

BASES DE DATOS Y SISTEMAS DE INFORMACIÓN: APLICACIONES EN BIOGEOGRAFÍA

por

Tania Escalante E.¹, Jorge Llorente B.^{1,2}, David N. Espinosa O.³,
Jorge Soberón M.⁴

El proceso ideal de toda investigación consiste en:

- 1) Un examen minucioso de cada uno de los elementos.*
- 2) Una oposición de los diferentes fenómenos entre sí, interacción y comparación, y,*
- 3) Unas conclusiones generales deducidas de las dos etapas anteriores.*

Punto y línea sobre el plano (1923) Wassily Kandinsky.

Resumen

Escalante T., J. Llorente, D. Espinosa & J. Soberón: Bases de datos y sistemas de información: Aplicaciones en Biogeografía. Rev. Acad. Colomb. Cienc. **24**(92): 325-341, 2000. ISSN 0370-3908.

Se describen algunos aspectos sobre la nueva instrumentalización y elementos metodológicos en la conformación de Sistemas de Información en Biodiversidad (SIB). El uso de datos georreferidos con precisión a partir de vastas fuentes disponibles: colecciones de historia natural y literatura científica obligan al uso de bases de datos y Sistemas de Información Geográfica (SIG). La conceptualización de SIB y SIG, con base en el uso de grandes bases de datos, han implicado modelación detallada y construcción de archivos de autoridad: catálogos exhaustivos de nomenclatura y sinonimia, listas bibliográficas completas y nomencladores o gazeteros histórico-geográficos con localidades y sus sinonimias ubicadas bajo un sistema de Geoposicionamiento Global (GPS en sus siglas inglesas) que se genera de una concepción geoesférica de la Tierra y su biota. Se explican algunos problemas en el desarrollo de los sistemas y en la construcción de las bases de datos bióticas: control de calidad de datos, por ejemplo. Se muestra que el uso de tales sistemas es fundamental para responder a numerosas preguntas de frontera en los actuales estudios de biodiversidad y bioconservación. En particular, se detallan algunas aplicaciones en biogeografía y su importancia

1. Museo de Zoología, Facultad de Ciencias. UNAM. Apdo. Postal 70-399. México 04510 D.F.
2. Instituto de Ciencias Naturales. Universidad Nacional de Colombia.
3. Facultad de Estudios Superiores Zaragoza, UNAM.
4. Instituto de Ecología, UNAM. México.

para modelar distribuciones, contrastar e identificar áreas de conservación de endemismos y riqueza biótica, y su uso como herramienta en lo que denominamos faunística predictiva y experimental. Finalmente se destaca el proceso en cuanto a su relevancia a niveles nacional y regional.

Palabras clave: Sistemas de Información Geográfica, Faunística predictiva, Sistemas de Información en Biodiversidad, colecciones científicas, bases de datos, áreas de distribución, Análisis de discontinuidades, biogeografía.

Abstract

Some aspects of the new instrumentalization and methodological elements that make up Information Systems in Biodiversity (ISB) are described. The use of accurate geographically referenced data allows a broad range of available sources: natural history collections and scientific literature require the use of data bases and Geographic Information Systems (GIS). The conceptualization of ISB and GIS, based in the use of extensive data bases, has implied detailed modeling and the construction of authoritative archives: exhaustive catalogues of nomenclature and synonymies, complete bibliographic lists, list of names proposed, historico-geographic gazetteers with localities and their synonyms united under a Global Positioning System which produces a geospheric conception of the earth and its biota. Certain difficulties in the development of the system and the construction of the biological data bases are explained: quality control of data, for example. The use of such systems is basic in order to respond to many questions at the frontier of current studies of biodiversity and conservation. In particular, some applications in biogeography and their importance for modeling distributions, to identify and contrast areas of endemism and biological richness for conservation, and their use as tools in what we identify as predictive and experimental faunistics are detailed. Lastly, the process as well as its relevance are emphasized at national and regional levels.

Key words: Geographic Information Systems, Predictive faunistics, Information Systems of Biodiversity, scientific collections, data bases, distribution areas, discontinuity analysis, Biogeography.

Introducción

Las tareas de proteger, manejar y usar la biodiversidad a nivel nacional llegan a ser simplemente imposibles sin el uso de modernas y poderosas herramientas informáticas (Soberón & Koleff, 1998a). En la actualidad, la información biológica alcanza magnitudes insospechadas y se encuentra en continua depuración. Además, día a día se generan grandes cantidades de datos que deben ser estandarizados, sistematizados y puestos a disposición de la comunidad científica mundial. Tan solo en México existen cerca de 80 sedes institucionales con poco más de 180 colecciones científicas, las cuales albergan aproximadamente 10,000,000 de ejemplares (Llorente & Koleff, 1997; Llorente *et al.*, 1999). Realizar una estimación de la cantidad de información o de los datos que contienen las colecciones, y la que es posible extraer de los ejemplares, es impresionante, en especial si se incluyen otras fuentes de datos como las notas de libretas de campo, los distintos catálogos de las colecciones y la bibliografía especializada. Duckworth *et al.* (1993) calculan 2500 millones de ejemplares en las colecciones de historia natural del mundo, además de varios millones de publicaciones relacionadas con la biota. Todos estos datos son susceptibles de organizarse para utilizarse en el desarrollo de estrategias de conservación y manejo

sustentable de los recursos, siempre y cuando estén disponibles y estructurados lógicamente para su consulta. Varios modelos de información taxonómica han sido publicados en años recientes y los aspectos bioinformáticos se han hecho una nueva especialidad (Berendsohn, 1997; Berendsohn *et al.*, 1999).

En los últimos años se han desarrollado Sistemas de Información sobre Biodiversidad (SIB) y Geográficos (SIG), los cuales son herramientas que permiten conjuntar los datos almacenados para hacer interpretaciones y realizar aplicaciones prácticas en el campo de la conservación, a partir de patrones biogeográficos descubiertos al integrar los datos. El proceso se puede concebir e ilustrar como un continuo flujo de información y retroalimentación (Figura 1). A continuación serán descritas las principales etapas en que se lleva a cabo este proceso.

Obtención de datos y sus fuentes

Al menos se han reconocido tres fuentes de datos básicas en taxonomía y biogeografía con influencia sobre biodiversidad: los provenientes de colecciones y museos,

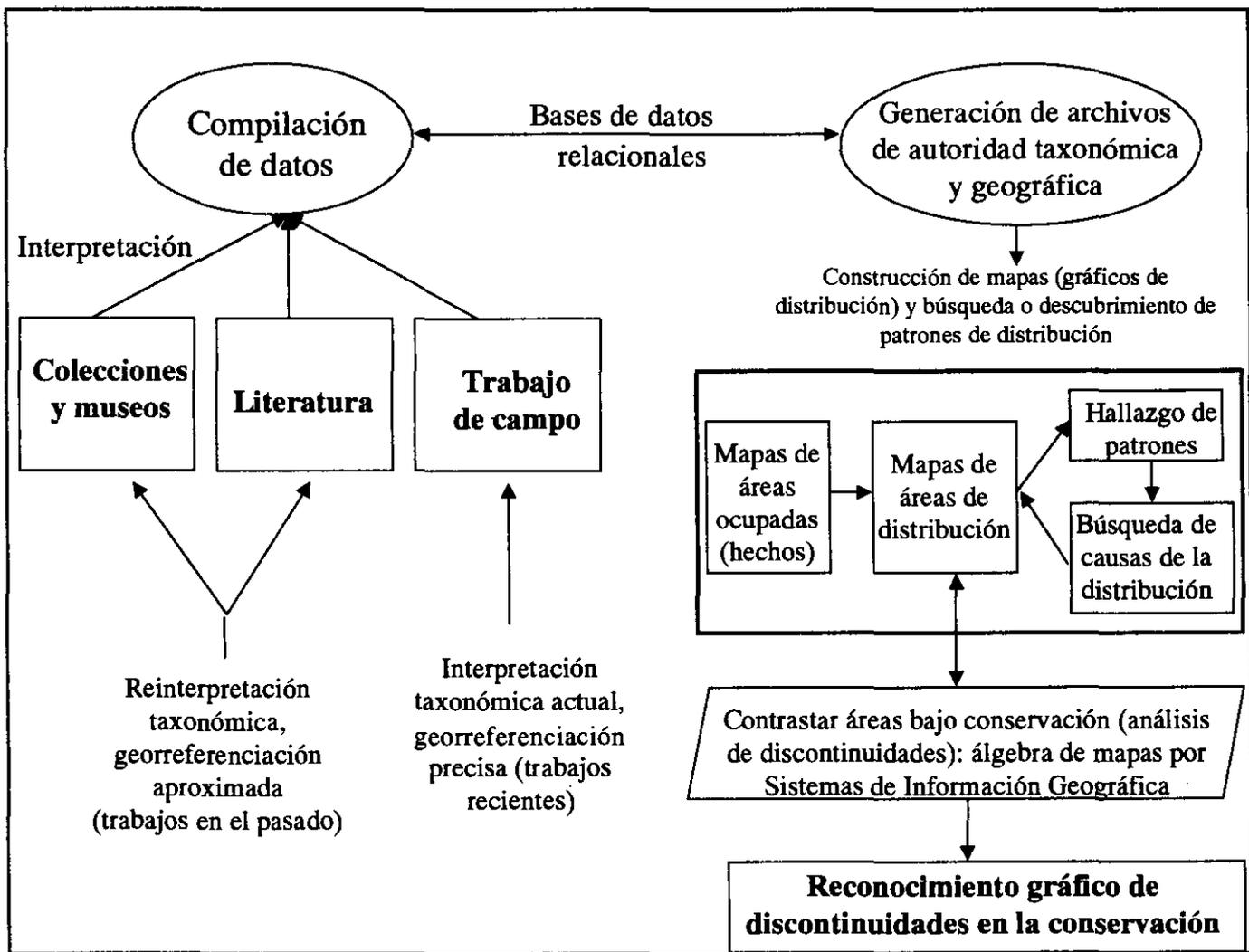


Figura 1. Proceso de integración, almacenamiento y aplicación de los datos biológicos en biogeografía.

la hemerobibliografía biológica, y el trabajo de campo reciente. En los dos primeros, debido a que en su gran mayoría se trata de trabajo histórico, es necesario considerar que puede existir una interpretación taxonómica diferente a la actual, y que la georreferenciación sólo se llevó a cabo de manera aproximada. Al contrario, los datos que se están generando hoy en día a partir del trabajo en campo, son realizados bajo una perspectiva taxonómica contemporánea (a menudo con mayor información), y las herramientas de georreferencia (por ejemplo: imágenes de satélite, Sistemas de Posicionamiento Global o GPS) permiten conocer con mayor grado de exactitud la ubicación y características de las localidades de recolecta.

