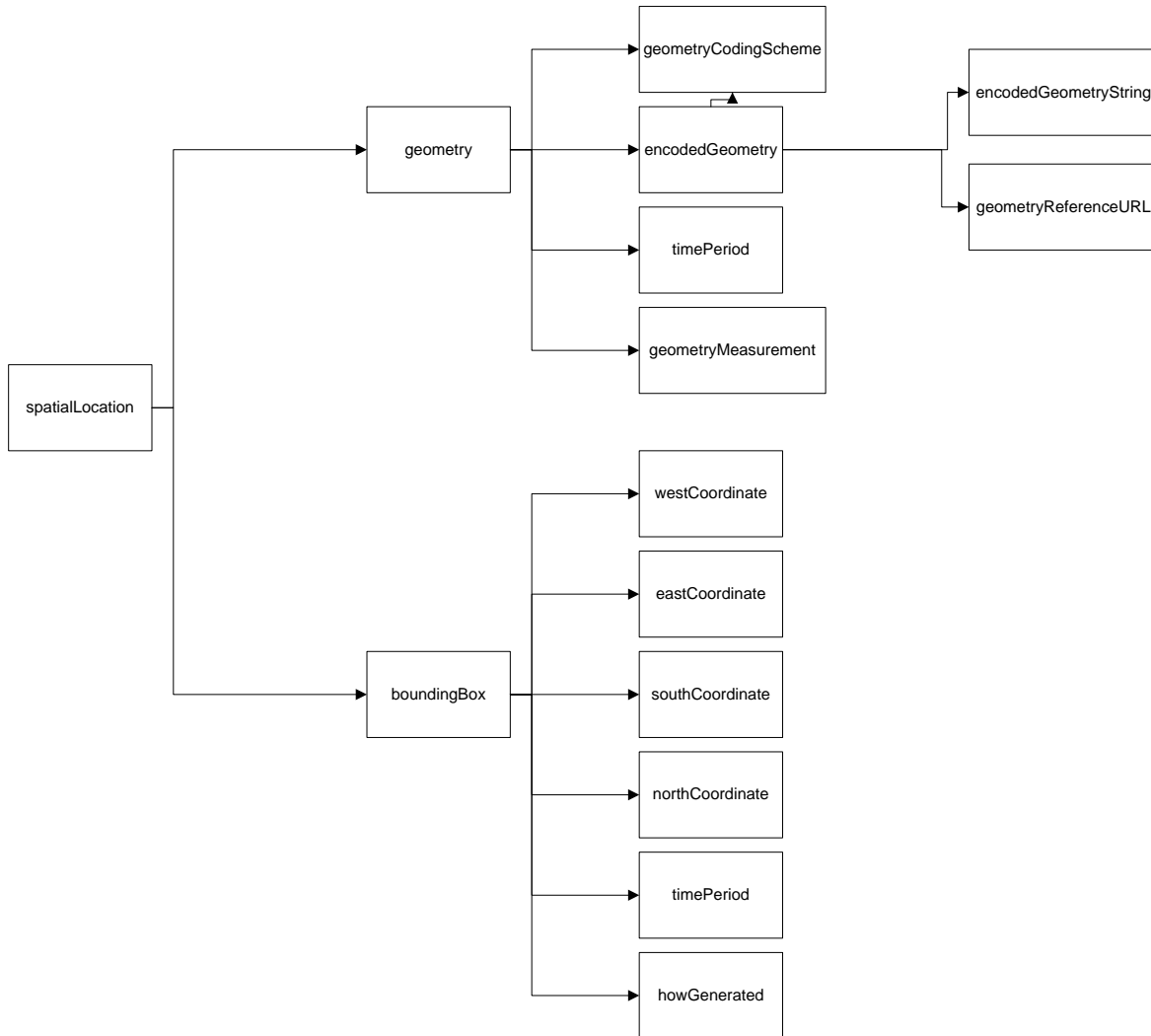


Figura 3.5 Partes de la documentación espacial del ADL Gazetteer Content Standard

La geometría es una representación de la huella de un rasgo, puede ser cualquier tipo de símbolo geoespacial y no necesita ser contigua. Por ejemplo, una huella para el estado de Hawaii podría consistir en una unión de los polígonos disjuntos, uno por isla. Un registro del gacetero digital puede tener más de una geometría, en cuyo caso representan diferentes interpretaciones, aproximaciones, o resoluciones de la ubicación de la huella y siendo una de ellas designada como la geometría primaria.

La geometría se expresa en forma codificada que está incluida, o con la entrada del gacetero como una cadena de puntos más bases geodésicas y el tipo de geometría (por ejemplo, punto, línea, polígono), o una referencia URL, de un archivo que contiene las coordenadas u otra representación de la ubicación geográfica. La geometría del

esquema de codificación utilizado en ambos casos, está documentado. Las Geometrías tienen periodos de tiempo (requerido) y la documentación se puede añadir acerca de cómo y cuándo fue medida la huella.

Cada rasgo tiene un cuadro de límite, que es generado con datos simples o múltiples. Las dos longitudes y las dos latitudes del cuadro de límite son los mismos a menos que alguna forma de búfer haya sido aplicado a ampliar el punto a un área. Si la geometría es solamente una caja, el cuadro delimitador es igual a la geometría del cuadro. Si hay varias geometrías, el cuadro delimitador puede representar hasta el máximo del conjunto o sólo la geometría primaria. El método utilizado para crear el cuadro delimitador está documentado.

Elementos de relación de GCS

La relación que se establece con otros rasgos puede ser a través de la descripción, esta descripción puede utilizar algunas palabras clave, por ejemplo “parte de”, “consta de”, “a partir de”, “desde la fecha”, “hasta la fecha”, “con el tipo”, etcétera, y aunque esto no es un elemento necesario de la descripción, dichas frases son la clave para entender ciertas dependencias entre dos o más rasgos en particular. Una relación está representada por el nombre de otro rasgo y opcionalmente, su característica única de identificación cuando se está disponible, así como los esquemas asociados de los tipos de rasgo. Estos esquemas puede ser un conjunto bien conocido a través de catálogos previamente definidos. Algunas relaciones pueden cambiar con el tiempo, es por eso que en el gacetero propuesto las relaciones temporales tienen mucha importancia, ya que deben ser almacenadas para poder explotar los datos más precisamente con el paso del tiempo.

Para cerrar este capítulo es necesario destacar que todos los elementos expuestos en el ADL GCS son la base para que el nuevo gacetero digital sea capaz de aceptar los datos que el Instituto de Geografía de la UNAM posee en sus acervos, tales como toponimias, rasgos de línea, polígonos y otros datos que sirvan para georeferenciar las colectas biológicas. Todos los elementos de este modelo ADL GCS posee información suficiente para auxiliar a la actividad de georeferenciación.

3.3 Diseño del sistema

La herramienta principal y la más usada en esta fase, es el modelo Entidad-Relación donde se destacan las relaciones entre los campos que conforman las entidades.

Diagrama de Entidad Relación

La técnica de entidad relación sirve para modelar una situación del mundo real desde el punto de vista de los datos que es necesario almacenar. Es decir, representar el modelo lógico o conceptual de los datos de la organización.

Para la técnica de modelado de Entidad-Relación se usan básicamente tres elementos:

- Las entidades, que se dibujan a través de un rectángulo.
- Las relaciones, que se dibujan a través de una línea.
- Los atributos, que son el contenido de las entidades, como el Código de Cliente.

Una entidad está identificada por medio de un sustantivo, porque refleja una idea, una cosa del negocio que necesita ser guardada. Por ejemplo, los datos del cliente son una situación comercial, los datos de un alumno es una situación académica. Tiene que ir en singular, porque básicamente lo que se representan son las ideas de una por una, lo más diferenciadas posible. Por ejemplo: cliente, empleado, alumno, etc.

Con la ayuda de un modelo entidad-relación y con los datos que han sido detectados en la fase de requerimientos del sistema, se ha podido llegar a conocer cuáles son los datos que deben estar almacenados en el gacetero digital. Si también se toma en cuenta la estructura de la ADL Gazetteer Content Standard para conformar adecuadamente los requerimientos de información, entonces es factible concebir una serie de entidades y atributos propios de un nuevo gacetero digital que cumple con los estándares de la ADL y a su vez con las necesidades reales de poder manipular datos de localidades para posteriormente georeferenciar las colectas biológicas descritas en capítulos anteriores.

3.3.1 Modelo Entidad-Relación del gacetero digital

Una vez que ha sido definido el concepto de modelo entidad-relación y que su uso expresa el alcance completo de la solución a los requerimientos del usuario, entonces es conveniente identificar los datos que deben incluirse en la base de datos para consultarse cuantas veces sea necesario.

De acuerdo a los requerimientos del nuevo gacetero digital, se separaron los datos en familias tales como:

- Rasgos.- son los elementos centrales del gacetero, y cada uno de estos representa a un objeto geográfico único e irrepetible dentro del gacetero.
- Clasificaciones de rasgos.- es la asignación de un tipo de dato a un rasgo concreto. Esta clasificación se ubica dentro de una estructura jerárquica donde cada uno de los tipos específicos corresponde a un término más amplio. Por ejemplo, una comunidad rural corresponde al término del lugar habitado.
- Descripciones de rasgos.- es información adicional tipificada acerca del rasgo, la cual explica situaciones específicas acerca del rasgo.
- Nombres de rasgos.- es el conjunto de nombres que posee un rasgo.
- Datos varios de rasgos.- son sus características propias, las cuales pueden pertenecer a diversos temas como topografía, orografía, información censal, etc.
- Notas suplementarias de rasgos.- es información que explica una circunstancia inusual con la entrada del gacetero.
- Localizaciones de rasgos.- es información espacial específica acerca del rasgo.
- Colecciones de rasgos.- son cada uno de los bloques básicos en los que se agrupa información proveniente de diversas fuentes.
- Geometría de rasgos.- es la información geométrica asociada a cada uno de las entradas del gacetero, la cual corresponde a polígonos o puntos.
- Esquemas.- es un catalogo con cada una de las representaciones empleadas dentro del gacetero.
- Términos de esquemas.- es un catalogo con todos los términos correspondientes a cada una de las representaciones empleadas en el gacetero.
- Códigos de rasgos.- es el identificador externo del rasgo.

- Rasgos relacionados.- es el conjunto de relaciones existentes entre los rasgos contenidos en el gacetero.
- Rangos de fechas.- es una entrada que define de manera breve el inicio y término de un evento temporal.
- Periodos de tiempo.- es el conjunto de relaciones entre registros y eventos de tiempo específicos.
- Nombres de periodos de tiempo.- es una entrada que asigna un nombre concreto a un periodo de tiempo.
- Codificación de fechas.- es una entrada que permite incorporar fechas codificadas en algún esquema ajeno al sistema, por ejemplo, calendario romano, calendario azteca, entre otros.
- Fechas de inicio y fin.- es el conjunto de entradas con información específica de los eventos temporales, semejante a los rangos de fecha, pero permitiendo asignar un valor de confianza a las fechas.

La figura 3.6 muestra las entidades encontradas junto con sus relaciones, las cuales son la base del nuevo gacetero digital.

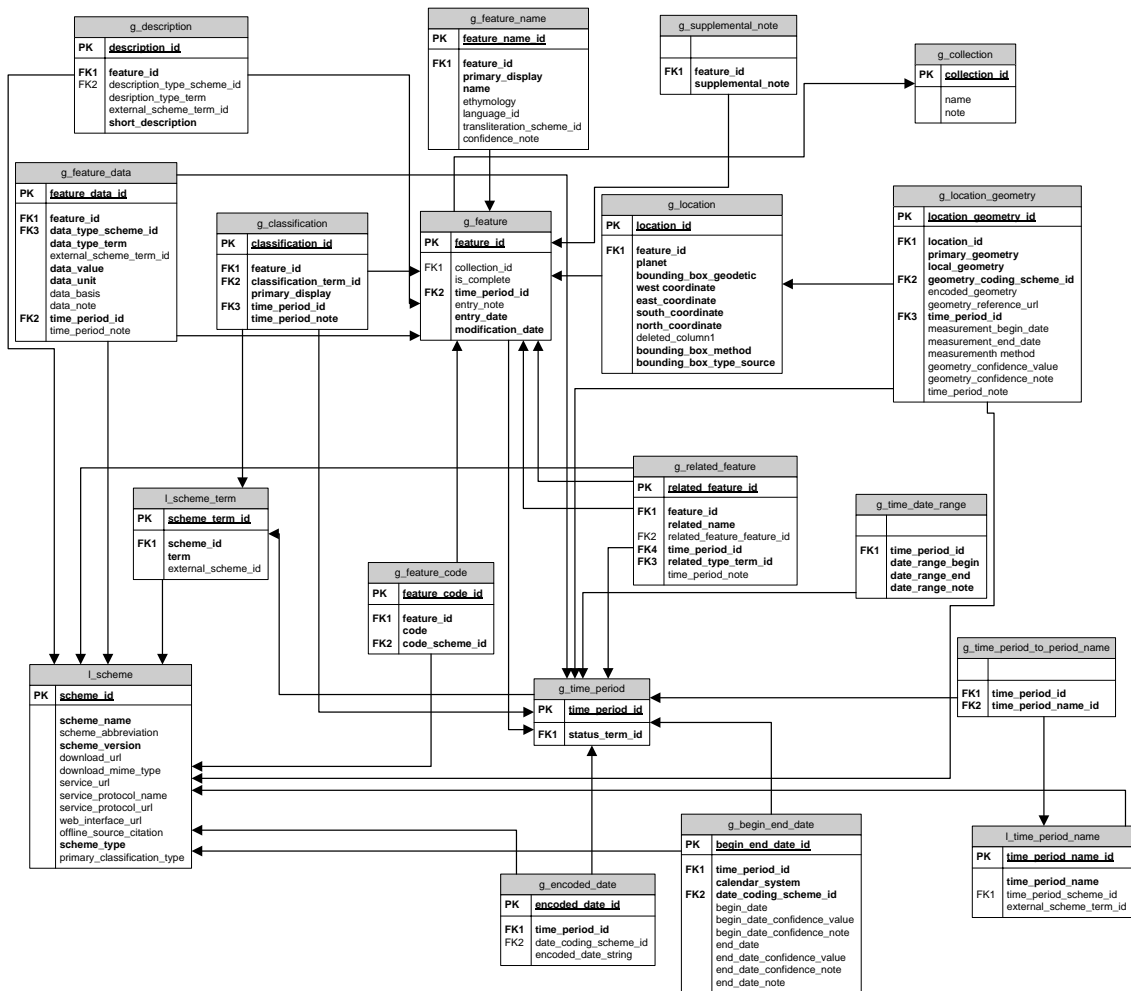


Figura 3.6 Modelo Entidad-Relación del gacetero digital

El modelo Entidad-Relación anterior enfatiza los atributos de cada una de las entidades encontradas; estos atributos son en realidad los elementos individuales que conforman la base de datos. Además cada uno de estos elementos se convierten en los llamados campos y cada campo puede ser capturado individualmente dentro de la base de datos a través de la interfaz de usuario.

