

La información sobre biodiversidad y su uso en procesos concretos de toma de decisiones

Julia Benavides-Molineros

Instituto de Investigación de Recursos Biológicos Alexander von Humboldt,
Sistema de Información sobre Biodiversidad de Colombia,
Bogotá, Colombia
jebenavides@humboldt.org.co

y

Ximena Franco-Villegas

Instituto de Investigación de Recursos Biológicos Alexander von Humboldt,
Sistema de Información sobre Biodiversidad de Colombia,
Bogotá, Colombia
xfranco@humboldt.org.co

Resumen

El Sistema de Información sobre Biodiversidad de Colombia (SiB) se ha venido implementando durante los últimos cinco años como una “alianza nacional desarrollada para facilitar la gestión de datos e información que apoyen oportuna y eficientemente procesos de investigación, educación o toma de decisiones relacionadas con el conocimiento, la conservación y el uso sostenible de la diversidad biológica de Colombia”. En esta labor, el país ha estructurado y puesto en marcha una iniciativa interinstitucional que cuenta con el apoyo de múltiples entidades y que facilita el acceso a datos, información y herramientas para que científicos y gestores de la biodiversidad involucrados en la toma de decisiones, tengan a la mano insumos, recursos de información e instrumentos que apoyen sus labores.

El Equipo Coordinador del SiB Colombia, con asiento en el Instituto de Recursos Biológicos Alexander von Humboldt, ha generado productos y servicios de información y ha definido vacíos existentes en esta materia conjuntamente con sus socios en la implementación del Sistema. En este trabajo se describen tres de estos desarrollos: 1. Servicio de información en línea sobre especies colombianas con potencial comercial; 2. Producto de información para la toma de decisiones en la evaluación de riesgos de OGM en Colombia, y 3. Herramienta para la extracción, transporte y carga de datos desde fuentes heterogéneas y distribuidas haciendo uso de lenguajes controlados. Las herramientas que se presentan en estos ejemplos han buscado establecer un acercamiento entre la producción científica y procesos concretos de toma de decisiones.

Palabras clave: Productos de información para la toma de decisiones, Sistema de Información sobre Biodiversidad, SiB Colombia, biodiversidad de Colombia.

Abstract

Biodiversity Information and its Use in Specific Decision Making Processes

The Colombia Biodiversity Information System (BiS) has been implemented during the last five years as a “national alliance developed to facilitate data and information management to efficiently support research, education or decision making processes related to the knowledge, conservation and sustainable use of biological diversity in Colombia”. In accomplishing this task, the country has structured and started an initiative that has the support of many institutions, with the aim of facilitating access to data and information, and of supplying tools so that scientists and administrators involved in biodiversity decision making processes can have information resources at hand to assist them in their work.

The Coordinating Team of Colombia BiS, based in the Alexander von Humboldt Biological Resources Research Institute, in a joint effort with its partners in the System’s implementation, has produced information products and services and has defined existing gaps in biodiversity decision making processes. Some examples of these activities are described in this work: 1. On-line Information Service about Colombian Species with Commercial Potential; 2. Information Product for Decision Making in GMO Risk Assessment, and 3. Tool for Extraction, Transport and Loading of Data from Distributed Sources Using Controlled Vocabularies. With the tools presented in these examples some approach has been achieved between scientific production and specific decision making processes.

Key Words: Information products for decision making, Colombia Biodiversity Information System, Colombia BiS, biodiversity in Colombia.

1. El Sistema de Información sobre Biodiversidad de Colombia - SiB

El Sistema de Información sobre Biodiversidad de Colombia – SiB, una alianza nacional de múltiples socios institucionales y liderada por el Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, está encaminada a “facilitar la gestión de datos e información que apoyen eficientemente procesos de educación, investigación y toma de decisiones relacionados con el conocimiento, la conservación y el uso sostenible de la diversidad biológica de Colombia”. En el cumplimiento de este propósito, el SiB es entendido como un Sistema desde un sentido amplio y no se ve sólo como una herramienta informática, un conjunto de procesos de análisis para la obtención de un recurso de información en particular, ni como el procesamiento de datos para una única institución o proyecto (todos estos, enfoques de *sistemas de información* ampliamente difundidos, válidos y útiles en sus contextos, pero que no son los utilizados en la implementación del SiB Colombia).