Colecciones y museos. La información que poseen los especímenes proporciona una de las bases para los inventarios

nacionales. Con frecuencia de ellos se tienen datos de localidad geográfica, colector, fecha y nombre, pero se reconoce que esa información contiene cierto grado de incertidumbre taxonómica y geográfica. A pesar de sus limitaciones, tales datos han constituido la base para muchos estudios biogeográficos clásicos, y son el campo común donde la sistemática, la biogeografía y a veces la ecología contribuyen a la biología de la conservación. Algunos resultados básicos que producen los datos de etiquetas de ejemplares en herbarios y museos son listas de taxones y áreas de distribución (ocupación) de taxones. Ambos tipos de información biogeográfica son esenciales para aplicaciones prácticas de políticas de conservación. Adicionalmente, la información de las etiquetas tiene importancia de tipo histórico, y su uso puede ser extremadamente útil para planeación y monitoreo ecológico (Soberón *et al.*, 1996).

En México, poco más del 65% del material depositado en las colecciones está incorporado a colecciones de consulta, pero su accesibilidad aún es muy reducida. Los propios especialistas se ven a veces abrumados por la magnitud de la información de su colección al no contar con herramientas que les permitan un manejo ágil y eficiente de la información (Llorente & Koleff, 1997; Llorente *et al.*, 1999).

De los especímenes albergados en las colecciones, en algunos casos se pueden desprender datos biográficos que están aunados a los datos de sus etiquetas, pues se encuentran parte de los itinerarios, y se puede contar con notas adicionales a partir de las libretas de campo de los colectores.

Bibliografía. Los medios impresos proporcionan datos de alta confiabilidad en la aportación de datos crudos, entre ellos están las monografías, revisiones, descripciones originales y redescrpciones, listas de referencia de especies (*check-list*), catálogos, y otros. Para una definición de estos tipos de trabajos puede consultarse el libro de Papavero & Llorente (1999).

Trabajo de campo. El trabajo de campo actual se encuentra bastante transformado ya que no está sometido a las limitaciones que existían en el pasado. Se han desarrollado técnicas, instrumentos y metodologías que permiten realizar muchos tipos de estudios, entre los que se encuentran las tradicionales exploraciones y trabajos florísticos y faunísticos, así como los modernos trabajos de distribución local bajo un parcelamiento ecológico o geográfico a distintas escalas y extensiones, o bien con transectos altitudinales.

Los datos obtenidos de las tres fuentes anteriores deben interpretarse para poder emplearse en conjunto. Actualmente la compilación de grandes cantidades de datos y su integración son posibles gracias a los desarrollos contemporáneos en los sistemas de información. Si es factible realizar lo anterior, entonces podemos hablar de una cuarta fuente de datos: las fuentes electrónicas, que incluirían en primer término bases de datos dentro de un sistema de información, discos y otros dispositivos de almacenamiento, *CD-ROMs*, y redes electrónicas (Figura 2).

Bases de datos y Sistemas de Información

Antes de enfocarse a la idea de un sistema de información, es necesario entender algunos conceptos básicos, que a continuación se expresarán.

Un dato es la representación simbólica de cualquier fenómeno que el ser humano puede conceptualizar, es

decir, se refiere a hechos concernientes a personas, objetos, eventos, u otros (Koleff, 1997). En una etiqueta de un espécimen de una colección, ejemplos de datos pueden ser el nombre que le fue asignado al realizar la determinación, la localidad en donde fue recolectado, la vegetación que lo rodeaba, entre otros muchos datos.

La información puede considerarse un conjunto de datos que al analizarlos nos permite tomar decisiones. Estos datos, que han sido procesados y presentados de una forma adecuada para su interpretación, frecuentemente tienen el propósito de revelar tendencias o descubrir patrones (Koleff, 1997). A partir de los datos de localidades de recolecta de varias etiquetas de especímenes y con otros datos de la literatura y estudios en campo, es posible proponer el área de distribución de un determinado taxón o su probabilidad de aparecer cuando existan determinadas condiciones ambientales (distribución potencial), esto es un constructo teórico y metodológico (Vuilleumier, 1999).

Entonces, una base de datos es un conjunto de información relacionada con un tema o propósito particular, la cual, si está bien estructurada y es precisa, puede recuperarse posteriormente y, en general se maneja a modo de archivos computarizados en uno o más sitios (Microsoft Corporation, 1994; WCMC, 1997). Dentro de una base de datos, los datos pueden compartirse por diferentes usuarios y mantenerse independientes de sus aplicaciones, también su definición y descripción deben almacenarse junto con ellos. Los procedimientos de actualización habrán de ser capaces de conservar la integridad, la seguridad y la confidencialidad del conjunto de datos (Escalante & Rodríguez, 1998).

Se han identificado principalmente dos tipos de bases de datos útiles para el trabajo de los taxónomos: las curatoriales (referentes a los especímenes de las colecciones) y las taxonómicas (referentes a las especies). Entre las bases de datos auxiliares de las taxonómicas y las curatoriales se encuentran las geográficas, las cuales contienen información sobre localidades de recolecta. A partir de las bases de datos curatoriales se pueden generar otras como las biogeográficas, que contendrían información acerca de la distribución de las especies con respecto a sistemas de clasificación biogeográfica establecidos y cartografiados (Peláez, 1994). Otra base de datos que siempre es importante pero puede ser parte de una taxonómica es la referente a la bibliografía.

Los sistemas de bases de datos pueden clasificarse de acuerdo con las estructuras de datos y los operadores que se presentan a los usuarios. En primer término, los sistemas más

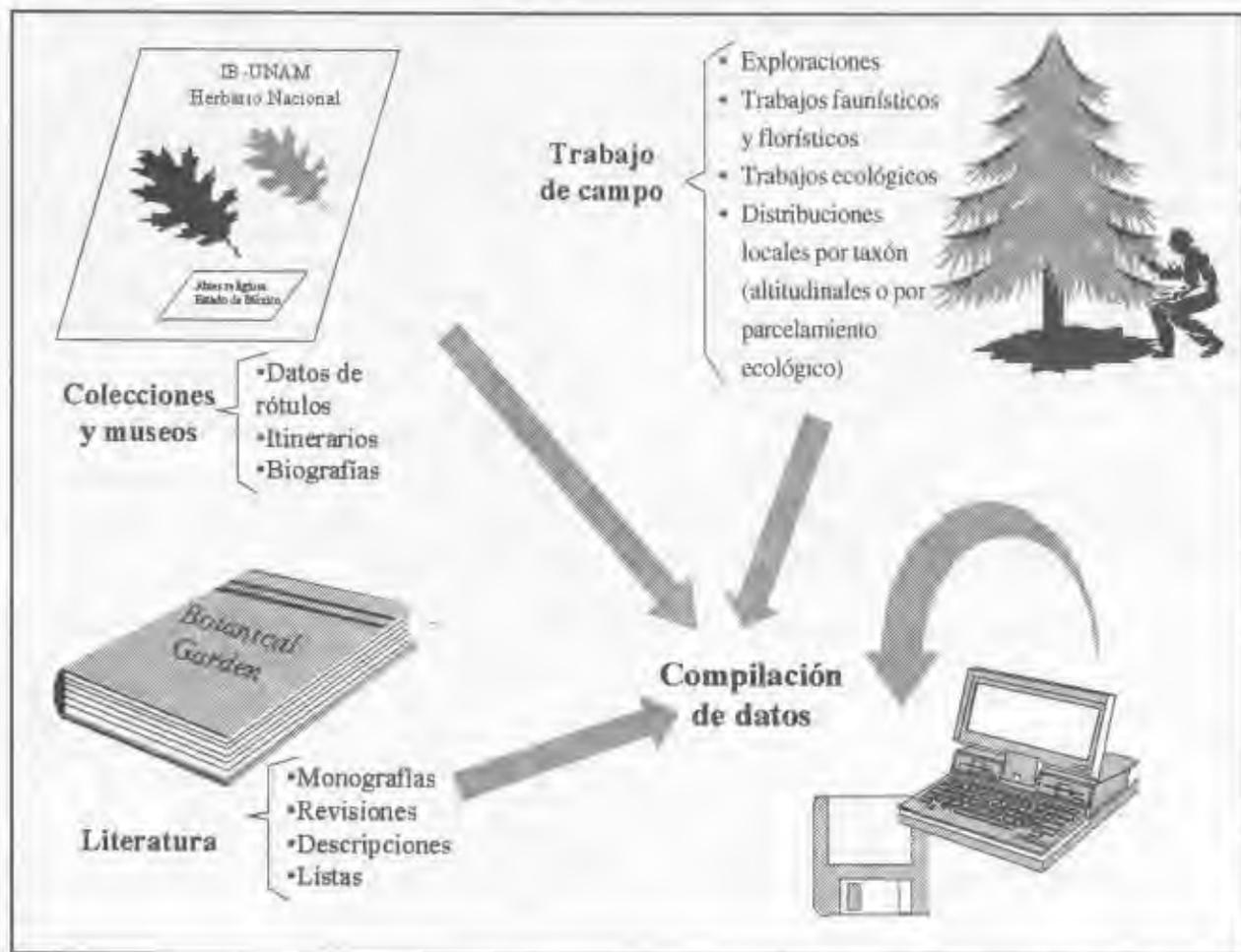


Figura 2. Principales fuentes de datos biológicos en la actualidad.

antiguos entran en tres categorías amplias: de lista invertida, jerárquicos y de red. En fechas más recientes se reconocen los sistemas relacionales y otros tales como los sistemas orientados a objetos (Date, 1998). Algunos de ellos ya se han aplicado al desarrollo de sistemas de cómputo para colecciones biológicas, pero indudablemente los sistemas relacionales son los más utilizados. En adelante, únicamente se hará referencia a bases de datos de tipo relacional.