Los campos mostrados en la figura 3.6 permiten que un registro dentro del gacetero digital arroje toda la información ordenada y dispuesta en la pantalla de la computadora, misma que describe a lujo de detalle una localidad, colonia, pueblo, país, municipio, etc.

3.3.2 Diseño de la base de datos

La etapa del diseño también incluye la conformación de los archivos o la base de datos que almacenará aquella información requerida por quien toma las decisiones en la organización. Una base de datos bien elaborada es fundamental para cualquier sistema de información. En esta etapa, el analista diseña la salida (en pantalla o impresa) hacia el usuario, de acuerdo con sus necesidades de información.

Existen en el mercado distintas tecnologías para la creación de bases de datos, pero cualquiera que se apegue a las necesidades del gacetero digital deberá cumplir con una característica fundamental: que sea una base de datos relacional. Ésta almacena los datos en tablas separadas en lugar de colocarlos en una sola de gran tamaño. Una base de datos relacional permite al Administrador de la Base de Datos (DBA's) definir las relaciones entre estas tablas apoyándose principalmente en un modelo entidad-relación previamente diseñado, con lo que garantizará que no haya repetición de datos, que éstos sean los adecuados y suficientes y sobre todo, que estén seguros.

Tal y como se observa en el modelo Entidad-Relación del gacetero, todas las entidades encontradas se convierten en tablas dentro de la base de datos y guardan una estrecha relación unas con otras. Estas relaciones permiten al DBA combinar datos de varias tablas de consulta y presentación de informes de manera rápida y eficiente.

La explotación de los datos almacenados se logra mediante el uso de claves o campos llave, que son utilizados para identificar los registros específicos de una tabla. Este es el potencial más notable de las bases de datos relacionales, ya que permite un acceso a los datos más grande, más rápido y más eficiente.

Dentro del marco de desarrollo de sistemas, el analista debe optar por usar una base de datos que se encuentre en el mercado y que sea económica, eficiente y lo suficientemente robusta para soportar sus aplicaciones. Una vez que se ha optado por una de ellas es indispensable que se utilice el lenguaje de consulta estructurado (SQL). SQL (Structured Query Language) que es el más común, ya que emplea un esquema normalizado para tener acceso a las bases de datos.

En la figura 3.7 se muestra la interfaz del administrador de la base de datos en donde se desarrolló el prototipo del gacetero. En realidad este tipo de herramientas normalmente se utilizan para la creación de las tablas a partir del modelo Entidad-Relación previamente diseñado y cada software de diseño contiene sus propias características en cuanto a la posibilidad de almacenar los datos.

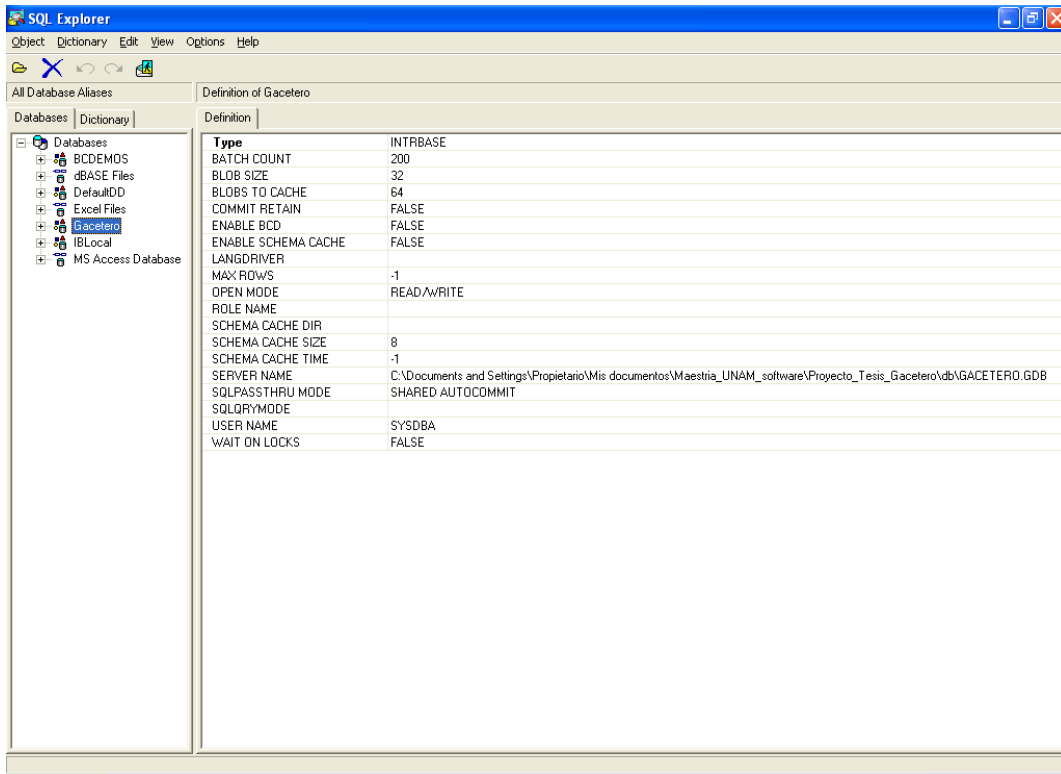


Figura 3.7 Interfaz del Administrador de la Base de Datos del Gacetero

De acuerdo al modelo entidad-relación de la figura 3.6, cada entidad es convertida en una tabla utilizando para ello una sentencia de SQL, por ejemplo, la siguiente sentencia crea una tabla llamada `g_address` junto con todos sus campos y sus tipos de datos:

```
CREATE TABLE g_address (
    address_id integer NOT NULL,
    feature_id integer NOT NULL,
    street_address character varying(80) NOT NULL,
    city character varying(80) NOT NULL,
    state_province character varying(80) NOT NULL,
    postal_code character varying(20),
    country character varying(20)
);
```

Esta instrucción en SQL permite que una vez que sea ejecutada, la base de datos posea un espacio dedicado al almacenamiento exclusivamente de direcciones o domicilios de las localidades; esto da un gran potencial ya que toda la información está separada de acuerdo al tipo o familia de datos. Por lo tanto, de las tablas que se han creado en la base de datos, cada una representa independientemente ciertas características de una localidad especial, la figura 3.6 denota todas estas características separadamente en tablas independientes entre sí, pero a su vez relacionadas unas con otras gracias a los campos que son comunes a un grupo en particular de tablas.

En la figura 3.8, se enlistan las tablas más representativas que conforman el gacetero digital, de acuerdo al programa utilizado.

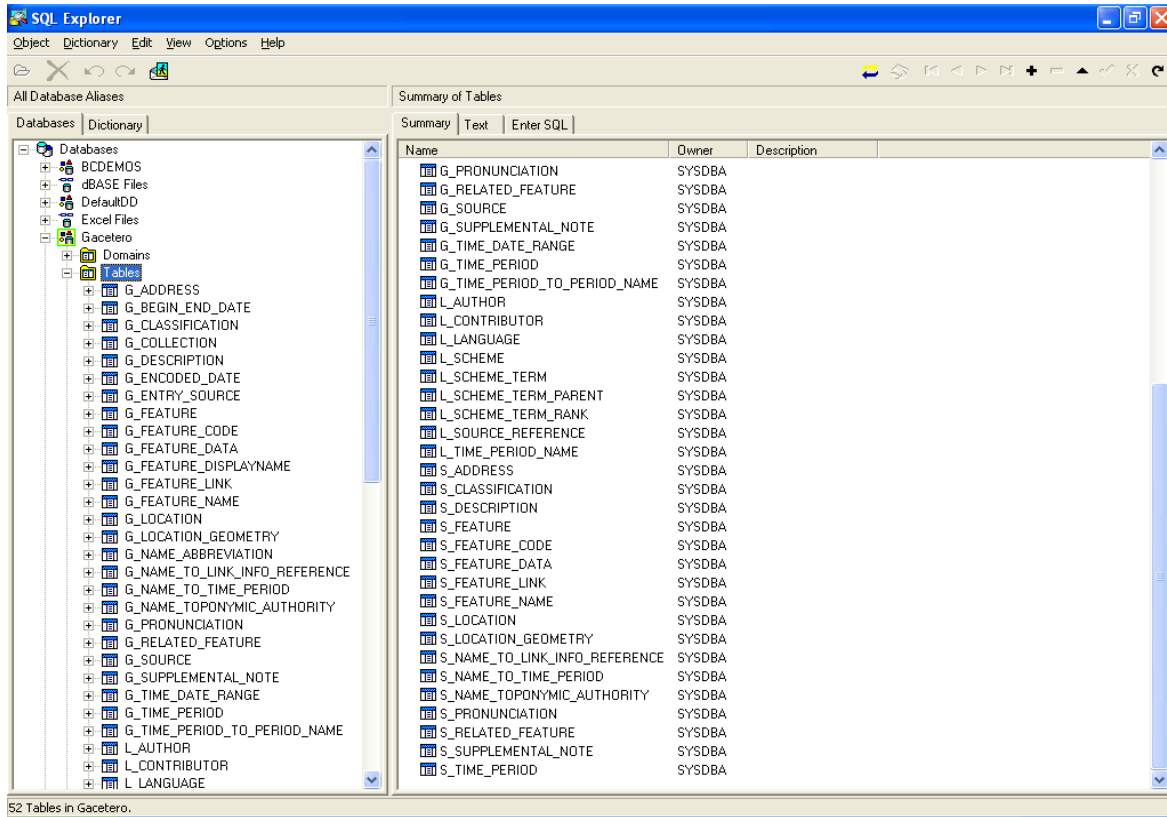


Figura 3.8 Tablas representativas del gacetero digital

Para mayor detalle del contenido de las tablas que se generaron en el gacetero digital, véase en el anexo B el listado completo de sentencias SQL que permitió la creación de la base de datos completa, en donde se especifican todos los nombres de los campos, su tamaño y los tipos de datos. Además, han sido separados por familias de datos, lo que ha permitido la creación de distintas tablas para tener un mejor control del sistema desde el punto de vista de desarrollo de sistemas de información.

3.3.3 Diseño de la Interfaz de acceso de datos

De acuerdo a las entrevistas con los usuarios finales del gacetero digital, se logró reunir una serie de datos que posteriormente fueron convertidos en campos, los cuales sirvieron de base para la creación de las tablas que se mencionaron en la sección anterior. Estas tablas conformaron la base de datos completa del gacetero digital, aun así, no es suficiente para cubrir las necesidades de los usuarios, ya que por si misma no es capaz de almacenar, recuperar y explotar los datos. Es necesario que además de almacenar la información, existan otros mecanismos que puedan ser controlados por los mismos usuarios con el fin de poder explotar y controlar la información que se tiene. Para poder contar con esta posibilidad de usar la información en el tiempo y la forma que el usuario desee, se procedió a diseñar y programar en un lenguaje moderno, de alto nivel y

orientado a objetos, un prototipo de interfaz que sea fácil de usar y que pueda ser vista por varios usuarios simultáneamente.

Una manera de presentar el diseño de la interfaz para acceder a los datos es poder ir comparando una serie de necesidades que fueron detectadas directamente desde los usuarios e ir verificando que satisfaga dichas necesidades.

En primer lugar, se desarrolló un sistema de bases de datos que contiene una ventana principal que sirve de entrada a todas las posibles actividades que puede realizar un usuario, a fin de obtener información proveniente de la base de datos desarrollada.

En la figura 3.9 se muestra la ventana principal, nótese que existen las llamadas barras de menú y de herramientas y la zona de trabajo para que al usuario le sea más fácil aprender su manejo. Por razones de simplicidad, en la figura solo se han dispuesto algunas de estas características principales.



Figura 3.9 Vista de la ventana principal del gacetero digital

De acuerdo al diseño previo del gacetero digital y de acuerdo al modelo Entidad-Relación presentado anteriormente, se deja claro que la entidad más importante es la de Rasgos, ya que con ella el usuario podrá controlar una enorme cantidad de datos inherentes al mismo, por este motivo el gacetero posee una ventana exclusivamente para mantener los datos de los diferentes Rasgos, tal y como se aprecia en la figura 3.10.

Figura 3.10 Ventana de captura de Rasgos

Como puede apreciarse en la figura 3.10, cada Rasgo posee una serie de atributos que lo distinguen de los demás, sin embargo, son muchos de ellos los que necesitan estar presentes en el gacetero, para lo cual, por cuestiones de facilidad en el manejo se han separado unos de otros tomando en cuenta la naturaleza de dichos atributos; por esta razón existe dentro de la ventana mostrada un acceso individual para capturar, respecto a un Rasgo, su descripción detallada, además de que existen varios niveles de detalle para hacer una descripción exacta, por lo que el usuario tiene la facilidad de capturar una lista completa de diferentes descripciones, véase la figura 3.11

Figura 3.11 Ventana para capturar los detalles de las descripciones de un rasgo

Obsérvese que es posible describir los esquemas y los tipos de términos adjudicados a un solo Rasgo.

En la figura 3.12 se ha preparado una interfaz para que el usuario pueda darle mantenimiento a los diferentes nombres con los que se conoce a determinado Rasgo.

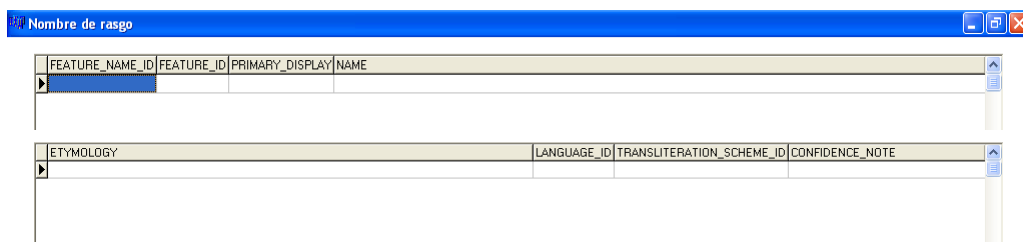


Figura 3.12 Interfaz para el control de nombres varios a un Rasgo

También cada Rasgo puede almacenar un sinfín de notas suplementales que se relacionan con el mismo, para lo cual se diseñó una interfaz como la que se muestra en la figura 3.13.