El SiB se ha estructurado a partir de tres ejes de implementación: (i) fortalecimiento de capacidades nacionales para la gestión de información sobre biodiversidad, (ii) infraestructura informática apropiada y, (iii) gestión eficiente de contenidos. Con estos tres componentes en los que se fundamenta el esquema de gestión de información propuesto por el SiB se busca, entre otros objetivos, facilitar el acceso a la información, promover la interoperabilidad y el intercambio de los datos, su integración.

En la gestión eficiente de contenidos de información se abordan, por un lado, la estandarización y el manejo de los datos básicos bajo el concepto de registros biológicos (para datos que representen cualquier nivel de organización biológica), manejo que implica el uso de lenguajes controlados y de metadatos, de estándares de documentación y de intercambio y de otros elementos ampliamente documentados en publicaciones previas; por el otro, se encuentra el aspecto de las herramientas de procesamiento y análisis de los datos para la generación de información en forma de productos de información y los servicios de descubrimiento, visualización y búsqueda de datos sobre biodiversidad. Este aspecto, denominado “productos y servicios de información” está enfocado principalmente a suplir necesidades de información útil a los tomadores de decisiones que no necesariamente cuentan con una formación en las ciencias biológicas y que requieren datos agregados, información procesada, o herramientas que apoyen el procesamiento de tales datos e información.

En este contexto se presenta este resumen, con base en el cual se busca dar a conocer tres ejemplos de productos de información que apoyarán de manera efectiva decisiones o procedimientos que requieren ser llevados a cabo en sectores de la producción (el primer ejemplo que se expondrá), o entre autoridades ambientales del orden nacional o regional en Colombia. En el documento se describen los servicios o productos, se presenta el contexto en el cual fueron desarrollados y las tecnologías en las cuales se basan.

2. Servicio de información sobre especies colombianas con potencial comercial

2.1. Descripción

El objeto del Servicio es facilitar el acceso desde un mismo sitio *web* a información biológica y comercial sobre las especies con potencial comercial en Colombia, con base en la lista de especies previamente identificadas como promisorias por la línea de investigación en Biocomercio Sostenible del Instituto Humboldt y a partir de la información disponible en recursos de información existentes. El rango de posibles usuarios del servicio va desde personas interesadas en tener un primer acercamiento al biocomercio y los mercados verdes, hasta empresarios con experiencia en el tema que requieren información especializada o entidades gubernamentales que quieren explorar posibilidades de uso de la biodiversidad.

El servicio de información desarrollado hace una **búsqueda federada** sobre cuatro recursos de información, el *Observatorio nacional de biocomercio (Obio)*, *DSpace*, el *Catálogo electrónico de los organismos presentes en Colombia y Yoscua*.

El *Observatorio nacional de biocomercio* contiene información básica y de análisis sobre empresas, mercados y tecnologías de biocomercio. Con *DSpace* se manejan documentos sobre las especies con potencial comercial y sobre otros temas de biocomercio y mercados verdes, tales como legislación o documentos de planeación. Esta herramienta es un repositorio de documentos en diferentes formatos digitales (texto, vídeo o audio) que permite organizarlos en categorías definidas de acuerdo con las necesidades de los usuarios, indexarlos y hacer búsquedas en línea sobre sus

fichas bibliográficas descriptivas y sus contenidos. El *Catálogo electrónico de los organismos presentes en Colombia* contiene fichas con información sobre distribución, nombres comunes, historia natural, taxonomía, estado de conservación y otros aspectos de las especies que habitan en Colombia. *Yoscua - Usos y saberes locales sobre biodiversidad* contiene información sobre los usos que diferentes comunidades hacen del medio natural: qué usan (recursos usados), para qué, cómo y dónde (espacios de uso).

El servicio de información hace uso de dos lenguajes controlados, el **tesauro sobre biodiversidad** (<http://www.siac.net.co/sib/tesauros2/WebModuleTesauros/index.jsp>) y los **archivos de autoridad taxonómica (AAT)** (<http://www.siac.net.co/sib/aat/WebModuleAAT/index.jsp>) del SiB.