Un sistema relacional es aquel en el cual el usuario percibe los datos únicamente en forma de tablas, y donde los operadores al alcance del usuario generan tablas nuevas a partir de las existentes (Date, 1998). Como característica adicional, las bases de datos de tipo relacional generalmente están compuestas de dos o más tablas que se comunican por medio de igualdades. Las tablas de las bases de datos relacionales están constituidas por dos elementos principales: registros o renglones, y columnas. En

el caso de bases de datos de especímenes, debe entenderse por registro uno o varios ejemplares recolectados en la misma fecha, en la misma localidad (aquel punto geográfico que tiene la misma georreferencia y nomenclatura), por la misma persona, que corresponden al mismo taxón, y que, en un momento dado, hayan sido incorporados dentro de un catálogo con el mismo identificador (por ejemplo: número de catálogo cronológico de la colección o museo).

De esta manera, un sistema de información se puede definir como un conjunto estructurado de procesos, personal, *hardware* y *software*, para convertir datos en información (WCMC, 1997). Es el medio por el cual los datos fluyen de un usuario a otro, además proporciona servicios a todos los demás sistemas de una organización y enlaza todos sus componentes en forma tal que éstos trabajen con eficiencia para alcanzar el mismo objetivo.

Sistemas de Información sobre Biodiversidad

Desde hace cerca de 20 años, se han realizado intentos en todo el mundo para elaborar Sistemas de Información sobre Biodiversidad (SIB) (ver **Berendsohn**, 1999). **Soberón & Koleff** (1998) presentan los factores críticos en el establecimiento de un SIB:

1. Los proveedores de los datos. Se incluyen a los grupos de personas que generan y proporcionan los datos "crudos" o agregados de datos de cualquiera de los niveles de biodiversidad (geógrafos, taxónomos, ecólogos, genetistas, forestales, físicos tradicionales, vendedores de productos naturales y otros).
2. Usuarios potenciales y actuales de la información. Las salidas de un SIB deben definirse en conjunto con los principales usuarios del sistema, es decir, la manera en como deban presentarse los resultados dependerá

del usuario al que son proporcionados y el fin que vayan a satisfacer.

3. Especificaciones técnicas. Dadas sus características multiescalares, los datos pueden aparecer en un sinnúmero de formatos (imágenes, archivos de texto, archivos de información genética, u otros). Así, la estructura de un SIB debe responder a la complejidad y magnitud del problema que se requiera y se presente. Según **Davis et al.** (1990) un SIB debe direccionarse a contestar las siguientes clases de preguntas:

-¿Cuál es el área de una especie o comunidad dadas?

-¿Cuáles especies tienen alto riesgo?

-¿Cuáles sitios tienen la más alta riqueza de especies?

-¿Cuáles sitios biológicamente importantes tienen alto riesgo (por ejemplo: "hot spots")?

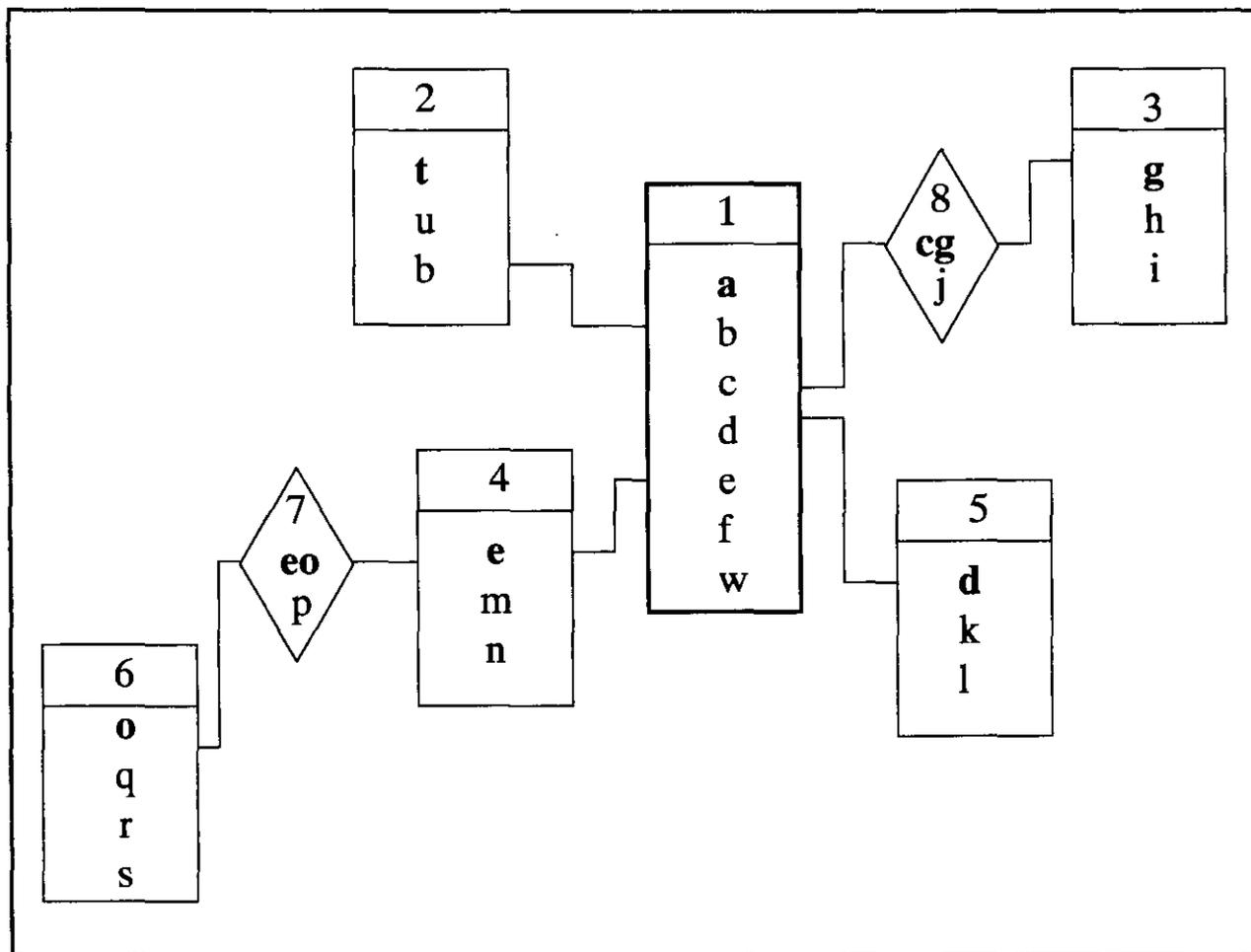


Figura 3. Esquema que representa un sistema de bases de datos de tipo relacional.

- ¿Cuáles ecosistemas están adecuadamente protegidos?
- ¿Qué factores ambientales están relacionados a sitios de alta diversidad? ¿Puede predecirse la diversidad sobre la base de esos factores?
- ¿Dónde puede ocurrir desarrollo sustentable ecológicamente con impacto aceptable sobre la biodiversidad?
- ¿Cuáles son las tendencias actuales de las especies y diversidad de ecosistemas?

Uno de los criterios importantes para caracterizar un SIB es si está basado en datos “crudos” o “atómicos” o sobre información interpretada. Un proyecto de un SIB hipotético, capaz de resolver preguntas de todo tipo sobre biodiversidad se muestra en la figura 4. El elemento

núcleo de los datos (o dato atómico) es el espécimen (Soberón & Koleff, 1998).

Generalmente, los SIB almacenan datos cuya estructura de información está organizada en forma de bases de datos que pueden ser centralizadas o distribuidas, según una arquitectura determinada. En una base de datos centralizada, todos los datos están localizados en un sitio único y los usuarios en sitios remotos pueden tener acceso a ella utilizando facilidades de comunicación (por ejemplo, redes). En cambio, una base de datos distribuida se encuentra dispersa físicamente a través de computadoras en múltiples lugares (McFadden & Hoffer, 1991).

Davis *et al.* (1990) identificaron cuatro aplicaciones primarias de un SIB:

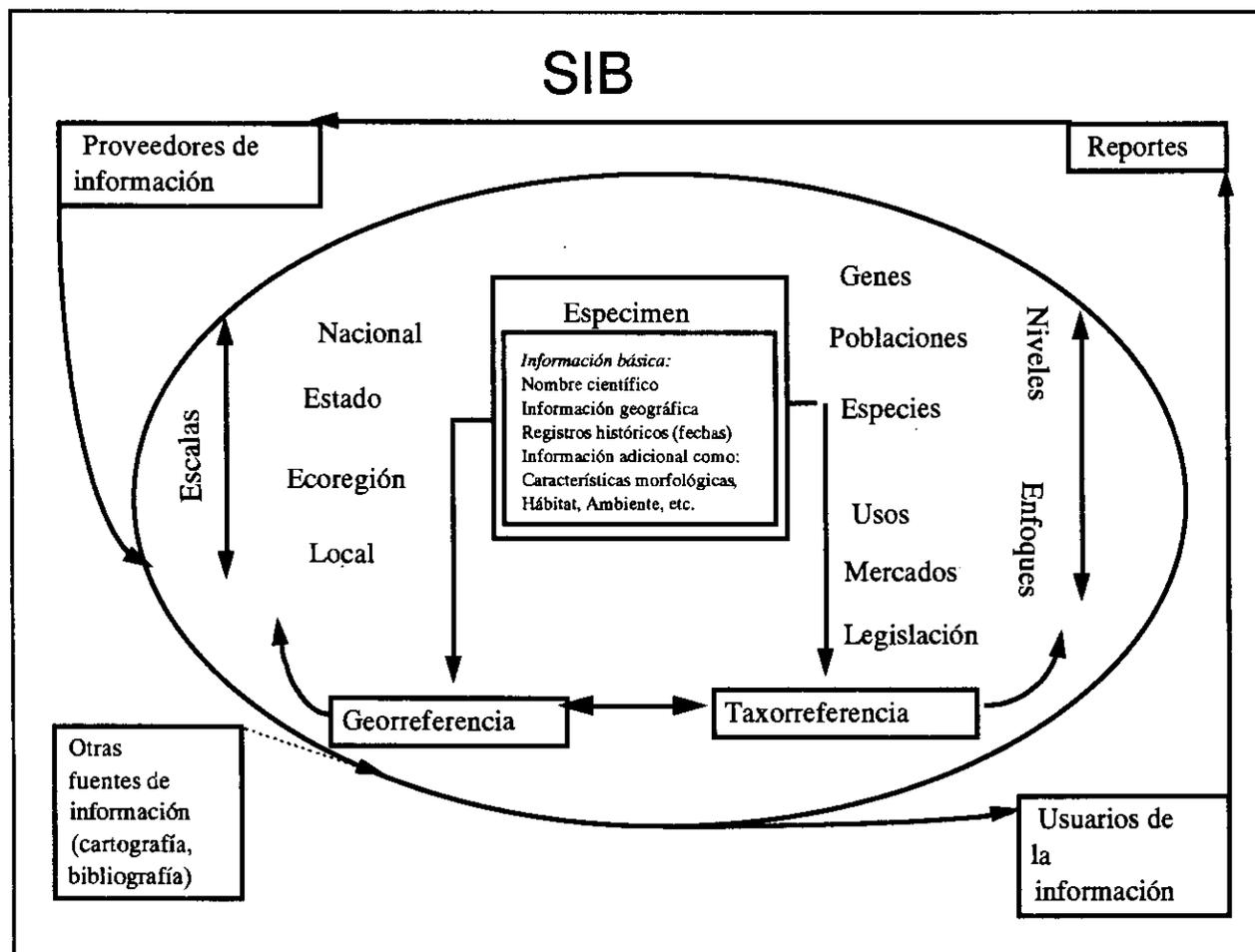


Figura 4. Proyecto de un Sistema de Información de Biodiversidad ideal, capaz de responder todo tipo de preguntas sobre biodiversidad (Proporcionado por Soberón y Koleff, 1998).

1. Evaluación y análisis de la biodiversidad. Incluye el solapamiento automatizado de mapas de riqueza de especies y otros temas como vegetación, uso de suelo y sustentabilidad para determinar la extensión de la pérdida de la biodiversidad.
2. Análisis de discontinuidades y conservación. Una estrategia ampliamente utilizada para conservar la diversidad biológica es localizar tierras para reservas naturales (ver más adelante).
3. Impacto de desarrollo. Ayudar a la evaluación de los efectos sobre la diversidad biológica en proyectos de impacto ambiental.
4. Investigación científica. Un SIB general debe dar la invaluable ventaja a los investigadores científicos, para minimizar el tiempo de análisis de datos, contrastar teorías y probar hipótesis de localización de riqueza o endemismo, así como la verificación y la actualización de datos.

La importancia de construir un SIB de envergadura nacional, y alcanzar una escala global, ha sido discutida en foros de todo el mundo, durante los últimos seis años, como el llevado a cabo en 1994, en relación con la Agenda Sistemática 2000. Entre las misiones que se propusieron en ese entonces, cabe destacar la tercera: "organizar la información derivada de este programa global en una forma de recuperación eficiente, con el fin de satisfacer mejor las necesidades de la ciencia y de la sociedad". Así, los beneficios de un sistema de información sistemática eficiente incluyen (*Systematics Agenda 2000*, 1994):

1. Permitir a quienes formulan las políticas aplicar decisiones con base en una mejor información en relación con el manejo sustentable de los recursos.
2. Proveer una mejor documentación sobre la extinción y los cambios en la distribución de las especies.
3. Manejar los recursos biológicos con un mayor factor costo-beneficio, dado que las bases de datos electrónicas proveen una comunicación eficiente de la información sistemática y asociada.
4. Proporcionar un acceso rápido al conocimiento sistemático para la solución de problemas.
5. Facilitar nuevas clases de comparaciones y asociaciones entre datos biológicos y de otras fuentes, especialmente en biotecnología.
6. Mejorar la comunicación y colaboración global y reducir la duplicación de esfuerzos científicos.

Sistemas de Información Geográfica

Desde su origen, los Sistemas de Información Geográfica (SIG) han sido considerados como un caso especial dentro de los sistemas de información general (López, 1998). Un SIG es un conjunto organizado de conocimientos sistemáticos sobre las diversas formas de integrar y analizar la información acerca de entidades geográficas (Maguire *et al.*, 1991). También se ha definido como una colección organizada de *hardware*, *software*, y datos geográficos, diseñada para captura, almacenamiento, actualización, manipulación o simulación, análisis, y despliegue de todas las formas de información geográficamente referenciada (ESRI, 1996).

Los principales componentes de un SIG se pueden agrupar en tres conjuntos (Burrough, 1989; López, 1998), figura 5:

1. Elementos físicos (*hardware*)

La unidad central de procesamiento está ligada a una unidad de almacenamiento, el cual proporciona espacio para guardar los datos y el *software*. Un digitalizador u otro instrumento es utilizado para convertir datos de mapas y documentos a formatos digitales y enviarlos a la computadora. Un *plotter* es usado para presentar los resultados del procesamiento de los datos, y una unidad de cinta guarda los datos o programas en cintas magnéticas, o para comunicarse con otros sistemas. El usuario controla la computadora y sus periféricos a través de una unidad de despliegue visual.

2. Elementos lógicos (conjunto de programas o *software*)

Los programas individuales permiten realizar una secuencia de funciones para obtener resultados del análisis de la información. Cada uno de esos programas individuales permite realizar un conjunto de órdenes estructuradas lógicamente que facilitan la solución de consultas (*queries*). Existen cinco grupos de funciones o subsistemas que se realizan:

- (a) Entrada y verificación de los datos.
- (b) Almacenamiento de los datos y manejo de la base de datos.
- (c) Salida y presentación de los datos.
- (d) Transformación de los datos.
- (e) Interacción con el usuario.

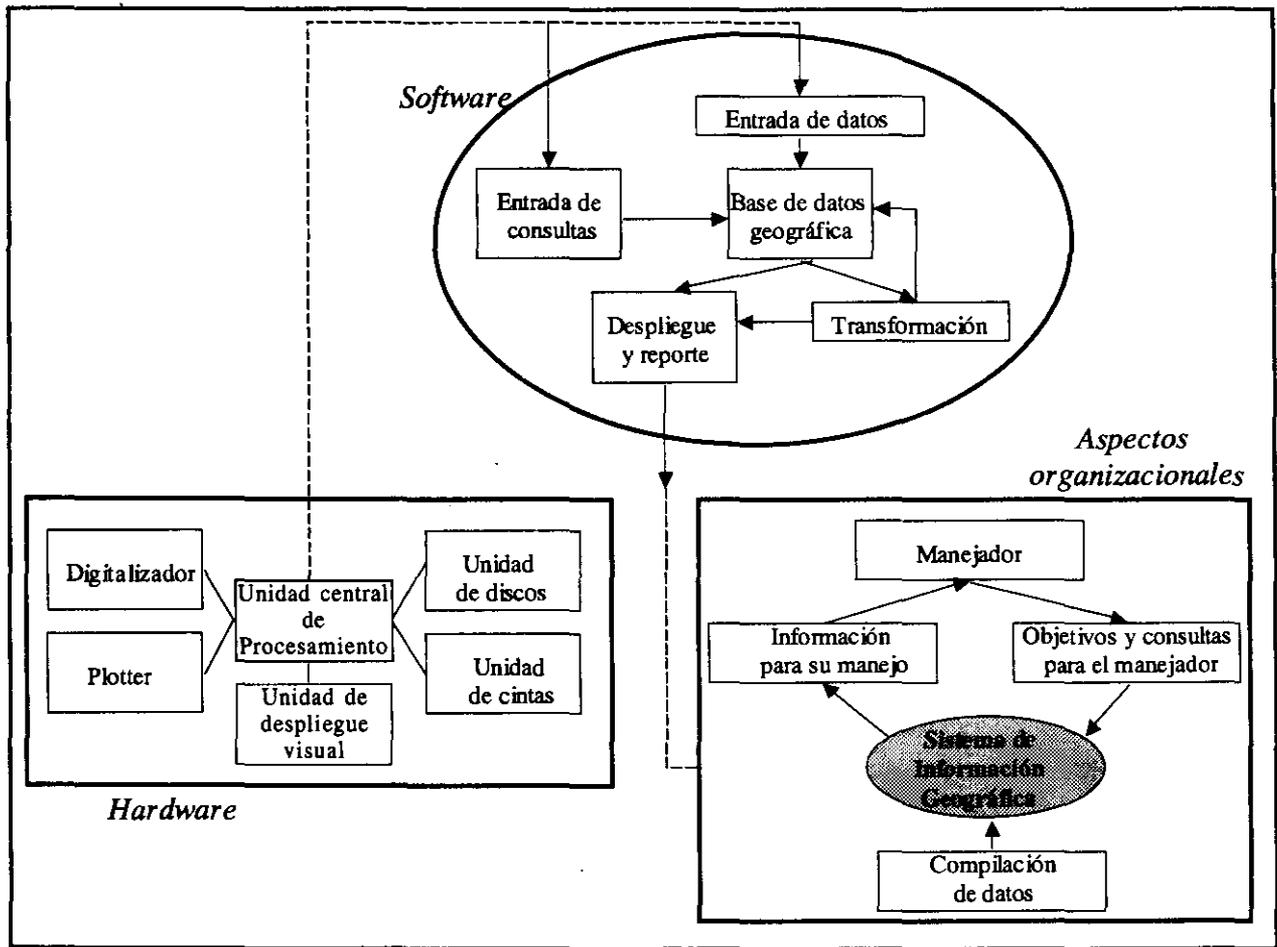


Figura 5. Componentes de un SIG: elementos físicos (*hardware*), elementos lógicos (*software*), y aspectos de organización (Redibujado de Burrough, 1989).

3. Aspectos de organización

Los cinco subsistemas técnicos anteriores determinan la forma en la cual la información geográfica puede procesarse, pero no por sí misma garantizan que sean efectivamente usados. Un SIG necesita estar localizado en un contexto organizacional adecuado, debe considerarse especialmente el entrenamiento del personal y los manejadores que usarán esta tecnología, así como modificaciones a la estructura institucional en donde se inicie el empleo de los SIG.