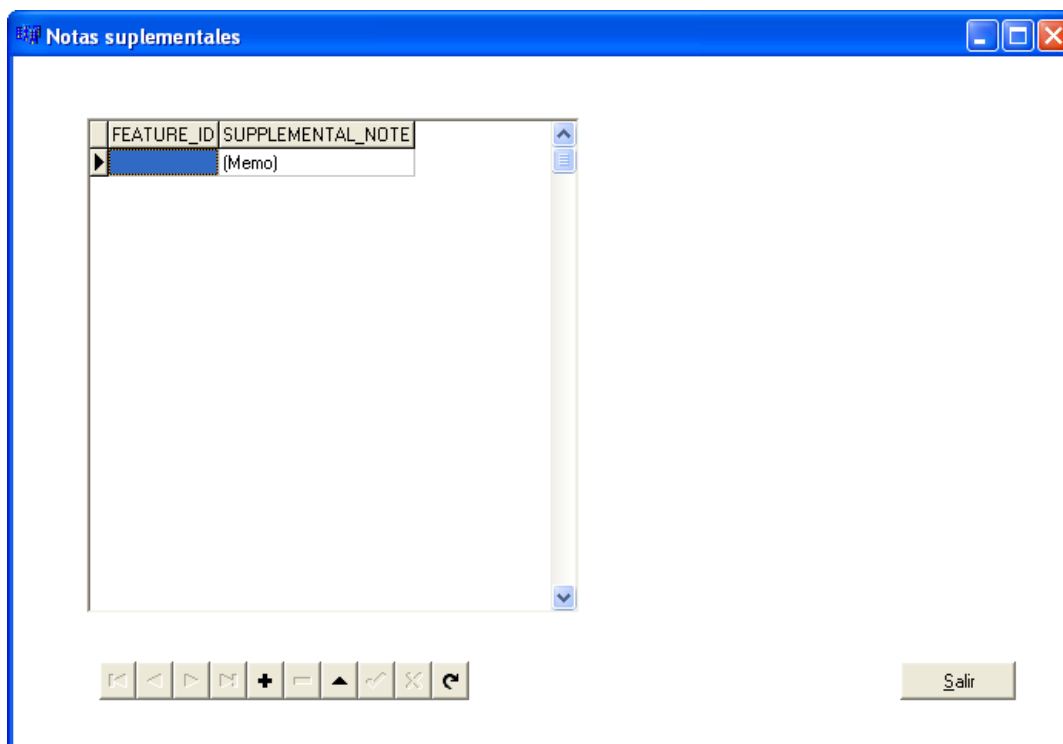


Figura 3.13 Interfaz para capturar una lista de notas suplementarias de un Rasgo.

También se ha creado una ventana especial para administrar la información referente a los datos más específicos de un Rasgo, tales como el tipo de esquemas, de término, etc., ver figura 3.14

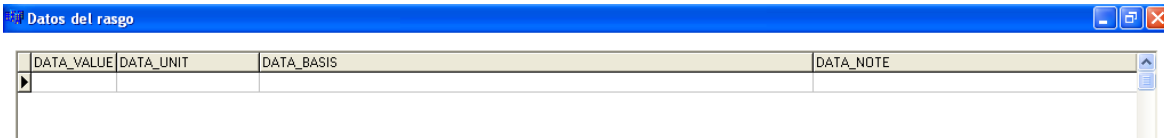


Figura 3.14 Interfaz para capturar Datos de los Rasgos

Una ventana especial donde se han dispuesto los campos para denotar las clasificaciones que tiene un Rasgo en particular; recuérdese que una clasificación es la asignación de un tipo de dato a un rasgo concreto, la cual se ubica dentro de una estructura jerárquica. Véase la figura 3.15

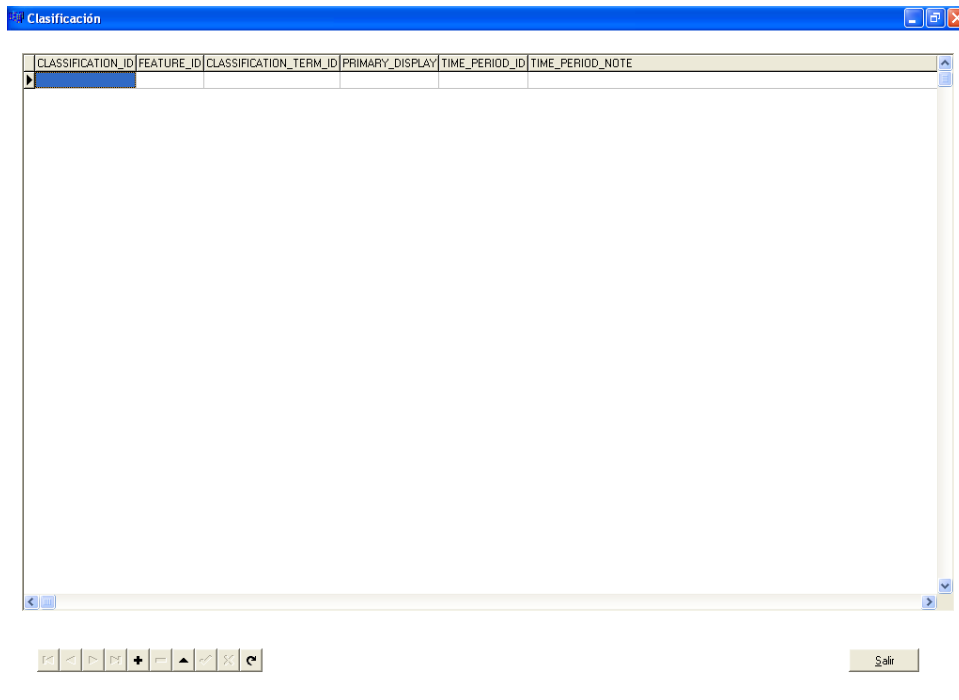
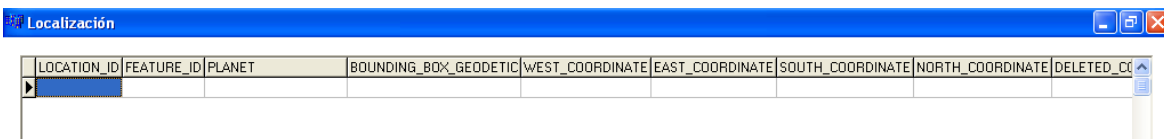


Figura 3.15 Ventana para capturar las diferentes clasificaciones de un Rasgo

La localización de un Rasgo es más específica, por lo que el gacetero posee una sección donde el usuario es capaz de contener información más específica o precisa del Rasgo en cuestión, para lo cual este tipo de información tiene que estar disponible cuando ésta exista y sin que tenga que revolveerse con otros datos, no precisamente de la misma naturaleza. En la figura 3.16 se destacan los campos más importantes donde el usuario controla estos datos.



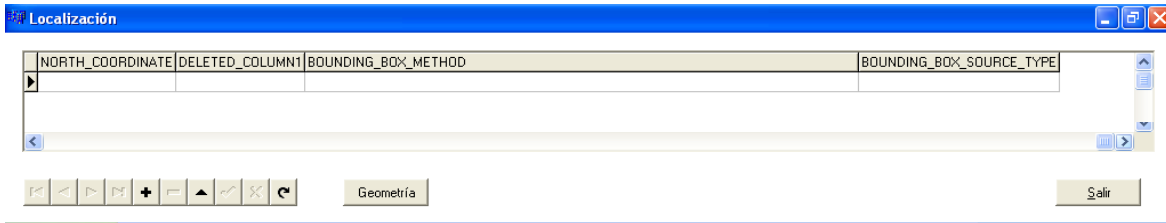


Figura 3.16 Ventanas para el almacenamiento de localizaciones de un Rasgo

De la misma manera, se diseñó una interfaz para capturar todos los datos relacionados a un rasgo específico, es decir, relaciones existentes entre los mismos rasgos contenidos en el gacetero que tengan que ver con uno de ellos en especial. En la figura 3.17 se muestra dicha interfaz.

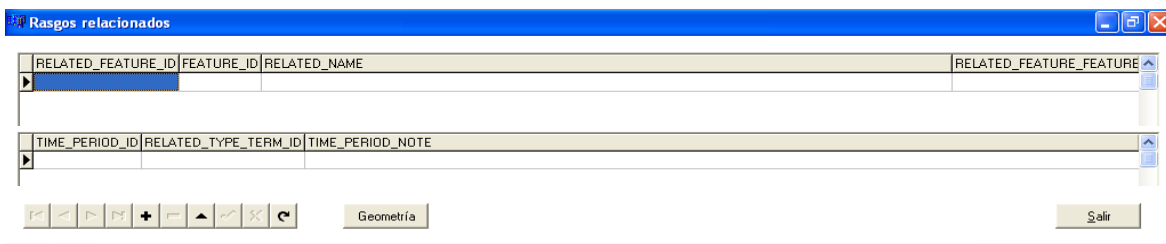


Figura 3.17 Interfaz para los rasgos relacionados a un Rasgo específico

Otro dato importante es que cada uno de los Rasgos, además de su clave interna, o sea, su llave, posee un código externo con el cual se conoce tradicionalmente. Por tanto, es posible que el usuario del gacetero digital lo pueda controlar, modificar, eliminar, aumentar, etc., sin que el gacetero se vea afectado dentro de sus tablas por alterar una llave primaria. En la figura 3.18 se muestra esta ventana.

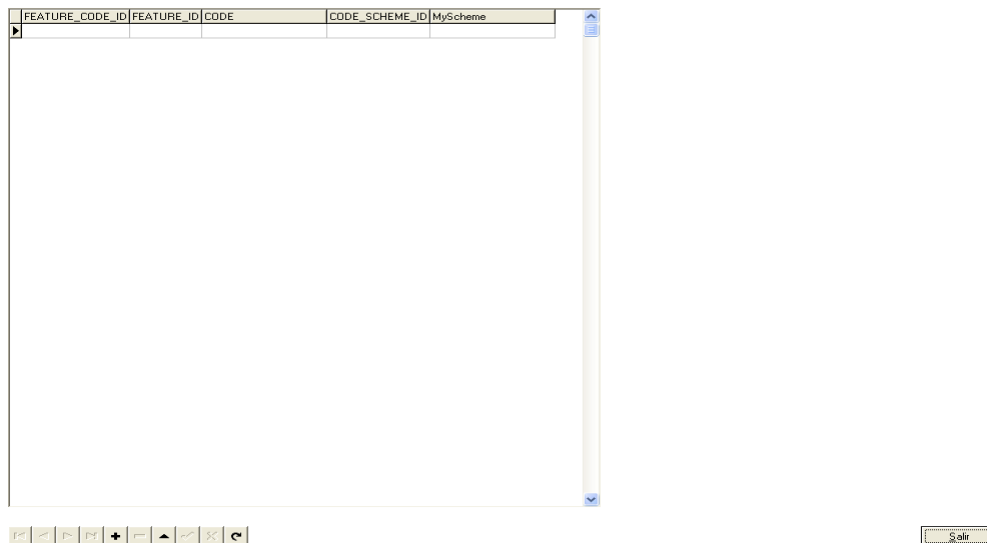


Figura 3.18 Interfaz por el control de códigos de rasgo.

Es muy importante aclarar que toda esta serie de ventanas utilizables por los mismos usuarios del gacetero, le permitirán conocer de una manera fácil y rápida la respuesta a las siguientes preguntas:

- ¿Existe la localidad con nombre x?
- ¿qué información existe acerca del rasgo x ?
- ¿Dónde está el rasgo x?
- ¿A qué categoría pertenece el rasgo x?
- ¿Con qué rasgo esta relacionado x?
- ¿Cuáles son los nombres asociados al rasgo x?
- ¿Cuál es el código externo del rasgo x?
- ¿Cuáles son los límites geográficos del rasgo x?
- ¿Cuál es el centro geográfico del rasgo x?
- ¿Qué atributos adicionales posee el rasgo x?

La respuesta a todas estas interrogantes se basa en la información que se puede almacenar en las distintas ventanas que conforman el gacetero digital, sin embargo, debe quedar claro que la información que alimenta al gacetero no siempre se introduce manualmente, ya que este contemplará otros mecanismos automáticos para la inserción de nuevos datos. Más adelante se explican algunos de estos mecanismos automáticos, pero por el momento solo resta detallar las interfaces más representativas de lo que es la extracción de datos basadas en consultas en línea o reporte predefinidos que pueden permitir la impresión de los datos en papel.

Implementación del Gacetero.

Esta etapa del proyecto consistió en revisar todos los trabajos realizados hasta la fecha, antes de iniciar su implementación: establecimiento de un equipo de cómputo para alojar los programas y la base de datos, configuración de cada computadora cliente donde se utilice el gacetero, así como cada uno de los procedimientos automáticos para almacenar la información inicial, tales como catálogos, índices de las tablas que conforman la base de datos, etc. Como se ha mencionado, no toda la información inicial debe ser alimentada al gacetero manualmente, es decir, debe haber ciertos mecanismos para que los usuarios no tengan que llenar los catálogos con los datos que ya se tienen de INEGI, cuya gran cantidad de datos normalmente no están estructurados de la misma manera como lo hace el gacetero propuesto. Esto denota que las estructuras de datos de INEGI y el gacetero son totalmente incompatibles; de hecho lo son, pero esto no significa que no se pueda tomar información existente de los sistemas del INEGI y convertirlos a través de procesos de transformación, a estructuras como las que contiene el propio gacetero.

Todos estos procesos de automatización deben efectuarse al menos la primera vez para evitar una captura manual masiva de los datos hacia el gacetero. Estos procedimientos no son parte de este trabajo, sin embargo, es importante aclarar que el gacetero puede empezarse a utilizar aún sin ellos; el requisito inicial sería que los usuarios deben capturar, por ejemplo, los catálogos mínimos para que el gacetero realice el resto de las tareas.

Con una captura inicial o automatizada, el gacetero posee una serie de ventanas con el mismo estilo de interfaz descrito en el capítulo anterior y en estas ventanas, también diseñadas para hacer consultas de los datos, se logra tener los catálogos básicos para iniciar su uso.

4.1 Catálogos del gacetero

Uno de los usos principales de los catálogos es poder asignarle a un conjunto de conceptos un código o clave, es decir, distinguir un concepto por medio de un código único para poder utilizarlo tantas veces como se haga referencia al concepto, en lugar de escribir el mismo número de veces la descripción del concepto, evitando así, la repetición innecesaria de datos dentro de las tablas que conforman la base de datos.

El gacetero cuenta con algunos catálogos en diferentes ventanas, de acuerdo a su naturaleza. Éstos son parte del mismo gacetero y son indispensables para que se produzcan los resultados esperados en las consultas de los usuarios. Además están dispuestos bajo un mismo rubro dentro de la interfaz, véase la figura 4.1.



Figura 4.1 Acceso a los catálogos desde la ventana inicial del gacetero digital

Los principales catálogos del gacetero son los siguientes:

- Esquemas
- Periodos de tiempo
- Colecciones
- Nombres de periodos de tiempo

En la figura 4.2 se muestran los campos específicos del Catálogo de Esquemas.

Figura 4.2 Catálogo de esquemas del gacetero digital

Estos serán utilizados en repetidas ocasiones dentro del gacetero, sobre todo relacionándolos con cada uno de los Rasgos y bastará con tener cada uno de ellos capturados solo una vez para poderlos relacionar muchas veces. En la figura 4.3 se muestra el Catálogo de Periodos de Tiempo, los cuales hacen referencia al conjunto de relaciones entre registros y eventos de tiempo específicos.

Figura 4.3 Catálogo de Periodos de Tiempo

Un catálogo más es el de Colecciones, el cual especifica cada uno de los bloques básicos en los que se agrupa la información proveniente de diversas fuentes, véase la figura 4.4.

Figura 4.4 Catálogo de Colecciones de Rasgos

En la figura 4.5 puede verse el Catálogo de Nombres de Periodos de Tiempo, el cual es una entrada que asigna un nombre concreto a un periodo de tiempo.

Figura 4.5 Catálogo de Nombres de Periodos de Tiempo

En todos estos catálogos se puede acceder a los datos bajo una misma política, ya que el gacetero completo tiene en sus ventanas las mismas capacidades de navegación; por ejemplo, botones de salida y de navegación entre registros.