La búsqueda en los contenidos de los recursos de información mencionados se hace usando el indexador *web Nutch*. Cuando el usuario hace una búsqueda, el servicio de información presenta los resultados encontrados en cada recurso, en un formato similar al de *Google* (título del recurso con hipervínculo y un extracto del texto en el que se encuentran las palabras buscadas). Además de estos resultados, también se obtiene dinámicamente una **lista de términos relacionados con la búsqueda**, proveniente del tesauro sobre biodiversidad y de los AAT. Lo que se pretende con esta función es sugerir nuevos términos al usuario para (i) ampliar el espectro de las búsquedas y obtener resultados que no están explícitamente establecidos en ellas, y (ii) refinar los resultados obtenidos.

En esta lista se incluyen los nombres científicos provenientes de los AAT asociados a los nombres comunes de las especies y los términos del **microtesauro de biocomercio**, construido de modo que sus contenidos y relaciones responden a las necesidades de información identificadas y a la estructura de búsquedas del Servicio. Actualmente el microtesauro contiene 290 términos organizados en siete clases que comprenden categorías de sectores, actividades y productos, productos propiamente dichos, componentes de los productos, usos y técnicas de producción. El usuario puede usar los términos de la lista de términos relacionados para generar nuevas búsquedas en los recursos de información con tan solo un clic.

El servicio también presenta permanentemente una **lista de las especies identificadas por su potencial comercial** y de las cuales el Obio contiene información.

2.2. Herramientas tecnológicas utilizadas

Servicio de información sobre especies colombianas con potencial comercial: Aplicación *web* en J2EE, servidor JBoss 4.0.2. Consultor para el desarrollo informático: Ingenian Software Ltda.

Observatorio nacional de biocomercio (Obio): Aplicación *web* en Java 1.5 con arquitectura J2EE, base de datos PostgreSQL, servidor JBoss 4.0.2. Desarrollo de la línea de investigación en Biocomercio Sostenible del Instituto Humboldt, conjuntamente con Proexport Colombia y el Fondo Biocomercio Colombia (FBC). Consultores para el desarrollo informático: Andrés Torres y Carlos Parra. <http://www.humboldt.org.co/obio/>

DSpace: Aplicación *web* de código abierto en Java 1.5 con arquitectura J2EE, base de datos PostgreSQL, servidor JBoss 4.0.2. Desarrollo conjunto de las bibliotecas de MIT (Massachusetts Institute of Technology) y los laboratorios de Hewlett-Packard (<http://www.dspace.org>). Mejoras en el motor de búsqueda usando Lucene implementadas por Ingenian Software Ltda. para el SiB - Instituto Humboldt.

Catálogo electrónico de los organismos presentes en Colombia: Aplicación *web* en Java 1.5 con arquitectura J2EE, uso de Hibernate, Struts, base de datos PostgreSQL, servidor JBoss 4.0.2. Desarrollo del Sistema de Información sobre Biodiversidad de Colombia y el programa de Inventarios de la Biodiversidad del Instituto Humboldt. Consultor para el desarrollo informático: Ingenian Software Ltda. <http://www.siac.net.co/sib/catalogoespecies/welcome.do>

Yoscua - Usos y saberes locales sobre biodiversidad: Aplicación *web* de tipo wiki en Java 1.5 con arquitectura J2EE, base de datos PostgreSQL, servidor JBoss 4.0.2. Desarrollo del SiB y el programa de Uso y Valoración de la Biodiversidad del Instituto Humboldt. Consultor para el desarrollo informático: Ingenian Software Ltda. <http://www.siac.net.co/yoscua>

Componente de búsqueda *web* e indexación: Nutch versión 0.8. Componente de código abierto desarrollado dentro del proyecto Apache Lucene (<http://lucene.apache.org/nutch/>).

Conexión con las bases de datos de tesauro y AAT: Servicio *web* en Java 1.5 que hace uso de protocolo SOAP y Axis2 0.94, servidor JBoss 4.0.2. Desarrollo del SiB - Instituto Humboldt.

El proyecto completo es un desarrollo conjunto del SiB y la línea de investigación en Biocomercio Sostenible del Instituto Humboldt para el grupo de Mercados Verdes del MAVDT. El servicio está disponible en <http://www.humboldt.org.co/b3/>.