La tecnología de los SIG frecuentemente se ha confundido con otras ciencias y tecnologías que manejan datos espaciales, tales como percepción remota, cartografía, geodesia, fotogrametría, y, por supuesto, la ciencia "padre" de los SIG, la geografía. Como un ejemplo, se define a la percepción remota como una tecnología de

fuente de datos, y a un SIG como una tecnología de procesamiento de datos (Parker, 1988).

Las posibilidades de aplicaciones de los SIG son muy amplias. Los límites más comunes son los que dan la profundidad de los conocimientos, la claridad de los objetivos, la experiencia, el ingenio, la imaginación y la creatividad del grupo de usuarios que trata de aplicar dicha tecnología (López, 1998). El desarrollo y uso de los SIG pueden facilitar problemas de integración de datos y el consumo de tiempo que toma el procesamiento de síntesis de enormes cantidades de información para el examen de datos espaciales (Walsh, 1985).

Los SIG se caracterizan por los elementos espaciales de entrada de datos: puntos, líneas o polígonos que representan hechos geográficos que deben convertirse a un formato que pueda ser almacenado, manipulado, y

desplegado por un sistema de información basado en computadoras. La información espacial representada en un mapa primero debe transformarse dentro de un arreglo de coordenadas *X*, *Y* y *Z* que describen los datos (Walsh, 1985). Sobre un mapa plano, las coordenadas *X* y *Y* se refieren a una localización que corresponde a la misma localización sobre la superficie esférica de la Tierra. Así, el eje *X* representa la dirección este-oeste, y el eje *Y* o vertical la norte-sur (ESRI, 1996). La coordenada *Z* indica el valor de la celda dentro de una retícula sobre un mapa plano (Walsh, 1985).

Un mapa representa gráficamente una porción de un espacio real sobre la superficie de la Tierra, entonces, las coordenadas *X* y *Y* corresponderán a la georreferencia o referencia geográfica de un determinado elemento (por ejemplo, las coordenadas geográficas expresadas en latitud y longitud respectivamente). Sin embargo, la georreferencia está compuesta de un tercer elemento, la altitud, que no necesariamente corresponde a la coordenada *Z* en un mapa plano de retícula. Para que este dato sea empleado de la mejor manera, es necesario recurrir a otra forma de representación de datos geográficos, los modelos digitales de terreno.

Control de calidad de los datos: Archivos de autoridades

Una tarea fundamental dentro de un sistema de información, es la necesidad de llevar a cabo un control de calidad de los datos. La información de los especímenes en las bases de datos es obtenida de fuentes muy diversas, así, es necesario el proceso de control de calidad porque los datos presentan numerosos problemas, tales como determinaciones mal realizadas, inestabilidad taxonómica y nomenclatural, entre otros. A menudo, el proceso de control de calidad genera bases de datos más pequeñas pero más confiables (Soberón & Koleff, 1998; Soberón *et al.*, 2000).

Cuando existen errores en los datos, la computadora generalmente los puede detectar como series de caracteres diferentes, lo que ocasiona que se "duplique" la información. El corregir estos errores, e incluso visualizarlos, es difícil sin la consulta de índices profesionales reconocidos, bibliografía especializada, u otros estándares o patrones, sean nacionales o internacionales. Entonces, la solución es contar con diccionarios (estándares) incluidos en la base de datos a los cuales se pueda recurrir cada vez que se captura un registro nuevo, o cuando se tiene una base de datos elaborada y se requiere verificar la información (Escalante & Rodríguez, 1998).

Por lo tanto, para ser utilizados, los datos deben confrontarse con estándares, los cuales juegan un papel crucial en la actividad coordinativa (Green, 1994). Sin embargo, debe tenerse cuidado que la "homogeinización de los datos" elimine información de valor, el proceso de uso de diccionarios debe permitir comunicación. Los estándares son mecanismos usados en la recolección, manejo e intercambio de datos. Los beneficios de usar estándares son (WCMC, 1997):

1. Obtención de mayor consistencia, consolidación y comunicación.
2. Incremento en la eficiencia del conjunto de datos.
3. Menores costos al realizar transacciones de intercambio de datos.
4. Los datos pueden utilizarse en diversas aplicaciones.

El desarrollo de estándares generalmente requiere esfuerzos interdisciplinarios para revisar, formalizar y publicar los estándares, además deben establecerse procesos de revisión de los mismos (WCMC, 1997). No se trata de uniformizar un lenguaje sin considerar la naturaleza de los errores en la toma e interpretación de datos taxonómicos y geográfico-ecológicos; así, debe ponerse especial atención en el desarrollo de estándares.

Existen dos clases principales generales de estándares según Green (1994):

Atributos estándares: Se refieren a la información que es esencial para todos los datos.

Estándares de control de calidad: Son indicadores de validación y precisión de los datos, como referencias a glosarios o autoridades usadas para los nombres y códigos que incluyan las clases de error en el ingreso de los datos.

Se reconocen al menos dos tipos de estándares fundamentales para realizar el control de calidad de los datos de ejemplares: taxonómicos o nomenclaturales, y geográficos. Dentro de los primeros se encuentran los *archivos o diccionarios de autoridad taxonómica* (nomenclaturales), y de los segundos se consideran revisiones de nomenclatura de localidades y rutinas de validación de georreferencia. Adicionalmente, los diccionarios nomenclaturales pueden estar ligados a archivos bibliográficos, que de acuerdo con los códigos de nomenclatura vigentes regulan la validación y aceptación de nombres.

Los archivos o diccionarios de autoridad taxonómica, son listas de los nombres aceptados de un taxón, es decir,

cir, son archivos que contienen la información jerarquizada dentro de un sistema de clasificación para un determinado grupo de organismos. El diccionario consiste de los nombres correctos de los taxones de todas las categorías taxonómicas. En el caso de los nombres de las especies debe incluir además la información del autor y año (Escalante & Rodríguez, 1998). La idea de computarizar diccionarios taxonómicos e información de colecciones científicas fue desarrollada formalmente desde 1970. A partir de entonces ha habido una diversidad de iniciativas de diferentes instituciones en el mundo, y en México se han comenzado a desarrollar y aplicar ampliamente con éxito. Con tales diccionarios es posible detectar gran número de inconsistencias nomenclaturales, como son la escritura incorrecta (errores de "tecleo") y sinonimia.

El control de calidad geográfica incluye la revisión de los nombres de las localidades y la georreferencia de las mismas. Los nombres de localidades presentan algunos problemas para poder utilizarse en análisis posteriores. Papavero & Llorente (1999) han identificado algunos de ellos:

-no se tiene indicación alguna de la localidad; lo cual es muy frecuente en localidades tipo,

-para localidades antiguas, la citación es extremadamente vaga,

-la localidad dejó de existir y ya no aparece en mapas actuales,

-la localidad cambió de país y puede encontrarse en la literatura antigua,

-la localidad viene mal indicada, por insuficiencia de datos, o por estar mal escrita o transcrita,

-la localidad cambió de nombre,

-existen nuevas localidades homónimas en la misma área, y

-la localidad citada es tan pequeña que no se encuentra en diccionarios geográficos, listas toponímicas, mapas, u otros medios.

Para verificar la nomenclatura de localidades se cuenta con atlas, gaceteros o nomenclatores, cartografía a diferentes escalas, listas municipales y, en ocasiones, archivos electrónicos de autoridades con esta información. Todos ellos pueden permitir análisis de consistencia-inconsistencia de los datos geográficos, esto es su validación.

En la actualidad, la verificación de la georreferencia se puede llevar a cabo con rutinas sistematizadas de validación. Existen SIB que incluyen tales rutinas como Biotica⁴, desarrollado en México por la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO, 1998), donde se realizan rutinas de validación nacional y estatal. La validación por estado implica que se toman los sitios (definidos por sus coordenadas X, Y) de la base de datos, y se sobreponen a un tema de estados. El sistema entrega el número de sitios que indican un estado y sobre el mapa se encuentra en otro (número de inconsistencias), así como los sitios en los que se encontró correspondencia (consistencias) (CONABIO, 1998).

Además de los controles de calidad ya mencionados, debe considerarse un caso interesante de verificación de datos que se presenta en los datos llamados "históricos", es decir, datos recabados en el pasado. Algunos de ellos datan de dos siglos atrás y, por lo tanto, la forma en que eran recolectados dista mucho de parecerse a la actual, que es más organizada y sistematizada de acuerdo con los estándares actualmente establecidos. Los nombres de los taxones probablemente no estaban bien definidos, no había los Códigos de Nomenclatura que hoy conocemos, ni referencias a autoridades ni a sistemas de clasificación. Otros datos de las etiquetas con frecuencia eran omitidos, como fechas y tipos de vegetación o asociaciones, así como nombre del recolector y las iniciales que los acompañaban.

Los datos histórico-geográficos también carecían de referencia geográfica, o si existía, era de muy dudosa exactitud. Algunos podían provenir de localidades hoy alteradas o transformadas, lo cual puede significar que sólo son valiosos si se reconoce su contexto ecológico de origen, también algunos nombres de localidades están en desuso, se han transformado lingüísticamente o son de difícil ubicación geográfica por su historia accidentada (Llorente *et al.*, 1997). Papavero & Llorente (1999) presentan algunas formas de descubrimiento de las localidades-tipo antiguas o de las modernas pero imprecisas, como son el examen directo del material-tipo, el examen de los libros de registro (catálogos cronológicos), las libretas de campo o bitácoras, el levantamiento del itinerario de viaje del naturalista, incluso enciclopedias y bibliografías de viajes de exploración naturalista. Para localidades imprecisas también se puede recurrir al propio colector (si aún está vivo) o de la institución en donde trabajó, a los informes anuales publicados por dichas instituciones, e incluso, consultando a otros especialistas. Todo esto es, en buena parte, hacer historia de la historia natural, que puede ser de gran valor para rescatar datos aparentemente perdidos.

Aplicaciones de las bases de datos en biogeografía

La habilidad para reunir grandes bases de datos, junto con las herramientas computacionales requeridas para analizarlos, está incrementando el valor de las colecciones, museos, y el trabajo de los taxónomos alrededor del mundo. Las aplicaciones científicas de las bases de datos son de gran interés, como las implicaciones prácticas para prospección de especies importantes, restauración y reforestación, planeación ecológica, localización de áreas prioritarias a la conservación, y monitoreo o evaluación de estudios de impacto ambiental, entre otras (Soberón *et al.*, 1996).