4.2 Implementación del gacetero con datos de localidades

Después de haber desarrollado las interfaces para que los usuarios alimenten manualmente los datos de los catálogos, el siguiente paso en el desarrollo del proyecto es alimentar al gacetero con toda la información ya existente proveniente del INEGI, puesto que esta información además de ser muy extensa, es fácilmente accesible desde los mismos sistemas de información del INEGI, por tanto, sería una pérdida de tiempo y esfuerzo el capturarla manualmente dentro del gacetero.

Por este motivo debe existir un mecanismo que sea lo suficientemente ágil para traspasar los datos hacia el gacetero, principalmente los que se relacionan con las localidades, ya que estos datos son los más fáciles de controlar y su origen está perfectamente estudiado. Respecto a este origen de los datos, se conoce su estructura, su forma de ordenamiento, su longitud de campos, su tipo de datos y su número de registros, por lo que en realidad, solo hace falta un mecanismo que pase los datos de una estructura a otra.

En la siguiente sección de este documento se describe este proceso de traspaso de información, el cual está basado en algoritmos de programación puesto que se accede a bases de datos, tanto de los datos de origen como los de destino para posteriormente mostrarse en las interfaces diseñadas en el gacetero.

4.2.1 Importación de datos

Una consulta típica de localidades en el sitio oficial de INEGI proporciona normalmente información centralizada y estructurada de tal manera que el usuario en una sola ventana tiene todo el resumen de la localidad que desea referir. Pero internamente esta información es extraída de una base de datos con una estructura típica, la cual no es conocida por el usuario, desde el punto de vista de diseño. Recuérdese que la captura manual de datos provenientes del INEGI puede ser evitada si se vacían estos datos en forma automática en el nuevo gacetero.

Aquí surge la primera problemática a resolver, ya que si se quiere tener esta misma información, pero ordenada de otra manera, tal que sea compatible con la estructura de la base de datos del gacetero propuesto, es necesario conocer las dos estructuras simultáneamente para conseguir el traspaso lo más directo posible de una base de datos a otra. Por tanto, habrá que interpretar todos los datos obtenidos en la página del INEGI a partir del siguiente criterio e ir insertándola en el nuevo gacetero.

A continuación se describe el procedimiento de creación de registros nuevos en la estructura de la base de datos del gacetero que se menciona en la Figura 3.6. Este procedimiento refiere únicamente los campos esenciales en el proceso de migración de los datos y además debe ser repetido tantas veces como registros se tengan en la información del INEGI:

1. Se toma y se examina una línea completa de los datos del INEGI y se particiona en segmentos independientes para crear registros en varias tablas de la nueva base de datos.
2. Se obtiene y se define la colección, es decir, un bloque básico en el que se debe agrupar la información del punto anterior y se abre un nuevo registro en la tabla *g_collection*, según el modelo Entidad_Relación propuesto en la Figura 3.6, del capítulo anterior. Este nuevo registro posee un identificador asignado por la computadora para distinguir a la colección como única entre las demás, creándose un registro completo con estos campos: *collection_id*, *name* (nombre de la colección) y *note* (nota acerca de la misma).
3. A continuación se crea un registro en la tabla *g_time_period*, para lo cual se especifican sus campos *time_period_id* automáticamente y *status_term_id*, los cuales identifican un nuevo periodo de tiempo activo para un rasgo nuevo. Simultáneamente se crea un registro en la tabla *g_time_period_to_period_name*, *l_time_period_name*, *g_encoded_date*, *g_begin_end_date* y *g_time_date_range*, quienes controlan el nombre del nuevo periodo de tiempo, el tipo de fecha guardada, sus fechas de inicio y fin, y un intervalo de tiempo probable.
4. En este paso se define un identificador único en la tabla *g_feature*, también asignado automáticamente por la computadora, y los campos *collection_id* y *time_period_id* para establecer la relación con la colección y periodo de tiempo correspondientes. La tabla *g_feature* es muy importante, ya que de ella se obtiene la mayor parte de las relaciones existentes dentro de la base de datos y por tanto, la posibilidad de extraer la información de forma adecuada. Sus campos más importantes permiten saber cuáles son las fechas de creación de este rasgo, de la última modificación y una nota breve acerca de este rasgo y su estatus para saber si la información está completa o no.
5. Posteriormente a la creación de un nuevo rasgo, se crea un registro en la tabla de *g_description* para establecer el nombre y tipo de esquema, la descripción del tipo de término y una descripción corta del nuevo rasgo.
6. La siguiente tabla de importancia en este proceso es la llamada *g_feature_name*, la cual representa un potencial de búsqueda a los usuarios, ya que contiene algunos campos tales como *primary_display* que se usa para desplegar el nombre primario con el que se conoce el rasgo, otro llamado *name* proporciona el nombre tradicional con que se identifica en la localidad, el campo *etymology* es una derivación del nombre y el llamado *confidence_note* que almacena una expresión de confianza asociada con la forma del nombre del rasgo; por ejemplo el término “citing” que es el nombre que fue tomado de una fuente arqueológica que probablemente fue leída parcialmente.
7. Después de esto cada rasgo queda perfectamente definido a través de un registro en una tabla llamada *g_location*, quien se encarga de almacenar las coordenadas oeste, este, sur y norte del rasgo en cuestión. Dichas coordenadas no son otra cosa que los valores de longitud y latitud, es decir, los valores positivos o negativos respecto al meridiano de origen y el ecuador de cada uno de los rasgos almacenados dentro de la tabla *g_feature*.

8. Por último, y no por eso menos importantes, se crean los registros correspondientes en las tablas *g_supplemental_note*, quienes se encargan de almacenar las notas adicionales del rasgo; *g_feature_data*, almacena la categoría que describe al tipo de rasgo, por ejemplo, población, elevación, área, etc., la unidad de medida, por ejemplo pies o metros, la base de referencia del dato, por ejemplo, “sobre el nivel del mar”; *g_location_geometry*, guarda una oración acerca de la confidencialidad de la localidad, por ejemplo, una evidencia acerca de la localidad; *l_scheme*, URL’s relacionadas al rasgo correspondiente y un esquema particular y *g_related_feature*, rasgos que tienen algún tipo de relación con el rasgo principal.

Todos los pasos anteriores, construyen al final, una serie de relaciones, de tal manera que la información se encuentre distribuida de acuerdo a su naturaleza, dentro de la base de datos del nuevo gacetero para que al momento de su extracción se usen procedimientos más precisos y por tanto, más confiables.

En la siguiente sección se detallan los procesos más importantes en la extracción de los datos, para permitir al usuario cumplir con sus expectativas de información.

4.3 Consultas y reportes principales

Como se ha mencionado anteriormente en este mismo documento, la introducción de información a la base de datos del gacetero digital, está basada en procesos de captura manual, o bien, en procesos automáticos tomando los datos provenientes de INEGI. Estos deben introducirse adecuadamente en las diferentes tablas que conforman dicha base de datos; todo depende del método y cantidad de información a introducirse.

Una vez que los datos han sido depositados en el gacetero propuesto, el proceso de consulta resulta diferente ya que tienen que estar dispuestos en cierto modo para que puedan ser interpretados por el usuario ayudándolo a la toma de decisiones.

Pero la ventaja de este gacetero es que la mayoría de las ventanas que se utilizan para la introducción de los datos, sirven también para la consulta, puesto que en todas ellas siempre está presente un mecanismo para navegar por todos los registros. Por esta razón se verá una serie de íconos en todas las ventanas, como los que se muestran en la figura 3.19 para poder navegar por los registros.



Figura 3.19 Botones de navegación de las interfaces del gacetero digital

Las interfaces que se han desarrollado para la extracción de datos desde las tablas, ocupan internamente en todos los casos, el lenguaje de consulta estructurado (SQL), y toda esta serie de consultas están ocultas para el usuario final. Exclusivamente el personal de desarrollo de sistemas ocupará este tipo de herramientas, mientras que el usuario final solamente le interesará resolver sus requerimientos de información; todos los resultados obtenidos están diseñados para ser mostrados, ya sea en una ventana en el monitor de la computadora, o bien, por medio de reportes escritos.

El potencial del nuevo gacetero está basado en responder las siguientes preguntas:

- ¿Existe la localidad con nombre x?, esta pregunta es primordial puesto que la misma definición del gacetero implica que los registros son accesibles por medio de su nombre y para responder a esta pregunta solo se requiere consultar el listado correspondiente que el gacetero almacena. La instrucción SQL respectiva es:

```
Select g_feature.feature_id, name
From g_feature, g_feature_name
Where g_feature.feature_id = g_feature_name.feature_id and name like
'x%
```

Nótese que se imprimirá el identificador del rasgo, el nombre primario y el segundo nombre de la localidad, siempre que exista un nombre parecido al propuesto por el usuario, en este caso llamado x.

- ¿qué información existe acerca del rasgo x ?, esta pregunta regresa la información que se encuentre disponible para el rasgo en cuestión, como puede ser la colección a la que pertenece, datos asociados al rasgo, nombres adicionales, relaciones con otros rasgos, entre otros. Por ejemplo, para enlistar el conjunto de datos asociados al rasgo, así como los esquemas asociados a cada dato, se emplea la siguiente consulta que requiere de buscar en las tablas g_feature, g_feature_data y l_scheme:

```
select *
from g_feature_data , g_feature , l_scheme
where g_feature.feature_id = 22222 AND
      g_feature_data.feature_id = g_feature.feature_id AND
      g_feature_data.data_type_scheme_id =
      l_scheme.scheme_id
```

En este caso, nótese que todas las tablas donde se quiere obtener información tienen un campo en común llamado *feature_id* cuyo valor debe ser igual en todas las tablas involucradas, pero a su vez, el nombre del rasgo debe coincidir con el propuesto por el usuario.

- ¿Dónde está el rasgo x?, independientemente del nivel de detalle de la respuesta, es necesario conocer las coordenadas extremas del rasgo solicitado. La consulta SQL se puede ejecutar de la siguiente manera:

```
Select west_coordinate, east_coordinate, south_coordinate,
north_coordinate
From g_location, g_feature, g_feature_name
Where g_location.feature_id = g_feature.feature_id and
g_feature_name.feature_id = g_location.feature_id and primary_display like
'x%
```

- ¿A qué categoría pertenece el rasgo x?, esta pregunta debe entregar todas las categorías con las cuales está clasificado el rasgo x. La tabla *g_classification*,

junto con *l_scheme* y *l_scheme_term* poseen toda la información necesaria para responder la pregunta. La instrucción SQL es la siguiente:

```
Select l_scheme_term.term, l_scheme.scheme_name, term_id
From g_classification, l_scheme, l_scheme_term
Where g_classification.classification_term_id = l_scheme_term.term_id and
l_scheme_term.scheme_id = l_scheme.scheme_id
```

Obsérvese también que el rasgo de nombre *x* dado por el usuario, es el principal dato de comparación.

- ¿Con qué rasgos esta relacionado *x*?, esta pregunta entrega el conjunto de rasgos sobre los cuales se ha definido una relación cuyo significado es ajeno al alcance del modelo y ha sido definido de manera externa por el usuario. Cada relación definida dentro de la tabla *g_related_feature* está tipificada, asignándole un significado a dicha asociación entre rasgos. La consulta SQL siguiente responderá a la pregunta:

```
Select *
From g_feature, g_related_feature
Where g_feature.feature_id = g_related_feature.feature_id and
g_feature_name.primary_display like '%x%'
```

La selección de campos puede particularizarse de acuerdo a las necesidades a mostrar.

- ¿Cuáles son los nombres asociados al rasgo *x*?, dado que un rasgo puede ser identificado por más de un nombre, es conveniente saber todos los nombres que lo designan. La consulta SQL correspondiente es:

```
Select g_feature.feature_id, g_feature_name.*
From g_feature, g_feature_name
Where g_feature.feature_id = g_feature_name.feature_id and and
g_feature_name.primary_display like '%x%'
```

- ¿Cuál es el código externo del rasgo *x*?, esta pregunta es importante ya que este código relaciona al registro dentro del gacetero con su correspondiente externo y puede ser respondida con una consulta SQL de lo más simple, por ejemplo:

```
Select external_scheme_term_id, feature_id, name
From g_descripcion, g_feature, g_feature_name
Where g_descripcion.feature_id = g_feature.feature_id and
g_feature.feature_id = g_feature_name.feature_id and
g_feature_name.primary_display like '%x%'
```

Como puede observarse, siempre será necesario establecer una relación entre las tablas involucradas y efectuar una comparación con el nombre del rasgo *x* solicitado.

Existen muchas variantes y combinaciones de consultas y cada una de ellas usará una combinación de tablas diferente, por lo que las posibilidades de encontrar la respuesta a cada interrogante se hacen más grandes, a medida que se construyan consultas SQL cada vez más sofisticadas y no por ello imposibles.

Es muy importante destacar en este momento, que todas las preguntas anteriores si han podido ser respondidas perfectamente usando la estructura de base de datos del nuevo gacetero propuesto. En otras palabras, la respuesta a estas preguntas básicas, las cuales definen la funcionalidad del gacetero, muestran la pertinencia del modelo propuesto para implementar un gacetero digital empleando un modelo relacional.

4.4 Exportación de datos

Ya se han analizado por separado, tanto los principales procesos de entrada de datos al nuevo gacetero, como de extracción y explotación de éstos, a través de reportes en pantalla y en forma escrita, utilizando el lenguaje de consulta estructurada SQL. Aún así, existe otra forma muy importante de extraer la información almacenada en el nuevo gacetero, esta consiste en generar archivos llamados planos, los cuales, están constituidos de cadenas de caracteres, todos ellos legibles por cualquier editor de caracteres, pero con la peculiaridad de que están ordenados de acuerdo a los campos de las tablas de donde fueron extraídos por medio de SQL. Estos campos a su vez, están separados, ya sea por dos puntos (:), punto y coma (;), por tabuladores o cualquier carácter definido previamente como separador de campos.

Estos archivos planos pueden ser generados a partir de la petición de algún usuario y almacenados en algún sitio dentro de los archivos propiedad del usuario, y una vez que se han generado, podrán ser leídos, ya sea por una hoja de cálculo, un procesador de textos y especialmente por una base de datos distinta a la que generó la información del archivo plano. Todo esto con la finalidad de volver compatibles a los datos en relación a otros fabricantes de software para su adecuada manipulación y explotación a petición de otros nuevos usuarios que requieran información del gacetero.

El proceso de traspaso de datos a archivos planos está perfectamente controlado de acuerdo a los resultados de haber ejecutado consultas SQL, pero como es bien sabido, este lenguaje normalmente no está disponible para la mayoría de los usuarios, entonces se vuelve necesaria la existencia de un mecanismo previamente desarrollado para que solo se ejecute dicho proceso en forma transparente y se obtengan los resultados esperados. El proceso de generación de archivos planos consiste básicamente en ejecutar consultas SQL que rescaten la información de varias tablas y depositarlas en una sola estructura previa al archivo plano. Esta estructura puede ser una tabla temporal llamada vista, que contiene el resultado de la consulta SQL y una vez procesada la información, la vista es desechada a fin de no ocupar espacio en el disco duro de la computadora donde se esté ejecutando.