3. Producto de información para la evaluación de riesgos ambientales de organismos genéticamente modificados

3.1. Descripción

El MAVDT es en Colombia la autoridad nacional competente para autorizar o emitir pronunciamientos ambientales sobre las actividades de movimiento transfronterizo, tránsito, manipulación y utilización de los organismos genéticamente modificados (OGM) que puedan tener efectos adversos para el medio ambiente y la diversidad biológica, teniendo en cuenta los riesgos para la salud humana y la producción agropecuaria. En el proceso de toma de decisiones para dar curso a las solicitudes presentadas, el MAVDT debe realizar una evaluación de riesgos caso por caso, con el objeto de identificar y valorar los potenciales efectos directos e indirectos sobre la salud humana, el ambiente y la biodiversidad, la producción agropecuaria y, cuando se requiera, los potenciales efectos socioeconómicos.

El producto de información que aquí se presenta se desarrolló para manejar la información sobre biodiversidad requerida en los procesos de evaluación de riesgos, particularmente la información sobre parientes silvestres de los OGM.

En primer término, se desarrolló una herramienta informática dirigida a los evaluadores encargados de dar trámite a las solicitudes relacionadas con OGM, que posibilita el manejo de una estructura ordenada de criterios de evaluación de riesgos de las solicitudes de autorización. Esta herramienta permite construir y administrar **árboles de decisión**, y manejar las solicitudes de autorización haciendo uso de los árboles creados. Un árbol de decisión contiene preguntas (enunciados del árbol), opciones de respuesta a las preguntas/enunciados, las cuales determinan la ruta de navegación en el árbol, ayuda para responder a las preguntas, comentarios a las respuestas, alertas y ponderaciones de riesgo.

La evaluación de las solicitudes en trámite siguiendo de manera sistemática los criterios contenidos en los árboles de decisión creados, se hace a través de una aplicación asociada a una base de datos que almacena la información de cada solicitud evaluada. Como resultado final, después de haber recorrido todos los pasos del árbol, y con base en las respuestas y comentarios hechos y las ponderaciones asignadas, los evaluadores obtienen una orientación sobre el sentido de la decisión (negar o aprobar una solicitud, requerir al solicitante ampliación de la información o establecer las medidas de manejo correspondientes, de acuerdo con las circunstancias particulares del caso evaluado).

Como apoyo al proceso de desarrollo de la herramienta informática descrita, se compiló información biológica de base sobre tres especies de importancia económica en Colombia, **algodón, maíz y papa**. Se hizo una revisión de los parientes silvestres o especies relacionadas presentes en Colombia, y se recopiló información taxonómica de referencia, información sobre historia natural y datos de ocurrencia de las especies de interés, de sus parientes silvestres y de cultivos.

La información recopilada fue estructurada en las herramientas que hacen parte del esquema de gestión de información del SiB. Los registros biológicos¹ (especímenes y observaciones) de las especies y sus parientes se estructuraron de acuerdo con el *Estándar para intercambiar información al nivel de organismos* [2]. Las fichas nomenclaturales de las especies se hicieron de acuerdo con el *Estándar para la elaboración de archivos de autoridad taxonómica* [3]. Los conjuntos de datos se documentaron con metadatos de acuerdo con el *Estándar para la documentación de metadatos de conjuntos de datos relacionados con biodiversidad* [1] en el *Catálogo nacional de metadatos sobre biodiversidad*.

Con los registros biológicos recopilados en los herbarios del país se conformó un **nodo de parientes silvestres de OGM** en el **servicio de búsquedas distribuidas** del SiB. Este servicio permite realizar búsquedas distribuidas de registros biológicos en línea y en tiempo real, en los nodos de datos conectados al Sistema, y visualizar espacialmente estos datos, también en línea, usando la herramienta *BerkeleyMapper*. Los datos de este nodo están disponibles en http://www.siac.net.co/BancoConocimiento/S/sib_condiciones_acceso_distribuida/sib_condiciones_acceso_distribuida.php. Se espera que los registros biológicos que se sigan recopilando en las evaluaciones de riesgos relacionadas con OGM en el país se pongan a disposición del público a través de este nodo.

Con la información estructurada en el producto, los evaluadores y el público interesado podrán visualizar la localización de los cultivos no modificados y modificados, las áreas en las que se han autorizado OGM y la distribución de los parientes silvestres de las especies de interés.

3.2. Herramientas tecnológicas utilizadas

¹ Un registro biológico es la observación y captura de información sobre unidades biológicas referidas a un espacio (y usualmente a un momento) determinado. Los especímenes de un herbario, las accesiones de un banco de tejidos o las observaciones de organismos en campo, son ejemplos de registros biológicos.