Las bases de datos taxonómicas no sólo proporcionan acceso rápido a una cantidad sin precedente de información de interés para el sistemático, sino que también pueden usarse en el trabajo de ecólogos, biogeógrafos y para otros propósitos o aplicaciones (Soberón *et al.*, 2000_a). Una de las partes más gratificantes de un trabajo taxonómico es la obtención de patrones de distribución, que sirven de base para interpretar la historia biogeográfica del grupo. Para realizar esto, es necesario contar con la mayor cobertura geográfica posible, es decir, examinar el mayor número posible de ejemplares, al estudiar el mayor número posible de colecciones (Papavero & Llorente, 1999_a), y la forma moderna de almacenar, recuperar y analizar la información recabada son las bases de datos. En principio, las bases de datos de especímenes deben ser capaces de resolver dos preguntas relacionadas, las cuales son centrales en biogeografía y macroecología: (1) ¿Qué especies se encuentran en una localidad dada?, y (2) ¿Cuál es la distribución geográfica de cada especie? Sin embargo, no hay procedimientos universalmente aceptados para evaluar las bases de datos de especímenes en cuanto al esfuerzo de recolecta que podría ocultar los patrones reales y evita contestar esas preguntas (Soberón *et al.*, 2000_a).

Se han realizado algunos trabajos para analizar bases de datos desde la perspectiva de la debilidad de su uso para dos importantes objetivos de conservación: la obtención de listas de especies y la estimación de áreas de distribución (Soberón, *et al.*, 2000_a). Otros estudios pretenden conocer la representación de especímenes de un determinado grupo en un área geográfica. En México se han llevado a cabo análisis para conocer la proporción de recolección de cada estado o a nivel nacional, con la finalidad de hacer comparaciones. Se ha encontrado que conviene utilizar cocientes del número de individuos entre el número de localidades o bien el número de registros entre el número de localidades, ya que esto consideraría

qué tan ampliamente distribuidas están en el espacio geográfico y, en términos temporales, cómo se manifiestan las recolecciones. Si uno quiere perfeccionar tales índices, sería recomendable considerar la complejidad geográfico-ecológica, tomando en cuenta valores de variabilidad climática, vegetacional y fisiográfica para un estado determinado, tal vez un factor sobre ello afinaría el índice, lo cual puede lograrse si se tienen buenos mapas digitalizados de tales componentes geográficos (Peterson, *et al.*, 1996; Llorente *et al.*, 1997; Soberón *et al.*, 2000_a; Soberón *et al.*, 2000_b).

Un ejemplo de bases de datos de ejemplares se encuentra en el libro de Llorente *et al.* (1997), donde se conjuntaron 55,000 ejemplares de mariposas, de los cuales 13,807 provenían de la literatura. En el caso de esta fuente de datos deben tomarse en cuenta las implicaciones que tiene, ya que con frecuencia se sigue un criterio de conteo mínimo de ejemplares al incorporar los datos de la literatura en la base de datos, es decir, si no se indicaba la cantidad de ejemplares, sólo se registraba uno en la base de datos, lo cual implicaba una estimación mínima, pero al fin y al cabo una subestimación. Por otra parte, también debe advertirse que es posible que algunos datos registrados de ejemplares en la literatura, a menudo están depositados en colecciones que pudieron haber sido consultados (Llorente *et al.*, 1997), lo cual conduce a redundancia de ejemplares, pero no necesariamente de registros.

Aplicaciones de los SIG en biogeografía

Como se vio anteriormente, las bases de datos son un elemento constituyente de los SIG, lo cual implica más opciones de análisis de los datos cuando ya se cuenta con ellas.

La tecnología de los SIG está revolucionando la elaboración de mapas de áreas ocupadas ("puntos" georreferidos con latitud, longitud y altitud) y su trazo o interpretación en áreas de distribución. La incorporación de técnicas areográficas sistematizadas electrónicamente pronto será de uso cotidiano; en la actualidad, la oferta de programas computacionales para el diseño de mapas biogeográficos que ilustren o grafiquen la distribución de los taxones, especies o poblaciones ya es notable. A través de estas herramientas se ganan precisión, rapidez y homogeneización, pues facilitan la comparación entre escalas, y dan la posibilidad de introducir varios parámetros geográficos (clima, suelo, vegetación, topografía, u otros) que contextualizan los limitantes ecológicos en las distribuciones (Papavero & Llorente, 1999_b).

El análisis espacial es la tarea más común que realiza un biogeógrafo al utilizar los SIG, el cual es el estudio de las localizaciones y formas de los elementos o características geográficas y las relaciones entre ellos (ESRI, 1996); dentro de los análisis espaciales encontramos el álgebra de mapas. Generalmente se realizan las operaciones de unión y solapamiento espacial. Se entiende como solapamiento espacial al proceso de superposición de capas (temas) de datos geográficos que ocupan el mismo espacio con el fin de estudiar las relaciones entre ellos; en la unión, los atributos de los elementos de dos diferentes temas se juntan o unen basándose en las localizaciones relativas de los mismos (ESRI, 1996). Ver figuras 6 y 7.

Al utilizar los SIG se han descubierto algunas cuestiones interesantes en los datos de los ejemplares que antes eran difíciles o casi imposible de detectar. Los fenómenos de apiñamiento de puntos en los mapas a causa de los

síndromes de carreteras, de los colectores y localidades clásicas, o de cercanía a ciudades, universidades o estaciones de campo, repercuten en la redundancia observada en muchos de los mapas y a la vez quedan muchas áreas vacías. No obstante, los puntos disponibles con su caracterización geográfico-ecológica podrían ayudar a determinar, en gran parte, las restricciones para definir las áreas potenciales o áreas de distribución (Llorente *et al.*, 1997). Las redundancias pueden verse como aparentes, pues reafirman observaciones de la valencia ecológica para el trazo de la distribución potencial.

Análisis de discontinuidades para conservar la biodiversidad

Los biólogos de la conservación por años han utilizado, consciente o inconscientemente, el proceso llamado *gap analysis* para establecer prioridades de conservación a corto o largo plazo (Burley, 1990). Para el manteni-

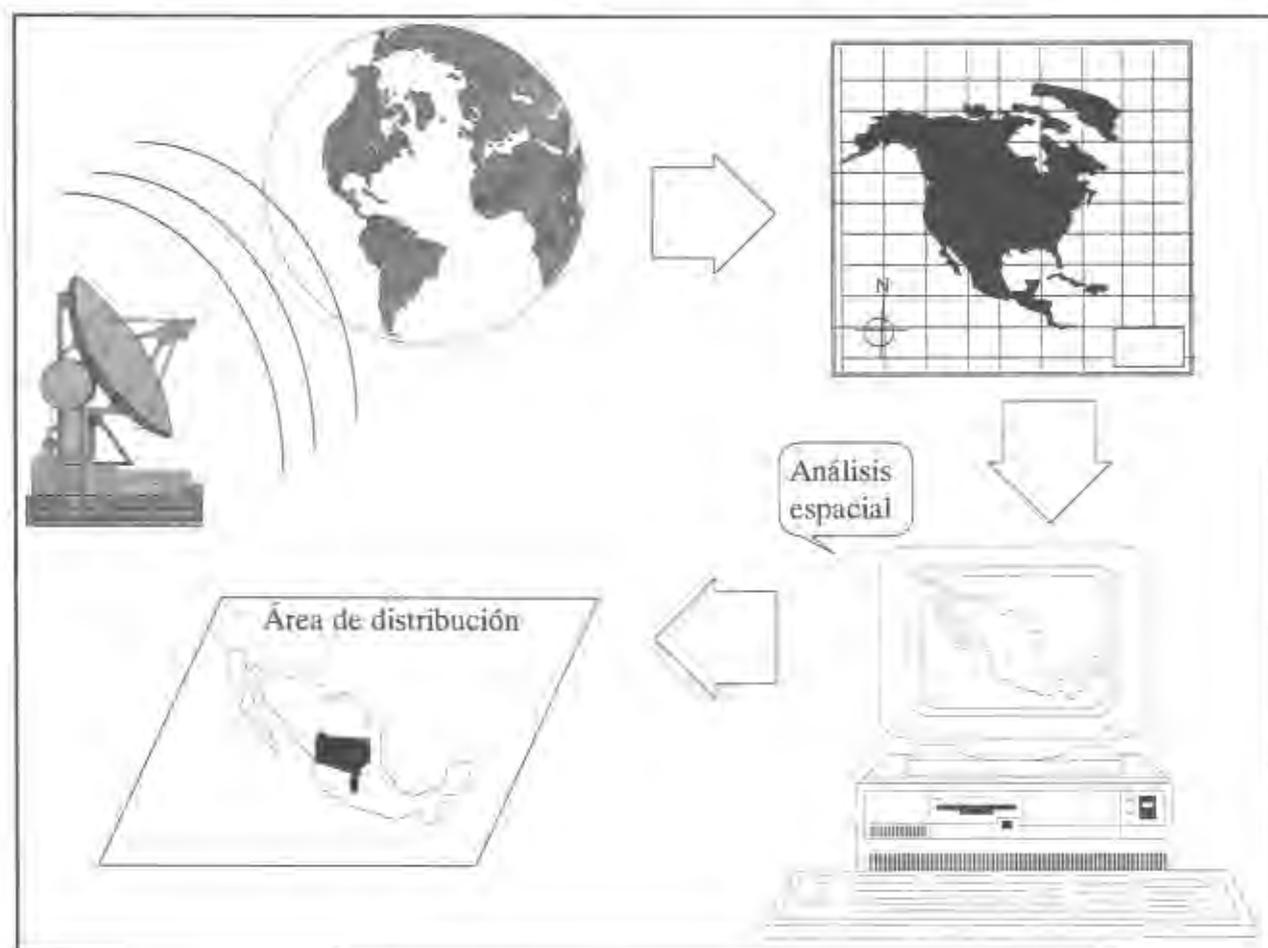


Figura 6. La tecnología de los SIG se ha aplicado al campo de la biogeografía para cartografiar la distribución de taxones.