Por tanto, cualquier consulta SQL puede convertirse en una vista y de esta forma hacer que los usuarios solo conozcan el conjunto de vistas disponibles y sus resultados, sin necesidad de entender las instrucciones. Por ejemplo, para la pregunta de ¿Con qué rasgo está relacionado x?, recordemos que su respectiva consulta SQL que generaba la

información que respondiera a esta necesidad, estaba dada por la instrucción SQL siguiente:

```
Select *
From g_feature, g_related_feature
Where g_feature.feature_id = g_related_feature.feature_id and
g_feature_name.primary_display like '%x%
```

Entonces, una vista que pueda ser utilizada subsecuentemente llamada "rasgos_relacionados" se crea con la siguiente instrucción SQL:

```
Create view rasgos_relacionados as
Select *
From g_feature, g_related_feature
Where g_feature.feature_id = g_related_feature.feature_id and
g_feature_name.primary_display like '%x%
```

Esta vista, es publicada al usuario directamente, o también puede ser insertada en alguna interfaz de usuario y hacer que el resultado vaya directamente a una tabla temporal y a continuación a un archivo plano.

Todo lo anterior indica que una buena manera de exportar datos que sean leídos por otro tipo de software con otros formatos diferentes, es la de crear archivos planos a partir de vistas.

4.5 Utilización adecuada de índices

Ya se ha hablado mucho de la explotación de la información y se ha mencionado que el usuario puede disponer de ésta cuando así lo desee, pero también está presente la necesidad de:

- Realizar búsquedas más rápidas, sobre todo cuando el número de registros es muy elevado y el tiempo de rastreo puede hacerse muy grande.
- Verificar que la información que se han dado de alta en la base de datos no esté duplicada al momento de insertar un nuevo registro.
- Presentar la información, ya sea en el monitor de la computadora, en reportes o en archivos planos, ordenada bajo un criterio específico para su mejor lectura o comprensión.

En cualquiera de los tres casos anteriores, es necesario el uso de índices para cumplir adecuadamente con estos mismos requerimientos. De esta manera queda ratificada la necesidad de que absolutamente todas las tablas que conforman la base de datos del gacetero propuesto deben contener al menos un índice que coadyuve a las tres actividades anteriores.

Las tablas del gacetero propuesto deben verse como estructuras de datos formales. Éstas deben optimizar la velocidad de las transacciones que se realizan, por ejemplo, agilizando el acceso a los registros. Sin embargo, el incremento en la velocidad de las transacciones no sería posible si no existieran los índices, sobre todo en campos que se usen más frecuentemente para las búsquedas, ya que estos índices permiten el control de la posición de un registro dentro de la base de datos, accedando los mismos registros aleatoriamente o a través de un modo ordenado y más eficiente.

Los índices normalmente se construyen con base a algoritmos de árboles B o B+, también haciendo una mezcla de ellos. Estos algoritmos tienen como objetivo ordenar datos y controlar las eliminaciones e inserciones de los mismos. Para el caso del gacetero propuesto, los índices creados en los diferentes campos que conforman las tablas, no están exentos de la aplicación de estos algoritmos y en gran parte es conveniente que así sea, ya que el gacetero debe cumplir con cierta capacidad de búsqueda de información.

El gacetero propuesto debe cumplir con ciertos requisitos para ser considerado como un buen buscador de información. Las búsquedas de los datos dentro de las tablas serán su principal función, para ello el uso de los índices será de gran ayuda en el cumplimiento de esta tarea.

Para lograr esto, técnicamente hablando, es necesario que al momento de la construcción de las tablas se especifiquen los campos llaves y los índices correspondientes. Por esta razón, se han implementado índices en la mayoría de las tablas, esto debido a las ventajas que puede ofrecer al momento de realizar las búsquedas, pero se debe resaltar que no necesariamente las tablas deben tener índices creados aún cuando se tengan campos llave, puesto que las llaves solo son necesarias en ciertas tablas clave. En consecuencia, las tablas más importantes en donde es requerido un índice independientemente de que sus campos sean o no llaves son los siguientes:

- `g_feature_name`, cuyo campo para crear un índice es *name*. La razón de crear un índice basado en este campo es porque la mayoría de las consultas están apoyadas en el nombre del rasgo; este es el principal potencial del gacetero puesto que lo que al usuario le va interesar mucho más, es la búsqueda por nombre de manera rápida.
- `g_feature_code`, en este caso el campo usado para crear el índice es *code*, ya que este campo en el gacetero propuesto contiene información única e irrepetible que ha sido extraída con base en algoritmos especiales desde los datos de origen y al ser vaciados en esta tabla pueden ocuparse como un criterio de búsqueda apropiado.

Cuando se diseña un sistema de cómputo, especialmente en el nuevo gacetero, se deben tomar en cuenta los modos de ordenamiento de los datos, cuando éstos sean presentados en las diferentes ventanas o interfaces que conforman dicho gacetero. Particularmente, se han tomado en cuenta las necesidades de los usuarios para que de acuerdo a su petición, los datos sean ordenados bajo criterios variados, por ejemplo, podría ser que un conjunto de Rasgos en un reporte solicitado, deba enlistarse alfabéticamente por el nombre del rasgo, o bien enlistarse por su fecha de creación, para esto deben haberse creado internamente los índices en cada una de las tablas existentes.

Tomando en cuenta lo anterior, al momento de crear las tablas de la base de datos deben especificarse por medio del mismo Lenguaje Estructurado de Consulta (SQL), los campos que se utilizarán para ordenar la información. A manera de ejemplo: en el caso de la tabla de notas suplementarias se utilizó la instrucción siguiente de SQL, tanto para crearla, como para especificar los campos que estrictamente deben ser capturados dentro de la base de datos y que seguramente estos campos son los que se usarán como criterios de ordenamiento.

```
CREATE TABLE g_supplemental_note (  
    feature_id integer NOT NULL,  
    supplemental_note character varying(1000) NOT NULL  
);
```

Obsérvese que ambos campos han sido declarados como NOT NULL, ya que si no se ingresan datos en esos campos, un posible ordenamiento podría fallar con la ausencia de estos mismos datos, aún con los índices en uso. En consecuencia, si se utiliza esta filosofía de creación de tablas, será muy sencillo tener acceso a los nombres de rasgos, a las categorías y a sus registros geométricos, ya que cada una de estas características deben ser accedidas especialmente con la ayuda de índices. Ahora bien, para crear un índice llamado “supplementalNotes_ndx” que ordene las notas suplementarias de los rasgos por el criterio de feature_id se construye con una instrucción SQL como la siguiente:

```
Create unique index supplementalNotes_ndx on g_supplemental_note (feature_id)
```

Y está por demás volver a mencionar que todas las tablas, al menos deben tener un índice que pueda ser utilizado para resolver cualquiera de las tres problemáticas ya mencionadas al inicio de esta sección.

Resultados.

En el presente capítulo se hace una descripción detallada de los resultados obtenidos a partir de los procesos de análisis, diseño y desarrollo del gacetero propuesto. Estos procesos han permitido llevar a cabo la creación de las estructuras de bases de datos, interfaces y procesos especiales para el llenado de las tablas a partir de los datos generados por el INEGI.

El alcance del gacetero digital que ha sido propuesto se ha cumplido al pie de la letra, como puede observarse en la figura 5.1, se denota dicho alcance, el cual contempla principalmente el diseño de la base de datos basado en el modelo ADL, el diseño de la interfaz de usuario que a su vez activa varios procesos como pueden ser la gestión de altas, bajas, modificaciones y consultas de registros de localidades.

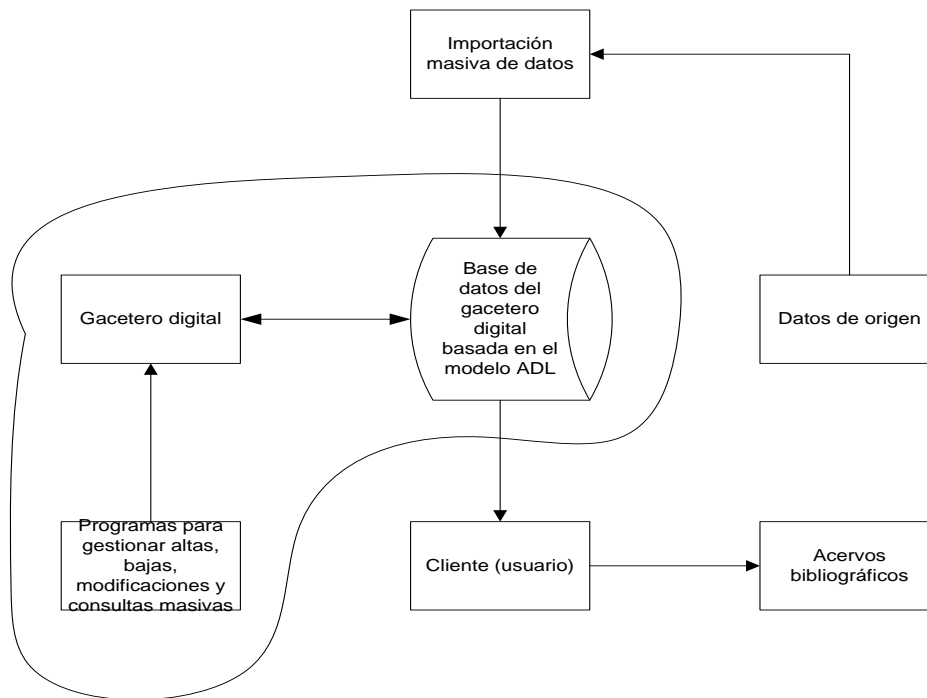


Figura 5.1 Alcance del gacetero propuesto

Por otro lado, se tienen contemplados otros procesos importantes tales como la importación masiva de datos provenientes del INEGI y los acervos bibliográficos. Estos

procesos en conjunto con los principales procesos del gacetero digital darán a los usuarios un gran potencial para poder explotar los datos de localidades. Los resultados obtenidos pueden ser catalogados en los siguientes rubros:

5.1 Metodologías resultantes

Respecto a los resultados en metodologías nuevas que los nuevos usuarios poseen o pueden adquirir paulatinamente, se destaca que la gran mayoría de la información de fuentes externas pudo ser resguardada en otro formato, además de que la aplicación que se desarrolló para el traspaso es capaz de realizar las siguientes funciones:

- Recolección de datos provenientes de las tablas del INEGI.
- Discriminación de datos en dos grandes rubros: importables y no importables. Para el caso de los primeros, son útiles según los nuevos criterios del gacetero, por ejemplo, nombre, clase y ubicación.
- Descomposición de esta estructura en los rubros espaciales, temporales, clasificación, etc., y separación de datos para ser incrustados en las nuevas tablas del gacetero.
- Almacenamiento de datos en tablas creadas con Postgres SQL que cumplen con las reglas de la normalización de tablas en una base de datos.
- Ejecución de respaldos y traspaso de datos a tablas de un nuevo gacetero.

Uno de los resultados más importantes en el desarrollo del nuevo gacetero fue el traspaso de datos hacia las nuevas tablas que conforman la base de datos, debido a que la información que se encontraba en un formato de una sola tabla se le ha dado una estructura en la cual se agrupa en categorías diversas; esto es importante porque le asigna un significado común a los elementos heterogéneos. Se le ha adjudicado una categoría a toda la información y también se ha estandarizado la manera de acceder a esta nueva estructura de información heterogénea.

Como parte del resultado de los datos reales de las colectas biológicas, se crearon millones de registros y se pudo demostrar que el software de diseño y desarrollo del nuevo gacetero cumplían con las especificaciones estándares para el manejo de los datos. Los servidores que contienen el gacetero digital y su base de datos se localiza físicamente en el Instituto de Geografía de la UNAM.

5.2 Directrices

El gacetero resultante coadyuva en el control y almacenamiento de información de localidades que permitan más adelante convertirse en información georeferenciada, ya

que como herramienta digital agiliza el acceso a datos espaciales, temporales y de otros tipos, evitando actividades laboriosas y propensas a errores.

Debido a que el gacetero está basado en el modelo ADL, sus características lo clasifican como muy flexible, de uso estándar y apropiado para sistemas de este tipo, entonces es muy fácil que la explotación de los datos sea más eficiente, que se tenga acceso a datos históricos, que la información esté categorizada y en general se evite la pérdida de información, en el momento de ser requerida.

El uso del nuevo gacetero también formaliza el procedimiento para que los usuarios satisfagan sus necesidades de información, de tal manera que con una simple consulta se obtenga un nombre geográfico, uno o más componentes prospectos a ser georeferenciados, un tipo de dato o clasificación dentro de un vocabulario controlado, información extra relacionando el rasgo, cero o más relaciones con otros rasgos y una serie de esquemas para catalogar la información contenida en el resultado de la consulta.

Es muy importante resaltar que la puesta en marcha del nuevo gacetero implica todo un plan de implantación, mismo que al momento de escribir esta tesis está en proceso de ser implantado, sin embargo ya se están haciendo pruebas con el millón y medio de registros que han sido importados del INEGI y se estima que su uso completo está en función de la liberación de todos los procesos que se mencionan en la figura 5.1 los cuales tomarán un poco más de tiempo.

5.3 Programación

Un resultado más tangible es la creación de una aplicación adaptada a las necesidades del usuario que incluye interfaces para el acceso, edición y manipulación de registros dentro de un gacetero digital, basado en el estándar ADL. La creación de una aplicación así implica por lo tanto, la ejecución de varios mecanismos para que su contraparte, en un servidor atienda las peticiones de los usuarios, como pueden ser las consultas de datos albergados en el gacetero y explotar la información eficientemente, permitiendo además cruzarla en el orden que el usuario lo disponga.

La figura 5.3 muestra un detalle de la interfaz resultante de este desarrollo, mismo que está siendo utilizado en el Instituto de Geografía de la UNAM en su carácter de prototipo, ya que se han cargado suficientes registros para iniciar su operación.