Componente para creación de árboles de decisión: XForms 1.0 sobre una estructura básica estándar XSD. Consultor para el desarrollo informático: Ubiquando Ltda.

Componente para administración de los árboles de decisión y las solicitudes: Aplicación *web* cliente-servidor en Java 1.5 con arquitectura J2EE, base de datos Microsoft SQL Server 2000, capa de presentación JSF MyFaces, servidor *web* Apache-Tomcat 5.5. Consultor para el desarrollo informático: Ubiquando Ltda.

Servicio de búsquedas distribuidas de registros biológicos: Servicio *web* en Java 1.5 con arquitectura J2EE que hace uso de protocolo SOAP y Axis2 0.94. Opera con cualquier servidor de aplicaciones que soporte Java y base de datos relacional. Desarrollo del SiB.

http://www.siac.net.co/BancoConocimiento/S/sib_condiciones_acceso_distribuida/sib_condiciones_acceso_distribuida.php

BerkeleyMapper: Aplicación *web* de código abierto integrada al servicio de búsquedas distribuidas de registros biológicos (<http://berkeleymapper.berkeley.edu/>).

El proyecto completo es un desarrollo conjunto del SiB - Instituto Humboldt, y la Dirección de Ecosistemas del Ministerio de Ambiente, Vivienda y Desarrollo Territorial de Colombia, en el marco del proyecto *Desarrollo de capacidades institucionales para implementar en Colombia el Protocolo de Cartagena sobre seguridad en la biotecnología*, financiado por el Global Environment Facility (GEF) - Banco Mundial y el gobierno de Colombia.

4. Herramienta para la extracción, transporte y carga de datos desde fuentes heterogéneas y distribuidas haciendo uso de lenguajes controlados

4.1. Descripción

En Colombia existen 33 **autoridades ambientales regionales** (corporaciones autónomas regionales y de desarrollo sostenible, en adelante abreviadas como CAR). Éstas son entidades autónomas de carácter público, integradas por entes territoriales que conforman una unidad geopolítica, biogeográfica o hidrogeográfica, y encargadas de administrar los recursos naturales renovables y no renovables. En ejercicio de su función, manejan, generan y demandan información sobre biodiversidad.

Parte de la información se maneja de manera semiestructurada o estructurada (en hojas de cálculo o en bases de datos) y cada CAR usa diferentes herramientas para ello, por ejemplo, Excel, Access, FoxPro, Informix, PostgreSQL, Oracle, etc. Por otro lado, los términos usados para ciertos atributos o entidades no están estandarizados ni siquiera en las bases de datos de una misma CAR.

Una de las principales necesidades de gestión de información identificadas es que ciertos tipos de datos sean interoperables. Con base en ello, se determinó comenzar a desarrollar los mecanismos para estructurar y compartir información de dos tipos, **registros biológicos** y **trámites** (licencias, permisos y autorizaciones)²; ambos tipos pueden manejarse de forma estructurada. Esta prioridad se estableció porque esta información es fundamental, puesto que con base en ella se pueden estimar las presiones de uso de los recursos naturales y se puede hacer seguimiento y evaluación del estado de la biodiversidad.

En una primera fase se diseñó y desarrolló un sistema para extraer ciertos elementos mínimos de la información de registros biológicos y trámites de las CAR localizada en fuentes múltiples y heterogéneas, homologarla con base en unos vocabularios controlados y cargarla en un repositorio central, para que luego pueda estar disponible a través de Internet. Los componentes del sistema se describen a continuación (Figura 1).

Un componente para ejecutar las funciones de **mapeo** de los datos en las fuentes, **extracción** desde diferentes bases de datos, archivos de texto o de Excel, **transporte** de un repositorio de datos a otro, **transformación** de los datos y **carga** de la información en un repositorio o en una bodega de datos. Este componente es la herramienta de código abierto **Kettle**.