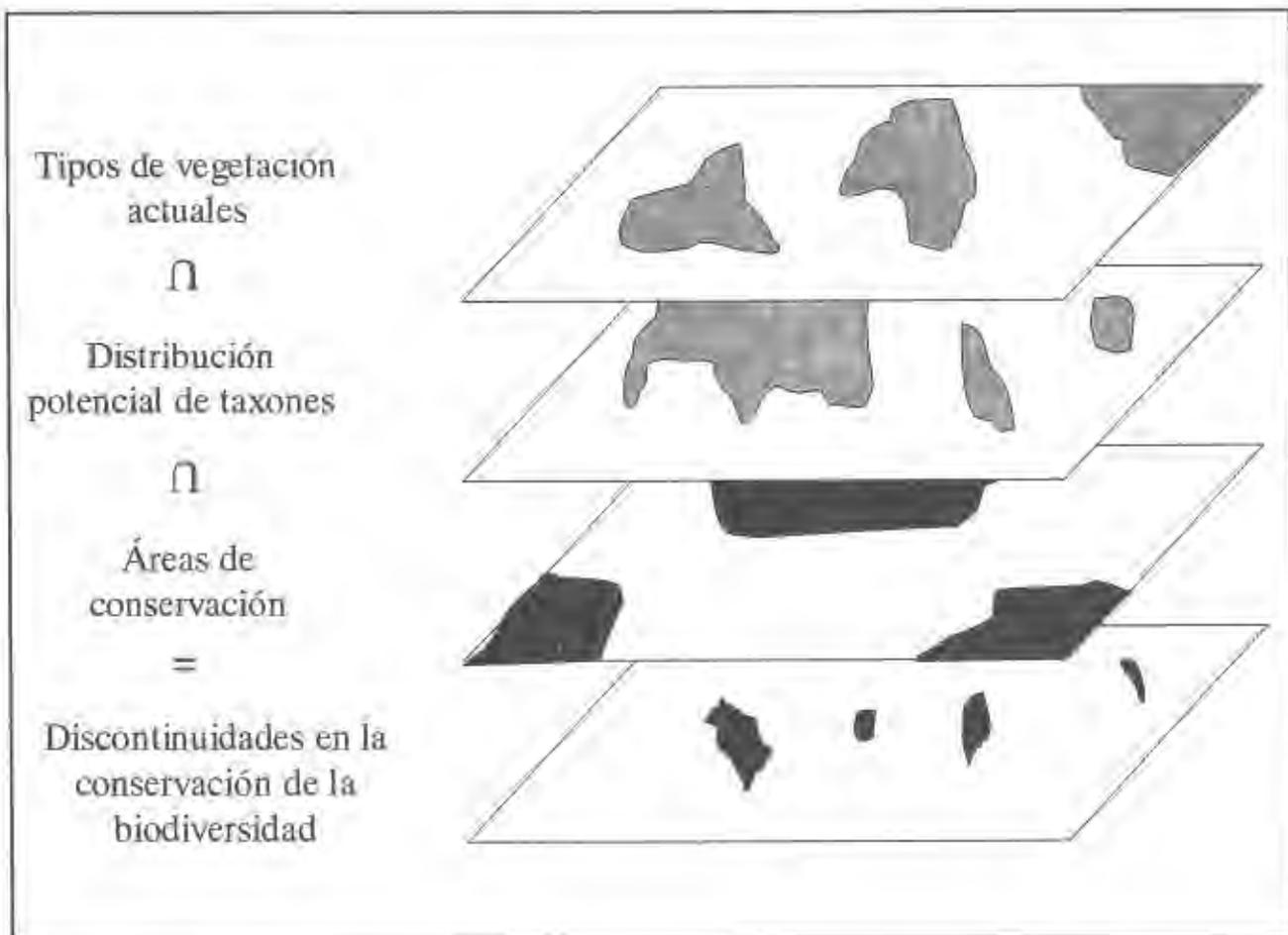


Figura 7. Análisis de discontinuidades (Redibujado de Upper Midwest Gap Analysis Home Page, 1998).

miento de la diversidad biológica, de manera convencional se ha procesado una especie cada vez. El *gap analysis* es una metodología para identificar "vacíos" o discontinuidades en la representación de la biodiversidad en áreas manejadas exclusivamente para un mantenimiento a largo plazo de poblaciones de especies nativas y ecosistemas naturales (Upper Midwest Gap Analysis Home Page, 1998).

El *gap analysis* (que aquí en adelante llamaremos análisis de discontinuidades o AD) inicialmente fue un programa del "National Biological Service" de los Estados Unidos de América, y actualmente su uso se ha difundido en todo el mundo. Utiliza dos tecnologías relativamente nuevas; detección por satélites remotos y SIG (figura 6), para asistir en la evaluación del estatus y distribución de varios elementos de la biodiversidad. Mientras no exista un sustituto para las medidas biológicas tradicionales, se piensa que el AD puede proporcionar una evaluación pa-

notámica preliminar de la distribución de diversidad de especies y biocenosis, el cual puede usarse para guiar investigaciones de campo futuras y proporcionar un sistema espacial para una estrategia nacional preliminar de conservación de la biodiversidad. El AD promete tener un desarrollo rápido de información sobre la distribución de diversos indicadores de biodiversidad en áreas específicas del mundo (Scott & Csuti, 1997).

El AD se efectúa por solapamiento de tres coberturas primarias de datos de SIG: cubierta de vegetación, mapas de distribución de especies, y mapas de posesiones públicas y de manejo para especies nativas; así, los vacíos en el manejo de la biodiversidad pueden identificarse. Estas capas de datos son desarrolladas, desplegadas y analizadas usando SIG. La meta del AD es garantizar que todos los hábitats o biomas, y las áreas ricas en diversidad de especies (riqueza y endemismo) estén adecuadamente representadas en la planeación, elección y manejo de la

biodiversidad. También puede usarse información de otros factores ambientales, tales como elevación, inclinación, aspecto, suelos, características acuáticas, y clima, con el fin de mejorar la precisión de los mapas de vegetación y de distribuciones de las especies. Igualmente, puede simularse o examinarse información adicional sobre atributos socioeconómicos de la tierra (por ejemplo: tendencia a proteger poblaciones, proyectos de vivienda iniciados, propietarios de tierras por estados y agencias federales, restricciones, etc.) para refinar los planes de manejo del uso del suelo (Scott & Csuti, 1997; Upper Midwest Gap Analysis Home Page, 1998). Ver figuras 6 y 7.

La metodología que sigue el AD es realmente simple: dentro de un país o región particular, primero se identifican y clasifican los diferentes elementos de la diversidad biológica. Entonces se examina la existencia de los sistemas propuestos de áreas protegidas y otras unidades de manejo de tierras que permiten conservar la biodiversidad. Finalmente, al usar varias clasificaciones se determina cuáles elementos (biomas mayores, tipos de vegetación, tipos de hábitats o especies) no están representados o están escasamente representados en el sistema actual de áreas de conservación. Una vez conocido esto con precisión razonable, se establecen prioridades para acciones de conservación. El proceso continúa indefinidamente, y el sistema de conservación se refina de acuerdo con los cambios de uso de suelo, de mejor información acerca de la distribución y estado de las especies, hábitats y biomas (Burley, 1990).

Faunística predictiva

Dentro de los estudios que realizan los taxónomos, los biogeógrafos y los conservacionistas resaltan varias preguntas importantes como ¿cuántas especies hay en un área dada?, ¿cuáles de esas especies son endémicas a la región?, y ¿qué esfuerzo de recolecta se requiere emplear para conocer la fauna en una región determinada? (Soberón & Llorente, 1993; Colwell & Coddington, 1994; Moreno & Halfter, 2000). Colwell (1997) instrumentó estadísticamente formas de estimación automatizada para riqueza de especies a través de muestreos de distintos tipos.

Los trabajos de exploración e inventario, desde hace más de dos siglos, han sido las técnicas principales que permiten conocer las biotas locales. La exploración se inicia cuando no se tiene conocimiento alguno de lo que existe en un área determinada, y el propósito de los inventarios ha sido generar una lista de las especies que habita en un lugar. Sin embargo, actualmente se piensa

que no son las únicas formas de poder contestar las preguntas que se plantearon en el párrafo anterior.

La faunística predictiva, es decir, la capacidad de hacer conjeturas respecto a listas de especies en áreas donde se desconoce absolutamente su fauna, es un campo de amplia investigación futura, ya que es posible iniciarla a partir del conocimiento previo de las faunas de localidades geográficas determinadas. En la actualidad se ha comenzado a instrumentar con el uso de bases de datos y con el apoyo de los SIG. Algunas metodologías son las siguientes:

a) Curvas empíricas

b) *Garp*

c) Estimación de la riqueza de especies en un área dada por modelamiento teórico de curvas de acumulación de especies a partir de datos de museos (para más detalles ver los artículos de Soberón & Llorente, 1993, y Colwell, 1997).

d) Monitoreo de fauna

La formalización del trabajo taxonómico, especialmente el que conduce a la creación de listas locales de especies, fácilmente puede proveer de una potente herramienta predictiva, cuyo uso sería invaluable en la conservación y el manejo de los recursos bióticos. Esta formalización requiere, por una parte, del acuerdo de ciertos estándares mínimos para recolectar y referir las listas (curvas de acumulación de especies por área y por tiempo; distribución del esfuerzo de recolecta) y, por otra, de la sistematización y análisis de los datos así obtenidos para la proposición de reglas empíricas con valor predictivo (Llorente *et al.*, 1994). Un ejemplo de cómo se han aplicado diferentes modelos de acumulación usados para extrapolar predicciones de tamaños de las faunas, con base en datos de campo y colecciones en museos, se encuentra en el artículo de León-Cortés *et al.* (1998).

Otra pregunta que resulta de gran interés es ¿existe el mismo número de especies en áreas del mismo tamaño que tienen las mismas características ecológicas pero en zonas diferentes?