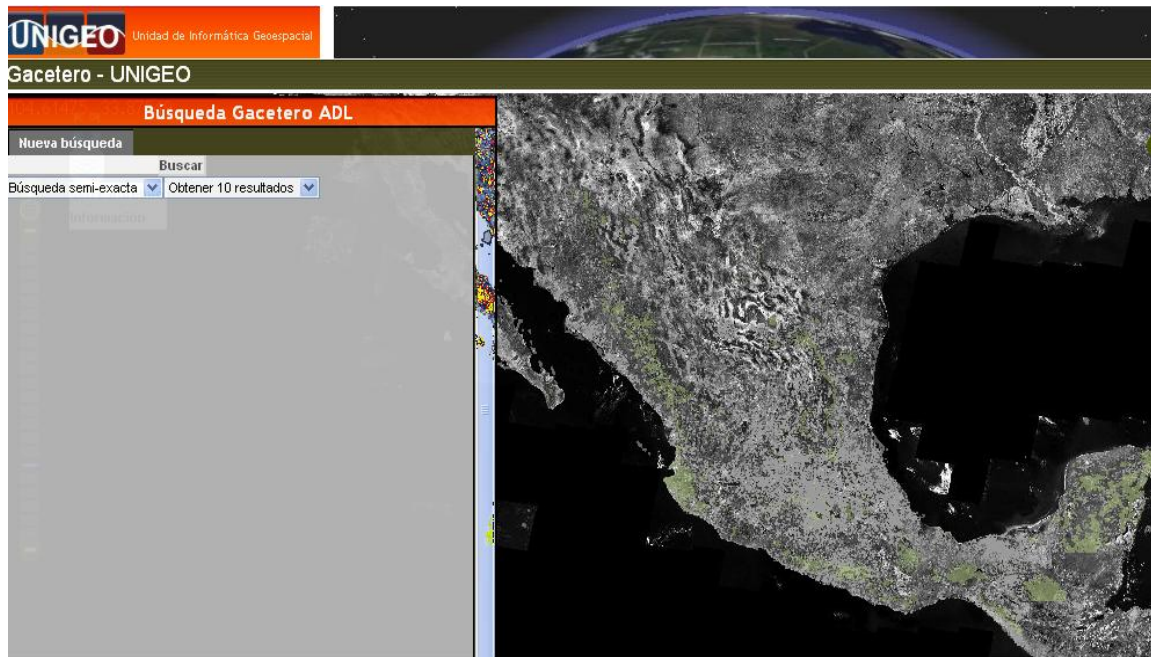


Figura 5.3 Ventana inicial del gacetero digital desarrollado
(Fuente: <http://datum.unigeo.igeograf.unam.mx:57212/gacetero/Gacetero.html>
recuperado el 12 de junio de 2010)

Como parte de la descripción de los resultados obtenidos, es indispensable mostrar el funcionamiento de las herramientas y procesos, producto del diseño de un nuevo gacetero. Recordemos que cronológicamente en la fase de análisis y diseño del nuevo gacetero, se obtuvo un modelo entidad-relación basado en estándar ADL, y a su vez, se pudo determinar la cantidad de campos a utilizar, sus tipos de datos, sus longitudes y sus nombres; pero lo más importante de esto, es que se crearon las tablas definitivas que conforman la nueva base de datos.

El gacetero digital posee cierta funcionalidad para que el usuario determine la manera en que desea visualizar los resultados. En la figura 5.4 se observa la manera de habilitar y deshabilitar el modo de visualización: mostrar carreteras, cuerpos de agua, estados, municipios, islas, etc., etc., etc.

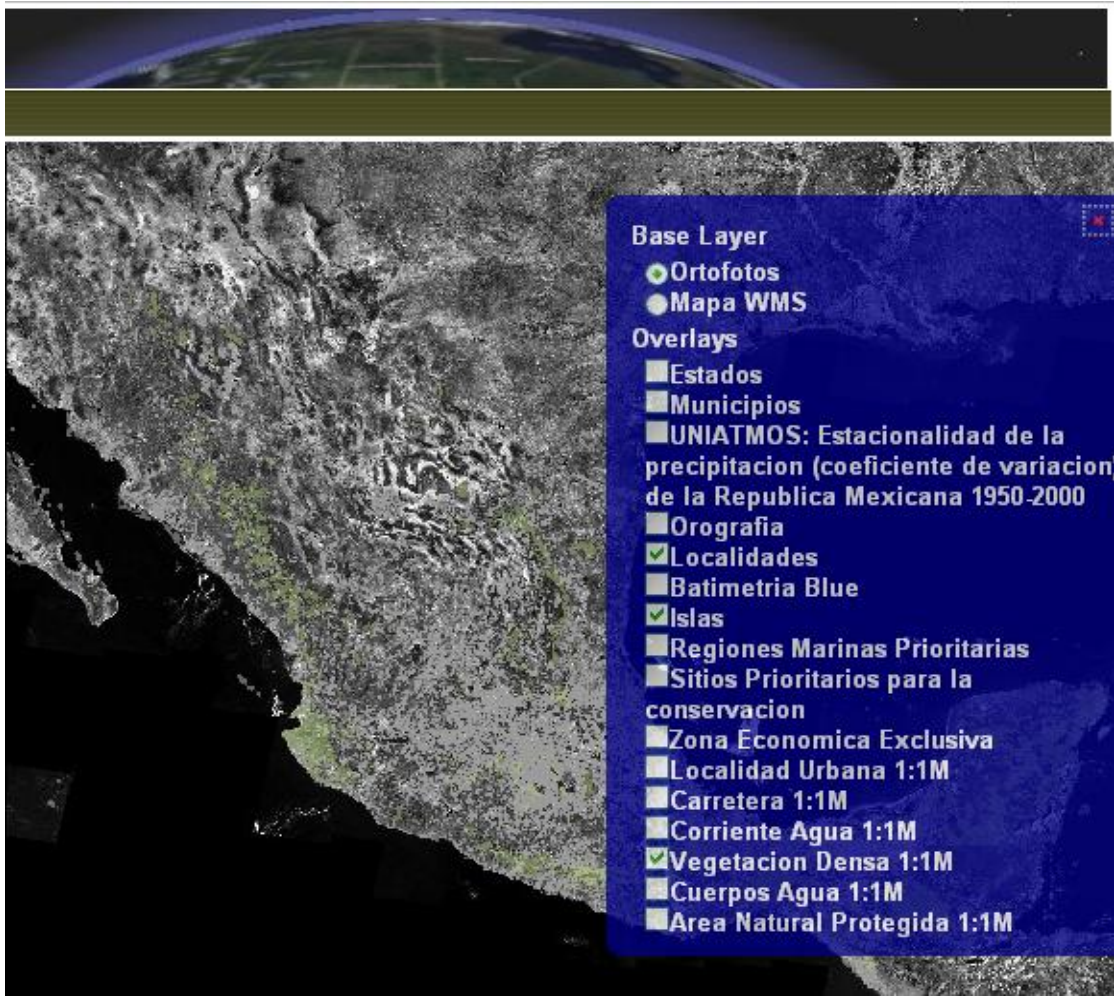


Figura 5.4 Configuración de resultados
 (Fuente: <http://datum.unigeo.igeograf.unam.mx:57212/gacetero/Gacetero.html>
 recuperado el 12 de junio de 2010)

Una vez que el usuario ha decidido la manera de visualizar el mapa, también puede elegir los criterios de búsqueda de localidades, por ejemplo, en la figura 5.5 se observan algunos de estos criterios.



Figura 5.5 Criterios de búsqueda de localidades del gacetero propuesto

Cuando el usuario determina la localidad a buscar, entonces podrá administrarle al gacetero digital su petición capturando el nombre en el campo correspondiente, de esta forma se generan los resultados, tanto gráficamente, como en modo texto.

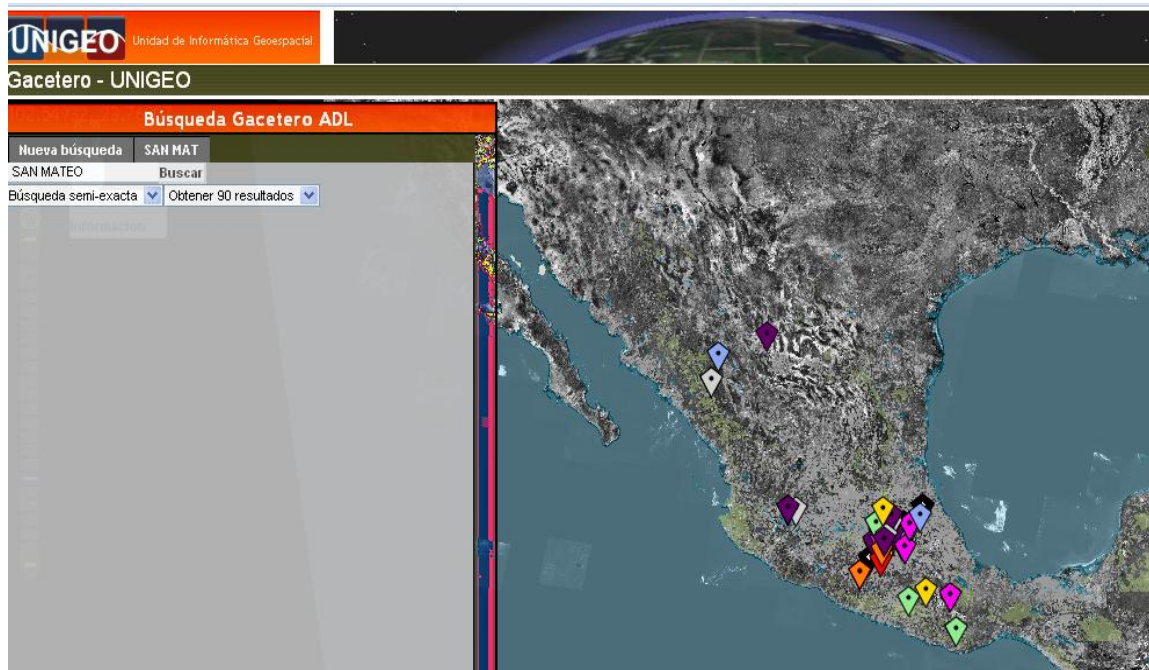


Figura 5.6 visualización gráfica de los resultados en el gacetero

La visualización de resultados en forma de texto permitirá al usuario conocer mucho más a detalle la localidad requerida, véase la figura 5.7

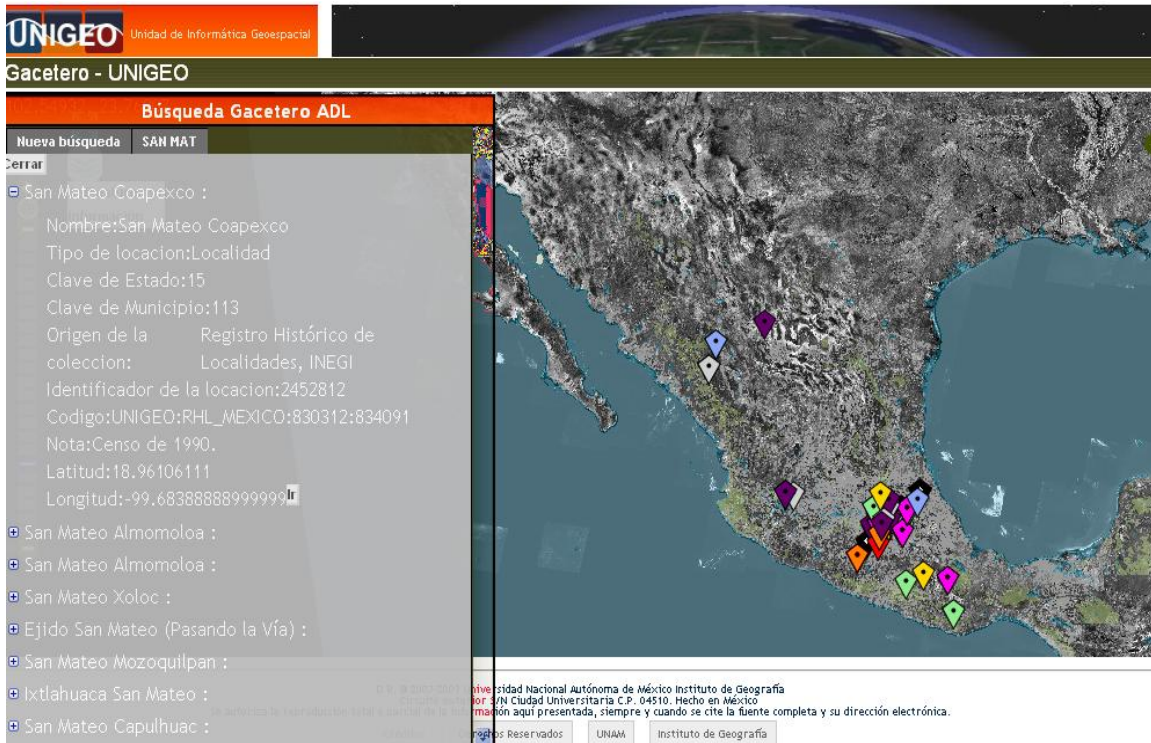


Figura 5.7 Detalle de resultados en forma de texto en el gacetero propuesto

Esta base de datos se desarrolló con un software de libre distribución llamado Postgres SQL versión 8.4, el cual es una herramienta lo suficientemente poderosa para alojar varios millones de registros en sus tablas, capaz de soportar diferentes tipos de datos que cumplieron con las necesidades de información de los usuarios del nuevo gacetero. Postgres es un *DBMS (Data base Management System)* robusto y entre sus ventajas está la de poder administrar bases de datos, dar mantenimiento, soporte y explotar los datos a través del lenguaje estructurado de consulta SQL, etc.

Cada sistema manejador de bases de datos es capaz de soportar un gran número de registros, los cuales difieren entre uno y otro. Particularmente Postgres SQL tiene algunas propiedades que son adecuadas para soportar, en este caso 1,500,000 registros de localidades, entre las más importantes se destacan las siguientes características:

Conclusiones y Recomendaciones

6.1 Conclusiones.

La georeferenciación es un procedimiento para asignar coordenadas sobre la faz de la tierra, ya sea a una imagen, o a una colecta biológica. El sistema que aquí se presenta, lleva a cabo esta técnica, de manera indirecta, mediante el uso de información histórica de localidades, tomadas a partir de una base de datos de INEGI, esto hace que resulte una herramienta adecuada para este propósito.

Existe un gran número de sistemas automatizados para realizar esta tarea, sin embargo, hasta el momento no había uno más especializado en el que se pudiera tener la oportunidad de otorgar a un usuario el poder administrar su información científica de colectas biológicas y mucho menos al estilo de un gacetero digital.

El desarrollo de un nuevo gacetero ha permitido establecer datos muy precisos acerca de localidades o rasgos, facilitando el acceso a la información en forma rápida y veraz. Este software llamado gacetero es un sistema de cómputo que cumple y satisface todas las necesidades inherentes al almacenamiento, concentración y recuperación de datos históricos necesarios para georeferir colectas históricas ya que permite explotar la información eficientemente, cruzarla en el orden que el usuario lo disponga y presentarla en forma clara.

Este sistema o gacetero está conformado por módulos de captura individual, o en forma masiva de registros provenientes de INEGI, consultas en línea y despliegue de datos en pantalla y en papel. También permite el control preciso de datos, lo cual es posible gracias al diseño completo de campos, tablas y la base de datos. Fue diseñado tomando en cuenta o basado en herramientas modernas, bajo un esquema de programación orientado a la web. Se concluye, por tanto, que este sistema ha sido capaz de atender uno de los grandes objetivos de este trabajo de tesis que es el desarrollo de un software para la gestión de localidades para que a futuro sirva e apoyo para georeferenciar colectas biológicas.

Este nuevo gacetero en su fase de prototipo ya está instalado en los servidores del Instituto de Geografía de la UNAM y listo para ponerlo a disposición de cualquier persona que requiera de este tipo de información, sin importar el lugar, fecha y hora de acceso, ya que cumple con las características técnicas propias de ambientes web.

De manera puntual se pueden establecer las siguientes conclusiones:

- Se ha diseñado e implementado un gacetero que cumple con la función de diccionario de rasgos geográficos ya que permite localizar éstos de manera rápida.