Los datos se transportan desde las fuentes originales a una **base de datos estándar** localizada en cada una de las CAR, que contiene los campos de los **elementos mínimos de registros biológicos y trámites** que se requieren compartir. En esta base de datos, algunos campos de la información extraída se **homologan** con dos conjuntos de referencia del SiB, los **archivos de autoridad taxonómica y el tesoro sobre biodiversidad**. La función de homologación se ejecuta con

² Los siguientes son algunos ejemplos de los principales tipos de trámites: licencias ambientales, permisos Cites, permisos de caza, registros de circos, zoológicos, viveros y jardines botánicos, aprovechamientos forestales, salvoconductos de movilización, permisos de investigación y sanciones (decomisos, sanciones económicas, etc.).

un componente que compara los datos de las CAR con los conjuntos de referencia, encuentra y señala los datos no coincidentes, sugiere posibilidades de reemplazamiento con base en una función de similitud y reemplaza los valores de la base de datos original por los datos de referencia cuando el usuario lo determina.

Desde esta base de datos estándar, los datos homologados se **cargan** en la estructura del **metamodelo de Ideam**. Esta base de datos es el repositorio en el que Ideam pretende integrar la información meteorológica e hidrológica del país, la información de uso de los recursos naturales de la industria y parte de la información de suelos y de los estudios de impacto ambiental de ciertos sectores de la economía. El servicio de búsquedas distribuidas del SiB tendrá acceso al metamodelo para hacer búsquedas de los registros biológicos.

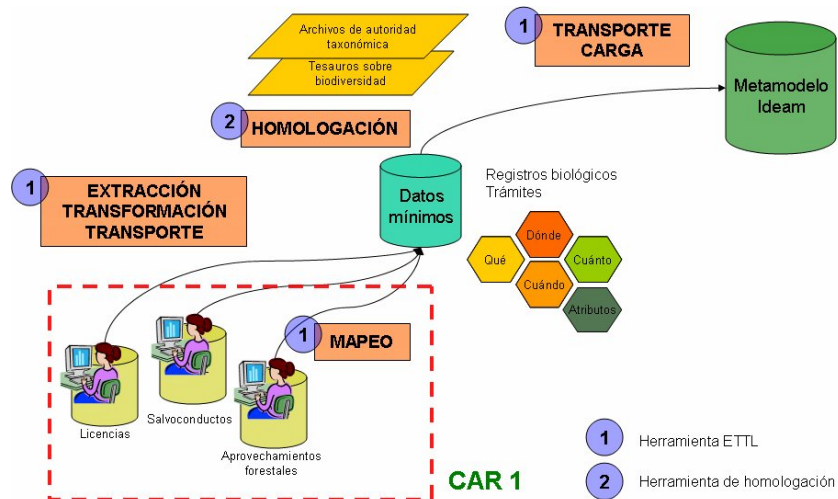


Figura 1. Esquema general del proyecto en la primera fase.

En el futuro se construirá un portal para la consulta de la información extraída de las CAR. En este portal será posible hacer consultas en línea de los datos básicos, así como análisis que requieren cierto grado de procesamiento de los datos.

4.2. Herramientas tecnológicas utilizadas

Herramienta ETL: Kettle - Pentaho Data Integration 2.3.1. Aplicación de código abierto de Pentaho Corporation, Open Source Business Intelligence (<http://kettle.pentaho.org/>).

Homologador: Aplicación *web* cliente-servidor en Java 1.5 con arquitectura J2EE, base de datos PostgreSQL 8.1, capa de presentación JSF MyFaces, servidor de aplicaciones JBoss 4.0.4. Consultor para el desarrollo informático: Ubiquando Ltda.

Cargador: Aplicación *web* cliente-servidor en Java 1.5 bajo Java SE, capa de presentación Swing, usa Kettle como herramienta ETL. Consultor para el desarrollo informático: Ubiquando Ltda.

Base de datos en CAR: PostgreSQL 8.1.

Metamodelo en Ideam: Oracle 9i.

El proyecto es un desarrollo conjunto del SiB - Instituto Humboldt, y el Ideam.

5. Referencias

- [1] Rivera-Gutiérrez H.F. y Suárez-Mayorga A.M. 2005. Estándar para la documentación de metadatos de conjuntos de datos relacionados con biodiversidad. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt. Bogotá D.C. 56 p.
- [2] Suárez-Mayorga A.M., Bernal R. y Cárdenas López D. 2005. Estándar para intercambiar información sobre biodiversidad al nivel de organismos. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt. Bogotá D.C. 26 p.
- [3] Trujillo-Motta D., Varón-Londoño A. y Bello-Silva J.C. 2005. Estándar para la elaboración de archivos de autoridad taxonómica. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt. Bogotá D.C. 51 p.