La distribución actual de una especie dada cambia de un momento a otro en el tiempo (Vuilleumier, 1999). A pesar de ello, existe un área de distribución potencial durante un tiempo determinado (Llorente, 1984). Esta área es el conjunto de áreas donde una especie puede obtener o exhibir inmediatamente hábitat y nicho. Hay algunos factores que determinan tales distribuciones; por lo tan-

to, una aproximación para determinar una distribución de especies es analizar factores a gran escala como características de clima, altitud y vegetación, que pueden ser ligadas a las especies para inferir su distribución potencial. Este estudio es el inicio para entender la historia biogeográfica de un taxón, así, las alteraciones ambientales pasadas y futuras pueden estar correlacionadas a distribuciones previas y futuras hipotéticas de las especies, siendo de importancia en la biología de la conservación y biogeografía evolutiva (Cerqueira *et al.*, 1998).

El uso de las bases de datos de especímenes son una herramienta de gran valor, ya que para muchos grupos se tienen registradas sus variables ecológicas, y es posible construir listas hipotéticas de las especies que habitan en un área dada, puesto que existen factores ecológicos que determinan la riqueza, siempre considerando la ausencia de barreras. Pero con la aplicación de teorías de la biogeografía histórica se puede ir más allá en las predicciones, al incluir la vicarianza de biotas se puede complementar e indicar el grado de diferenciación de especies estenoecias. De esta manera, a partir de áreas ocupadas es posible dibujar áreas de distribución de las especies, con lo que se formulan hipótesis de extrapolación o generalización en áreas desconocidas con el uso de álgebra de mapas.

Por lo tanto, de la faunística de muchos grupos más o menos conocidos pueden derivar hipótesis de cuántas y cuáles especies habitan en un sitio de acuerdo con los factores ecológicos existentes, a partir de información vertida en sistemas de información, y, adicionalmente, pueden someterse a comprobación dichas hipótesis. A este proceso lo hemos llamado faunística experimental.

Agradecimientos

Juan José Morrone y Rodrigo Medellín revisaron de manera crítica el manuscrito preliminar. Tania Escalante agradece a la Dirección General de Estudios de Posgrado de la UNAM, ya que con el apoyo de la beca otorgada pudo realizarse este trabajo. Jorge Llorente tuvo el apoyo del proyecto IN 21 1397, una beca de DGAPA (1999-2000) y el apoyo de CONACyT (32002) y el Instituto de Ciencias Naturales (UNC). John D. nos ayudó con el Abstract.

Referencias

- Berendsohn, W. G. 1997. A taxonomic information model for botanical databases: The IOPI Model. *Taxon*, 46: 283-309.
- Berendsohn, W. G., A. Anagnostopoulus, G. Hagedorn, J. Jakupovic, P. L. Nimis, B. Valdés, A. Güntsch, R. J. Pankhurst & R. J. White. 1999. A comprehensive reference model for biological collections and surveys. *Taxon*, 48: 511-562.
- Burley, F. W., 1990. Monitoring biological diversity for setting priorities in conservation. En: Wilson, E. O. & F. M. Peter (eds.). *Biodiversity*. National Academy Press. pp: 227-230.
- Burrough, P. A., 1989. *Principles of geographical information systems for land resources assessment*. Clarendon Press, Oxford. 194 p.
- Cerqueira, R., G. Marroig & L. Pinder, 1998. Marmosets and Liontamarins distribution (Callitrichidae, Primates) in Rio de Janeiro State, South-eastern Brazil. *Mammalia*, 62(2): 213-226.
- Colwell, R. K. & J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London. Series B*, 345: 101-118.
- Colwell, R. K. 1997. *Estimates. Statistical Estimation of species richness and shared species from samples. Version 5 User's Guide and Application*. <http://viceroy.eeb.uconn.edu/estimates>. University of Connecticut, Storrs, CT.
- CONABIO, 1998. Manual de usuario de Biotica⁴ v. 3.0. CONABIO, México.
- Date, C. J., 1998. Introducción a los sistemas de bases de datos. 5a ed. Vol. 1. Addison Wesley Longman. México. 860 p.
- Davis, F. W., D. M. Stoms, J. E. Estes & J. Scepán, 1990. An information systems approach to the preservation of biological diversity. *Int. J. Geographical Information Systems*, 4(1): 55-78.
- Duckworth, W. D., H. H. Genoways & C. L. Rose. 1993. *Preserving natural sciences collections: chronicle of our environmental heritage*. Washington.
- Escalante, T. & G. Rodríguez, 1998. Diccionario computarizado de nombres de taxones de plantas vasculares y vertebrados para bases de datos de colecciones biológicas. Tesis. Facultad de Estudios Superiores Zaragoza. Universidad Nacional Autónoma de México, México.
- ESRI, 1996. *Introduction to Arc View GIS*. Environmental Systems Research Institute, Inc. USA.
- Green, D., 1994. Databasing diversity- a distributed, public domain approach. *Taxon*, 43: 51-62.
- Koleff, P. 1997. *Introducción a las bases de datos en la Biología Comparada Contemporánea*, Publicaciones Docentes del Museo de Zoología "Alfonso L. Herrera", Número 1, UNAM. 37 pp.
- León-Cortés, J. L., J. Soberón-Mainero & J. Llorente, 1998. Assessing completeness of mexican sphinx moth inventories through species accumulation functions. *Diversity and Distributions*, 4: 37-44.
- López, J., 1998. Sistemas de información geográfica (SIG): conceptos, definiciones y contexto metodológico que involucra su uso. *Quivera*, 1(0): 27-38.
- Llorente, J. 1984. Sinopsis sistemática y biogeográfica de los Dismorphiinae de México con especial referencia al género *Enantia* (Huebner) (Lepidoptera: Pieridae). *Folia Entomológica Mexicana*, 58: 1-207.
- Llorente, J., I. Luna, J. Soberón & L. Bojórquez, 1994. Biodiversidad, su inventario y conservación: teoría y práctica en la taxonomía alfa contemporánea. En: Llorente-Bousquets,

- J. e I. Luna (comps.). *Taxonomía Biológica*. Fondo de Cultura Económica. pp.: 507-522.
- Llorente, J. & P. Koleff, 1997. La actividad taxonómica en México. *Biodiversitas* 13: 11-14.
- Llorente, J. E., L. Oñate-Ocaña, A. Luis-Martínez e I. Vargas-Fernández, 1997. *Papilionidae y Pieridae de México: distribución geográfica e ilustración*. UNAM - CONABIO. 226 p.
- Maguire, D. J., M. F. Goodchild & D. W. Rhind (eds.), 1991. *Geographical information systems: principles and applications*. Longman, U.K.
- McFadden, E. R., & J. A. Hoffer, 1991. *Modern database management*. The Benjamin/Cummings Publishing Company, Inc. U.S.A. 597 p.
- Microsoft Corporation, 1994. *Microsoft Access, Relational Database Management System for Windows, Version 2.0, User's Guide*, U.S.A.
- Moreno, C. E. & G. Halffter. 2000. Assessing the completeness of bat biodiversity inventories using species accumulation curves. *Journal of Applied Ecology*, 37: 149-158.
- Papavero, N. & J. Llorente, 1999. Levantamiento de localidades. En: Papavero, N. & J. Llorente (comps.). *Herramientas prácticas para el ejercicio de la taxonomía zoológica*. UNAM-Fondo de Cultura Económica. pp.: 189-198.
- Papavero, N & J. Llorente, 1999. Mapas de distribución geográfica. En: Papavero, N. & J. Llorente (comps.). *Herramientas prácticas para el ejercicio de la taxonomía zoológica*. UNAM-Fondo de Cultura Económica. pp.: 199-202.
- Parker, H. D., 1988. The unique qualities of a geographic information system: a commentary. *Photogrammetric Engineering and Remote Sensing*, 54(11): 1547-1549.
- Peláez, G., 1994. Bases de datos en taxonomía y colecciones científicas. En: Llorente-Bousquets, J. e I. Luna (comps.). *Taxonomía Biológica*. Fondo de Cultura Económica. pp.:259-277.
- Peterson. A. T., A. G. Navarro-Sigüenza, & H. Benítez-Díaz, 1996. The need for continued scientific collecting; a geographic analysis of mexican birds specimens. *Ibis*, 140: 288-294.
- Scott, J. M. & B. Csuti, 1997. Gap analysis for Biodiversity survey and maintenance. En: Reaka-Kudla, M. L., D. E. Wilson y E. O. Wilson (eds.). *Biodiversity II. Understanding and protecting our biological resources*. Joseph Henry Press, Washington, D.C. p: 321-340.
- Soberón, J. & J. Llorente, 1993. The use of species accumulation functions for the prediction of species richness. *Conserv. Biol.* 7: 480-488.
- Soberón, J., J. Llorente & H. Benítez, 1996. An international view of National biological surveys. *Ann. Mo. Bot. Gard.*, 83: 562-573.
- Soberón, J., J. Llorente & L. Oñate, 2000. The use of specimens-labels databases for conservation purposes: an example using mexican Papilionid and Pierid butterflies. *Biodiversity and conservation* (en prensa).
- Soberón, J., J. Llorente-Bousquets & A. Luis-Martínez, 2000. Modeling richness of Papilionid and Pierid butterflies (Lepidoptera: Papilionoidea) in Mexico on the basis of museum data. (manuscrito).
- Soberón, J. & P. Koleff, 1998. The mexican experience in the collection, organization and management of biodiversity data and information. En: *Framework for National Environmental Information Systems*. International Federation for Information and Documentation, Netherlands.
- Soherón, J. & P. Koleff, 1998. The National Biodiversity Information System of Mexico. en Contribution to the Second National Forum on Biodiversity and Human Society: The Quest for a Sustainable Future. National Academic Press.
- Systematics Agenda 2000, 1994. *Charting the biosphere*. Technical report. Produced by Sycematics Agenda 2000: a consortium of the American Society of Plant Taxonomists, the Society of Systematic Biologists and the Willi Hening Society in cooperation with the Association of Systematic Collections. USA. 34 pp.
- Upper Midwest Gap Analysis Home Page, 1998. <http://www.emtc.nbs.gov/umgap/home.html>.
- Vuilleumier, F. 1999. Biogeography on the eve of the twenty-first century: towards an epistemology of biogeography. *Ostrich*, 70(1): 89-103.
- Walsh, S. J., 1985. Geographic information systems for natural resources management. *J. of Soil and Water Conservation*, 40(2): 202-205.
- WCMC, 1997. *Darwin Initiative Handbook No. 2: Information Needs Analysis*. World Conservation Monitoring Centre, Cambridge, UK.