El criterio por el cual realiza las búsquedas puede ser variado, pero la fortaleza del gacetero reside en localizar los rasgos a través del nombre, aunque, se pueden usar otros criterios de búsqueda como por ejemplo, por fecha, antigüedad, clasificación, etc.

- La información se encuentra organizada jerárquicamente, por ejemplo, toda localidad se encuentra en un lugar de clasificación por debajo de municipios, los cuales a su vez se encuentran por debajo de país.
- El gacetero es apto para responder a las preguntas que cualquier usuario de este gacetero debe responderse, estas preguntas son: ¿Existe la localidad x?, ¿qué información hay de x?, ¿Dónde está x?, ¿A qué categoría pertenece x?, entre otras.
- La utilidad del nuevo gacetero se sustenta en la facilidad que éste tiene para almacenar datos generados por fuentes heterogéneas, una de ellas es el INEGI, ya que esta institución es la generadora de grandes cantidades de información en un formato no apropiado para facilitar la explotación de sus resultados. El gacetero puede incorporar estos mismos datos y presentarlos en un formato orientado en tres rubros: nombre, ubicación y categoría.
- El gacetero facilita el almacenamiento de datos históricos y es capaz de accederlos para ser presentados al usuario en forma rápida y eficiente.
- El modelo ADL y GCS efectivamente pueden ser empleados para tomarse de base en la implementación de gaceteros digitales con datos históricos.
- El gacetero es suficientemente flexible para incorporar datos que corresponden directamente a las tablas básicas del modelo E-R, tales como nombre, clasificación, ubicación, tiempo, etc., empleando el apartado “datos del rasgo” o `g_feature_data`, dentro del cual, cada dato es a su vez clasificado y almacenado acorde a un esquema general.
- El modelo es suficientemente versátil al permitir definir relaciones arbitrarias entre sus rasgos. Estas relaciones se pueden tipificar y permiten expresar relaciones no propuestas en el modelo ni por los datos existentes, por ejemplo, la pertinencia, las agrupaciones y las situaciones administrativas entre los rasgos. Las relaciones de los datos se pueden hacer tan variadas que el único límite está en el conocimiento que el diseñador tenga del lenguaje de consulta utilizado.

En general, el nuevo gacetero funciona de acuerdo al diseño propuesto, aún así sigue en proceso de mantenimiento por parte del personal del Instituto de Geografía y paulatinamente se hará la inclusión en las actividades cotidianas de los usuarios y del personal de desarrollo de sistemas, así como de la observación de su comportamiento.

6.2 Recomendaciones

A pesar de que este nuevo gacetero cumple con la función de ofrecer a los usuarios la posibilidad de poder tener a la mano los datos de las colectas biológicas y de las

localidades deseadas, junto con sus datos relacionados, a futuro deberán expandirse sus posibilidades de gestionar la información. Por este motivo se recomienda que en versiones futuras de este gacetero, no solo se creen relaciones comunes entre localidades, y que además son las que se deducen de las tablas creadas a partir del modelo Entidad-Relación que se mencionó en la figura 3.6, sino que también se implementen relaciones de alto nivel, por ejemplo entre toponimias y sus rasgos, también queda abierta la posibilidad de establecerse entre elementos espaciales, elementos administrativos, como son cercanía, relaciones jurídicas, contenidos de intersección e involucrar esquemas y nuevos términos.

Todo lo anterior no será otra cosa más que implementar otras metodologías no existentes para este tipo de casos, es decir, deberán construirse aplicaciones que aprovechen las relaciones entre datos de origen diferente entre ellos.

También se recomienda seguir el uso de Postgres como el manejador de bases de datos, puesto que también es un software moderno, poderoso y que es capaz de soportar el tipo de datos que los usuarios del nuevo gacetero necesitan.

Por otro lado, el equipo de cómputo necesario para este gacetero debe contar con características adecuadas para manipular registros del orden de los cientos de millones, es decir, servidores con una gran capacidad de disco duro, memoria *RAM* y virtual y control de usuarios con las políticas tradicionales de este tipo de sistemas.

El diseño de nuevos reportes dependerá en gran medida de los requerimientos de los usuarios, quienes con el paso del tiempo tendrán mejor visión de la capacidad del nuevo gacetero por lo que habrá que simplificar muchas de sus tareas con base en procesos automáticos que estén basados en la información de las tablas originales del gacetero, es decir, al nuevo diseño.

Por último, es recomendable que absolutamente todos los diseños de nuevos reportes, interfaces y procesos que manipule el usuario sigan basados en la web, es decir, que cumplan técnicamente con los mecanismos que Internet ofrece en el manejo de sistemas de información, todo esto para estar a la vanguardia en tecnología, velocidad y utilidad de las nuevas herramientas disponibles en el mercado.

Bibliografía

- Kenneth E. Kendall (2006). Análisis y Diseño de Sistemas. Prentice Hall Hispanoamericana, S.A. Sexta Edición
- Silberschatz U. (2004). Fundamentos de Bases de Datos. McGraw Hill Tercera Edición
- Roger S. Pressman (2001). Ingeniería del Software, Un enfoque práctico. McGraw Hill
- Ian Sommerville (2005). Ingeniería del software. Pearson-Addison Wesley Séptima Edición
- Elmasri / Navathe (2001). Sistemas de Bases de Datos Conceptos Fundamentales. Addison Wesley Iberoamericana Segunda edición
- Hernandez Sampieri Roberto, Fernández-Collado Carlos, Baptista Lucio Pilar (2006). Metodología de la Investigación. Mc Graw Hill Cuarta edición
- García Córdoba Fernando (2005). La Tesis y el trabajo de tesis, recomendaciones metodológicas para la elaboración de los trabajos de tesis. Limusa Noriega Editores, México
- Rivera Heredia María Elena (2009). Competencias para la Investigación. Trillas, México
- Hill Linda I. (2006). Georeferencing, the geographic associations of information. The Massachusetts Institute of Technology Press Cambridge, Massachusetts London, England
- Georeferenciación de localidades de colecciones biológicas, manual de procedimientos, CONABIO, México 2008, pp. 34

Referencias electrónicas

http://www.siba.unam.mx/index.php?option=com_remository&Itemid=2

http://www.siba.unam.mx/index.php?option=com_remository&Itemid=2&func=startdown&id=26

http://www.siba.unam.mx/index.php?option=com_remository&Itemid=2&func=download&id=38&chk=da72bfefab34641e35cf910520d3b722&no_html=1

http://www.siba.unam.mx/index.php?option=com_remository&Itemid=2&func=startdown&id=38

<http://datum.unigeo.igeograf.unam.mx:57212/gacetero/Gacetero.html>

<http://gborg.postgresql.org>

<http://pgfoundry.org/>

<http://www.biodiversidad.gob.mx/region/georeferenciacion.html> recuperado el 5 de octubre de 2009

<http://mapserver.inegi.gob.mx/rnng/> recuperado el 30/julio/2009

http://mapserver.inegi.gob.mx/dsist/ahl2003/AHL_proyecto.cfm?c=424 recuperado el 7/agosto/2009

<http://www.museum.tulane.edu/geolocate/default.aspx> recuperado el 2 de octubre de 2009
<http://74.6.146.127/search/cache?ei=UTF-8&p=Guide+to+Best+Practices+for+Georeferencing&rd=r1&fr=yfp&u=arobis.cenpat.gov.ar/recursos-tecnicos/georeferenciacion/material-taller/BestPractices.pdf&w=guide+best+practices+georeferencing&d=EDw9OBIMTG6O&icp=1&.intl=mx> recuperado el 7/agosto/2009

(<http://www.biogeomancer.org/about.html>) recuperado el 7/agosto/2009

<http://www.alexandria.ucsb.edu/gazetteer/> recuperado el 7/agosto/2009

<http://www.unigeo.igeograf.unam.mx/unigeo/index.php/UNIGEO.html> recuperado el 6/agosto/2009

Evaluación 2009.pdf (Informe de la evaluación anual de SIBA durante el periodo 2008 - 2009, realizada del 13 al 16 de abril.)

http://siba.ibiologia.unam.mx/index.php?option=com_remository&Itemid=2&func=startdown&id=42 recuperado el 29/julio/2009

<http://www.dlib.org/dlib/january99/hill/01hill.html> recuperado el 7/agosto/2009

<http://www.alexandria.ucsb.edu/gazetteer/ContentStandard/version3.2/GCS3.2-guide.htm>
recuperado el 7/agosto/2009

<http://www.dlib.org/dlib/may04/goodchild/05goodchild.html> recuperado el 17 de octubre de 2009

<http://www.acm.org/crossroads/espanol/xrds5-2/diglib.html> recuperado el 20 de octubre de 2009

<http://www.desarrolloweb.com/articulos/449.php> recuperado el 11 de noviembre de 2009

http://74.125.47.132/search?q=cache:8WI_B77PrxQJ:www.herpnet.org/herpnet/document/s/georrefparadum.xls+qu%C3%A9+es+la+georeferenciaci%C3%B3n+de+localidades%3F&cd=9&hl=es&ct=clnk&gl=mx recuperado el 5 de octubre de 2009

Glosario

Abiótico(a).- En el ámbito de la biología y la ecología, el término abiótico designa a aquello que no es biótico, es decir, que no forma parte o no es producto de los seres vivos.

Altitud.- Altura en metros, con relación al nivel medio del mar.

Atributo.- Es una propiedad de los objetos, la cual describe características geométricas, topológicas u otras. Es una característica que califica y describe un aspecto de un objeto.

Biodiversidad.- La variabilidad de organismos vivos de cualquier fuente, incluidos, entre otros, los ecosistemas terrestres, marinos y otros ecosistemas acuáticos y los complejos ecológicos de los que forman parte; comprende la diversidad dentro de cada especie, entre las especies y de los ecosistemas.

Carta topográfica.- Mapa o carta que conforme a los procesos de producción aplicados por el INEGI se obtiene a partir de fotografías aéreas, levantamientos geodésicos, clasificación de detalles y captación de toponimia en campo, además de procesos fotogramétricos y de edición y reproducción. Suele llamársele también carta fotogramétrica. Se distingue de las cartas o mapas planimétricos por contener información sobre el relieve del terreno.

Cartografía.- La representación en cartas de la información geográfica.

Coordenadas geográficas.- Ángulos que determinan la posición de un punto o lugar. Sistema universal para la localización de puntos sobre la superficie terrestre. Se basa en un conjunto de anillos imaginarios que rodean a la esfera terrestre (paralelos y meridianos). Sus coordenadas se denominan Latitud y Longitud.

Colecta.- Actividad que consiste en la captura, remoción o extracción temporal o definitiva de material biológico del medio silvestre, con propósitos no comerciales, para la obtención de información científica básica, integración de inventarios o para incrementar los acervos de las colecciones científicas.

Dato espacial.- Objeto o entidad que resulta de una abstracción del espacio geográfico real y que puede ser un rasgo natural, obra humana o alguna abstracción numérica derivada del tratamiento de cifras relacionadas con tal objeto o entidad. Su característica definitiva es la referencia espacial, en dos o tres dimensiones. En algunos caso se consideran sinónimos de dato espacial los siguientes términos: dato geográfico, dato geoespacial, entidad geográfica, objeto espacial, rasgo geográfico y demás que correspondan con esta definición.

Datos geoestadísticos.- Son aquéllos ya procesados estadísticamente y referenciados a su espacio geográfico al menos a nivel municipal. Comprenden los censos permanentes previstos en la ley y demás estadísticas socioeconómicas de interés nacional.

Datum.- El datum es una serie de parámetros que definen un sistema de coordenadas, y un grupo de puntos de los cuales son conocidas sus relaciones geométricas. Un Datum

es definido por un esferoide. Existen varios esferoides que representan la forma de la Tierra y muchos más Datums basados en ellos. Un Datum horizontal da el marco de referencia para realizar medidas locales o regionales sobre la superficie terrestre. Este define el origen y la orientación de las líneas de latitud y longitud.

Diccionario de datos.- Son documentos normativos que están dedicados a establecer especificaciones a nivel de objetos espaciales. Describen cada objeto en términos de su definición, sus atributos, los dominios de valores permitidos para cada atributo, así como su representación espacial y las restricciones de integridad.

Entidad.- Es una abstracción de un fenómeno del mundo real asociada a una localización relativa a la Tierra.

Gacetero.- Diccionario geográfico que proporciona acceso a datos geoespaciales indexados por nombres de lugares más que por coordenadas de localización.

Georeferenciación.- La georeferenciación es el posicionamiento de un objeto espacial que puede estar representado mediante puntos o líneas. Si se trata de una obra relacionada con escuelas, pozos de agua, puentes, pasos laterales, entre otros, se trata de un proyecto puntual. Mientras que si la obra se refiere a la construcción o rehabilitación de una vía, se identifican sus coordenadas de inicio y final para representarla mediante una línea.

Información geográfica.- El conjunto de datos, símbolos y representaciones organizados para conocer y estudiar las condiciones ambientales y físicas del territorio nacional, la integración de éste en infraestructura, los recursos naturales y la zona económica exclusiva.

Informática.- Tecnología para el tratamiento sistemático y racional de la información mediante el procesamiento electrónico de datos.

Latitud geodésica o latitud.- Ángulo que la normal al elipsoide en un punto forma con el plano del ecuador, positivo si está dirigido hacia el Norte.

Localidad.- Es todo el lugar ocupado con una o más edificaciones utilizadas como viviendas, las cuales pueden estar habitadas o no, este lugar es reconocido por un nombre dado por la ley o la costumbre.

Localidad rural.- Localidad con población menor a 2 500 habitantes, y no son cabeceras municipales.

Localidad urbana.- Son localidades que tienen una población igual o mayor a 2500 habitantes o es cabecera municipal, independientemente del número de habitantes registrado o en el último Censo o Censo Nacional de Población y Vivienda.

Nomenclátor.- Catálogo de entes del mundo real que contiene alguna información sobre su posición.

Ortofoto.- es una imagen fotográfica de una zona de la superficie terrestre, en la que todos los elementos presentan la misma escala, libre de errores y deformaciones, con la misma validez de un plano cartográfico.

Raster imagen.- También llamada mapa de bits, imagen matricial o bitmap, es una estructura o archivo de datos que representa una rejilla rectangular de píxeles o puntos de color. Se las suele caracterizar por su altura y anchura (en píxeles) y por su profundidad de color (en bits por pixel), que determina el número de colores distintos que se pueden almacenar en cada pixel, y por lo tanto, en gran medida, la calidad del color de la imagen.

Toponimia.- Rasgos puntuales de tipo hidrográfico, rasgos ortográficos, áreas naturales y culturales, formas litorales y obras de infraestructura

Topónimo.- nombre derivado de un lugar o región, también puede ser el nombre biológico de una planta.

Vectorial imagen.- Este tipo de imagen se crea a través del uso de objetos geométricos como curvas de Bézier y polígonos, no del simple almacenamiento del color de cada pixel.

Anexo A

Guía para georeferenciación de lugares

Entre las principales actividades dentro de la georeferenciación están las siguientes a manera de prontuario: La tabla completa puede ser consultada en Internet²²

TIPO DE LUGAR	PROCEDIMIENTO DE GEORREFERENCIACIÓN	DETERMINACIÓN DE LA EXTENSIÓN Y DEL ERROR
<p><u>Lugar citado:</u></p> <p>Entidad por Área Urbana:</p> <p>La localidad consiste en una referencia a un elemento geográfico que tiene una extensión espacial definida</p> <p><i>Ejemplo: "Las Vegas"</i></p>	<p>Lo mejor es estimar el centro geográfico (el punto medio entre los valores extremos de latitud y longitud) como las coordenadas de los lugares citados. Si el centro geográfico no está dentro del área urbana marcada, entonces se selecciona el punto más cercano al centro geográfico que esté dentro del área de la figura.</p> <p>Si se consulta un índice geográfico (<i>gazetteer</i>) tenga en consideración que muchos índices registran las coordenadas de la principal oficina postal de un lugar poblado. En el caso de la cabecera de un distrito, se anotan las coordenadas del palacio de justicia.</p>	<p>Extensión: Use la distancia entre las coordenadas del lugar citado y el punto más lejano del mismo. Algunos índices pueden dar áreas delimitadas para describir la extensión de lugares vastos.</p> <p>Error: En la Calculadora de Georeferenciación, seleccione los siguientes ajustes: Tipo de Cálculo: "Sólo Error - introduzca la Lat/Long para la localidad actual" Tipo de Localidad: "Sólo un Lugar citado"</p>
<p><u>Lugar citado:</u></p> <p>Dirección de una calle</p> <p><i>Ejemplo: "1 Orchard</i></p>	<p>Localice la dirección usando buscadores electrónicos o sitios web como Maporama o Mapquest.</p> <p>Localice la dirección en el mapa o con ayuda de software cartográfico para obtener las coordenadas si el</p>	<p>Extensión: La extensión es el área más pequeña posible que no se pueda confundir con ninguna otra dirección. Muchas veces la distancia será la mitad de la longitud de una cuadra de una</p>

22

http://74.125.47.132/search?q=cache:8WI_B77PrxQJ:www.herpnet.org/herpnet/documents/georrefparadum.xls+qu%C3%A9+es+la+georeferenciación+C3%B3n+de+localidades%3F&cd=9&hl=es&ct=clnk&gl=mx recuperado el 5 de octubre de 2009

<p><i>Lane, Berkeley, CA"</i></p>	<p>buscador electrónico no los muestra automáticamente.</p>	<p>ciudad, debido a la escala de los mapas en línea en Internet.</p> <p>Error: Lo mismo que para un lugar nombrado</p>
<p><u>Lugar citado:</u></p> <p>Lago, Montaña, Parque Cultural, u Otra Entidad Geográfica</p> <p><i>Ejemplo: "Parque Nacional Yosemite", "Lago Mead", "Condado de Alameda"</i></p>	<p>Sitúe las coordenadas en el centro geográfico. Si no puede localizar el centro geográfico fácilmente, dibuje un rectángulo que incluya por completo el lugar citado y trace las dos líneas diagonales para determinar el centro.</p> <p>Recuerde que las coordenadas tienen que estar localizadas dentro de la propia entidad geográfica. Si la localidad está dividida en más de un fragmento (como ocurre en un área de recreo que esté dividida por terrenos privados), asegúrese de que las coordenadas estén situadas realmente sobre la localidad y no en el punto intermedio entre dos fragmentos.</p>	<p>Extensión: Use la distancia desde las coordenadas del lugar citado hasta el punto más lejano del mismo o, en caso de haberlas dibujado, la mitad de una de las líneas diagonales</p> <p>Error: Lo mismo como para un lugar nombrado</p>

Anexo B

Código fuente utilizado en la generación de las principales tablas que conforman la base de datos del gacetero propuesto:

```
CREATE TABLE g_begin_end_date (  
  begin_end_date_id integer NOT NULL,  
  time_period_id integer NOT NULL,  
  calendar_system character varying(20) DEFAULT 'Gregorian'::character varying NOT NULL,  
  date_coding_scheme_id integer NOT NULL,  
  begin_date date,  
  begin_date_confidence_value double precision,  
  begin_date_confidence_note character varying(80),  
  begin_date_note character varying(80),  
  end_date date,  
  end_date_confidence_value double precision,  
  end_date_confidence_note character varying(80),  
  end_date_note character varying(80));
```

```
CREATE TABLE g_classification (  
  classification_id integer NOT NULL,  
  feature_id integer NOT NULL,  
  classification_term_id integer NOT NULL,  
  primary_display boolean NOT NULL,  
  time_period_id integer NOT NULL,  
  time_period_note character varying(255)  
);
```

```
CREATE TABLE g_collection (  
  collection_id integer NOT NULL,  
  name character varying(100) NOT NULL,  
  note character varying(255)  
);
```

```
CREATE TABLE g_description (  
  description_id integer NOT NULL,  
  feature_id integer NOT NULL,  
  description_type_scheme_id integer,  
  description_type_term character varying(255),  
  external_scheme_term_id character varying(80),  
  short_description character varying(2000) NOT NULL  
);
```

```
CREATE TABLE g_encoded_date (  
  encoded_date_id integer NOT NULL,  
  time_period_id integer NOT NULL,  
  date_coding_scheme_id integer NOT NULL,  
  encoded_date_string character varying(255) NOT NULL  
);
```

```
CREATE TABLE g_entry_source (  
  entry_source_id integer NOT NULL,  
  source_id integer NOT NULL,  
  entry_date date NOT NULL  
);
```

```

CREATE TABLE g_feature (
  feature_id integer NOT NULL,
  collection_id integer,
  is_complete boolean,
  time_period_id integer NOT NULL,
  entry_note character varying(80),
  entry_date date NOT NULL,
  modification_date date NOT NULL
);

CREATE TABLE g_feature_code (
  feature_code_id integer NOT NULL,
  feature_id integer NOT NULL,
  code character varying(20) NOT NULL,
  code_scheme_id integer NOT NULL
);

CREATE TABLE g_feature_data (
  feature_data_id integer NOT NULL,
  feature_id integer NOT NULL,
  data_type_scheme_id integer NOT NULL,
  data_type_term character varying(80) NOT NULL,
  external_scheme_term_id character varying(80),
  data_value numeric NOT NULL,
  data_unit character varying(20) NOT NULL,
  data_basis character varying(80),
  data_note character varying(255),
  time_period_id integer NOT NULL,
  time_period_note character varying(255)
);

CREATE TABLE g_feature_displayname (
  feature_id integer,
  displayname character varying(255),
  source_feature_ids character varying(127)
);

CREATE TABLE g_feature_link (
  feature_link_id integer NOT NULL,
  feature_id integer NOT NULL,
  link_description character varying(80) NOT NULL,
  feature_link_scheme_id integer,
  feature_link_type_term character varying(255),
  external_scheme_term_id character varying(80),
  language_id integer NOT NULL,
  link_url character varying(50)
);

CREATE TABLE g_feature_name (
  feature_name_id integer NOT NULL,
  feature_id integer NOT NULL,
  primary_display boolean NOT NULL,
  name character varying(255) NOT NULL,
  etymology character varying(80),
  language_id integer,
  transliteration_scheme_id integer,

```

```

    confidence_note character varying(80)
);
CREATE TABLE g_location (
    location_id integer NOT NULL,
    feature_id integer NOT NULL,
    planet character varying(20) DEFAULT 'Earth'::character varying NOT NULL,
    bounding_box_geodetic character varying(20) DEFAULT 'WGS-84'::character varying,
    west_coordinate double precision NOT NULL,
    east_coordinate double precision NOT NULL,
    south_coordinate double precision NOT NULL,
    north_coordinate double precision NOT NULL,
    deleted_column1 integer,
    bounding_box_method character varying(80) NOT NULL,
    bounding_box_source_type character varying(20) NOT NULL
);

CREATE TABLE g_location_geometry (
    location_geometry_id integer NOT NULL,
    location_id integer NOT NULL,
    primary_geometry boolean NOT NULL,
    local_geometry boolean NOT NULL,
    geometry_coding_scheme_id integer NOT NULL,
    encoded_geometry character varying(32000),
    geometry_reference_url character varying(500),
    time_period_id integer NOT NULL,
    measurement_begin_date date,
    measurement_end_date date,
    measurement_method character varying(255),
    geometry_confidence_value double precision,
    geometry_confidence_note character varying(255),
    time_period_note character varying(255)
);

CREATE TABLE g_name_abbreviation (
    feature_name_id integer NOT NULL,
    name_abbreviation character varying(20) NOT NULL
);

CREATE TABLE g_name_to_link_info_reference (
    feature_name_id integer NOT NULL,
    source_reference_id integer NOT NULL,
    pages character varying(20)
);

CREATE TABLE g_name_to_time_period (
    feature_name_id integer NOT NULL,
    time_period_id integer NOT NULL,
    primary_time_period boolean NOT NULL,
    time_period_note character varying(255)
);

CREATE TABLE g_name_toponymic_authority (
    feature_name_id integer NOT NULL,
    toponymic_contributor_id integer NOT NULL
);

```



```
CREATE TABLE g_pronunciation (  
  pronunciation_id integer NOT NULL,  
  feature_name_id integer NOT NULL,  
  pronunciation_note character varying(80),  
  pronunciation_text character varying(80),  
  pronunciation_text_url character varying(500),  
  pronunciation_audio_url character varying(500)  
);
```

```
CREATE TABLE g_related_feature (  
  related_feature_id integer NOT NULL,  
  feature_id integer NOT NULL,  
  related_name character varying(100) NOT NULL,  
  related_feature_feature_id integer,  
  time_period_id integer NOT NULL,  
  related_type_term_id integer NOT NULL,  
  time_period_note character varying(255)  
);
```

```
CREATE TABLE g_source (  
  source_id integer NOT NULL,  
  source_mnemonic character varying(80) NOT NULL,  
  contributor_id integer NOT NULL,  
  source_reference_id integer NOT NULL  
);
```

```
CREATE TABLE g_supplemental_note (  
  feature_id integer NOT NULL,  
  supplemental_note character varying(1000) NOT NULL  
);
```

```
CREATE TABLE g_time_date_range (  
  time_period_id integer NOT NULL,  
  date_range_begin date,  
  date_range_end date,  
  date_range_note character varying(255)  
);
```

```
CREATE TABLE g_time_period (  
  time_period_id integer NOT NULL,  
  status_term_id integer NOT NULL  
);
```

```
CREATE TABLE g_time_period_to_period_name (  
  time_period_id integer NOT NULL,  
  time_period_name_id integer NOT NULL  
);
```

```
CREATE TABLE I_author (  
  author_id integer NOT NULL,  
  author character varying(100) NOT NULL  
);
```

```
CREATE TABLE I_contributor (  
  contributor_id integer NOT NULL,
```

```

organization_name character varying(100) NOT NULL,
site_title character varying(100),
contributor_url character varying(500),
contact_name character varying(100),
street_address character varying(80) NOT NULL,
city character varying(80) NOT NULL,
state_province character varying(80) NOT NULL,
postal_code character varying(20),
country character varying(20)
);

```

```

CREATE TABLE I_language (
language_id integer NOT NULL,
language_code character varying(3) NOT NULL,
language_scheme_id integer NOT NULL
);

```

```

CREATE TABLE I_scheme (
scheme_id integer NOT NULL,
scheme_name character varying(100) NOT NULL,
scheme_abbreviation character varying(20),
scheme_version character varying(20) NOT NULL,
download_url character varying(500),
download_mime_type character varying(20),
service_url character varying(500),
service_protocol_name character varying(80),
service_protocol_url character varying(500),
web_interface_url character varying(500),
offline_source_citation character varying(500),
scheme_type character varying(20) NOT NULL,
primary_classification_scheme boolean NOT NULL
);

```

```

CREATE TABLE I_scheme_term (
scheme_term_id integer NOT NULL,
scheme_id integer NOT NULL,
term character varying(255) NOT NULL,
external_scheme_term_id character varying(80)
);

```

```

CREATE TABLE I_scheme_term_parent (
scheme_term_parent_id integer NOT NULL,
scheme_term_id integer NOT NULL,
parent_scheme_term_id integer NOT NULL
);

```

```

CREATE TABLE I_scheme_term_rank (
scheme_term_id integer NOT NULL,
ftt_order1 integer,
ftt_order2 integer
);

```

```

CREATE TABLE I_source_reference (
source_reference_id integer NOT NULL,
citation character varying(500),

```

```

reference_author_id integer,
reference_date date,
ibsn character varying(80),
issn character varying(80),
reference_url character varying(50)
);

CREATE TABLE l_time_period_name (
time_period_name_id integer NOT NULL,
time_period_name character varying(100) NOT NULL,
time_period_scheme_id integer,
external_scheme_term_id character varying(80)
);

CREATE TABLE s_address (
address_id integer NOT NULL,
entry_source_id integer NOT NULL
);

CREATE TABLE s_classification (
classification_id integer NOT NULL,
classification_term_id integer NOT NULL,
time_period_id integer NOT NULL
);

CREATE TABLE s_description (
description_id integer NOT NULL,
entry_source_id integer NOT NULL
);

CREATE TABLE s_feature (
feature_id integer NOT NULL,
time_period_id integer NOT NULL,
entry_note integer
);

CREATE TABLE s_feature_code (
feature_code_id integer NOT NULL,
entry_source_id integer NOT NULL
);

CREATE TABLE s_feature_data (
feature_data_id integer NOT NULL,
entry_source_id integer NOT NULL
);

CREATE TABLE s_feature_link (
feature_link_id integer NOT NULL,
entry_source_id integer NOT NULL
);

CREATE TABLE s_feature_name (
feature_name_id integer NOT NULL,
name integer NOT NULL,
etymology integer,

```

```

    language_id integer,
    transliteration_scheme_id integer,
    confidence_note integer
);

CREATE TABLE s_location (
    location_id integer NOT NULL,
    bounding_box_source_entry_id integer NOT NULL
);

CREATE TABLE s_location_geometry (
    location_geometry_id integer NOT NULL,
    time_period_id integer NOT NULL,
    entry_source_id integer NOT NULL
);

CREATE TABLE s_name_to_link_info_reference (
    feature_name_id integer NOT NULL,
    source_reference_id integer NOT NULL,
    entry_source_id integer NOT NULL
);

CREATE TABLE s_name_to_time_period (
    feature_name_id integer NOT NULL,
    time_period_id integer NOT NULL,
    entry_source_id integer NOT NULL
);

CREATE TABLE s_name_toponymic_authority (
    feature_name_id integer NOT NULL,
    toponymic_contributor_id integer NOT NULL,
    entry_source_id integer NOT NULL);

CREATE TABLE s_pronunciation (
    pronunciation_id integer NOT NULL,
    entry_source_id integer NOT NULL
);

CREATE TABLE s_related_feature (
    related_feature_id integer NOT NULL,
    time_period_id integer NOT NULL,
    entry_source_id integer NOT NULL
);

CREATE TABLE s_supplemental_note (
    feature_id integer NOT NULL,
    entry_source_id integer NOT NULL
);

CREATE TABLE s_time_period (
    time_period_id integer NOT NULL,
    entry_source_id integer NOT NULL
);

```