

Proyecto: Migración de base de datos a un sistema
web para publicar observaciones o reportes de
especies en el Perú

Informe de Avance

Octubre-Diciembre

Centro de Datos para la Conservación
Universidad Nacional Agraria La Molina



Diciembre, 2008

RESUMEN

En este primer trimestre de trabajo se ha logrado cumplir con los objetivos de avance del proyecto trazados en la propuesta enviada a Inter American Biodiversity Information Network (IABIN). Actualmente tenemos la base de datos diseñada e implementada en el servidor del CDC-UNALM. Así como también, los programas necesarios para permitir el intercambio de información siguiendo el estándar DarwinCore. Tapirlink, Apache, PHP ya se encuentran instalados, configurados y funcionando en línea.

Así mismo, se ha avanzado con la migración a la nueva base datos diseñada en PostgreSQL de 33 863 registros. Estos registros están distribuidos en las tablas de Taxones, Fuentes Bibliográficas, Temas, Subtemas, Editorial, Tipos de Fuentes, Colecciones y Usuarios.

ABSTRACT

In this first quarter of the project, we have succeeded in meeting the progress goals outlined in the proposal sent to the Inter American Biodiversity Information Network (IABIN). Currently, we have the database designed and implemented in the CDC-UNALM server. The programs needed to allow the exchange of information according to the Darwincore standard (Tapirlink, Apache, PHP), are already installed, configured and are working on line.

Similarly, some progress with the migration of data to the new PostgreSQL designed database have been made, with 33 863 registers transfered. These registers are distributed in tables Taxones, Fuentes Bibliográficas, Temas, Subtemas, Editorial, Tipos de Fuentes, Colecciones y Usuarios.

OBJETIVOS DEL PROYECTO

- Incorporar información almacenada en la base de datos de observaciones de especies y de colectas del CDC-UNALM en el territorio nacional a un sistema de intercambio de información que favorezca al estudio y análisis de la biodiversidad.
- Mejorar las condiciones para mantener actualizada la base de datos, garantizando un flujo de información nueva, provista tanto por la propia actividad del CDC-UNALM como por otros investigadores y centros de investigación.

PRODUCTOS Y RESULTADOS ESPERADOS

- Base de datos implementada que permita responder a los estándares ABCD Schema o DarwinCore, empleando PyWrapper o algún otro software que permita publicar bases de datos en PostgreSQL siguiendo el estándar TAPIR.

- Formularios para el mantenimiento de la información, así como para hacer algunos reportes y consultas para buscar información y para evaluar el desempeño de la misma.
- Información de observaciones de especies disponible por Internet siguiendo el estándar de IABIN, junto con información complementaria. La información de observaciones contiene datos de: descripción taxonómica (reino, clase, orden, familia, género, especie y, de ser preciso, subespecie o variedad), ubicación espacial (incluyendo precisión), fecha de registro, fuente bibliográfica, entre otros.

METODOLOGÍA EMPLEADA Y ACTIVIDADES REALIZADAS

En tanto el mantenimiento de una base de avistamientos de especies a nivel nacional es parte de los quehaceres propios del CDC-UNALM. Por ello, era importante realizar un trabajo interno de revisión de la propuesta de la base de datos al interior del equipo. Para tal fin, una vez concluida la propuesta de base de datos, se realizó un taller para presentar las reformas propuestas al equipo del CDC-UNALM, recoger sugerencias, y evaluar diferentes temas, entre ellas, los mecanismos para el control de la calidad del ingreso de datos ingresados. Las sugerencias también ayudaron a terminar de perfilar el diseño de la base de datos.

Para completar la actividad del primer trimestre fijada en la propuesta enviada, se ha constituido un equipo de trabajo que incluye dos practicantes de informática para garantizar la instalación y configuración del servidor donde se va implementar la base de datos y el sistema web con la finalidad de ponerla en línea. En este primer informe se explicarán los avances alcanzados hasta el momento, previos al lanzamiento de la base de datos institucional.

RESULTADOS

1. Indicador: Diseño de la estructura de la nueva base

La base de datos ya se encuentra desarrollada e implementada en el servidor del CDC-UNALM utilizando el manejador de base de datos PostgreSQL. La base de datos se podrá acceder a través de la siguiente URL <http://cdc.lamolina.edu.pe:8080/>.

El diseño de la base de datos se hizo teniendo en consideración las exigencias de DarwinCore de una manera tal que permita proveer información sobre observaciones de especies y especímenes en el Perú. También se espera que favorezca un nuevo flujo de información, provista tanto por la propia actividad del CDC-UNALM como por otros investigadores y centros de investigación.

Este modelo de la base de datos es un diseño totalmente escalable, es decir, que permite incorporar nuevos módulos de información en el futuro tales como

un módulo de áreas naturales protegidas, monitoreo de biodiversidad y otros tipos de datos relevantes para el manejo de la biodiversidad y conservación.

Por otro lado, todos los programas necesarios para el funcionamiento del sistema web que se va a desarrollar en los próximos meses, como el servidor web Apache, el PHP y el Tapirlink, también se encuentran instalados, configurados y en funcionamiento en el servidor del CDC-UNALM. Actualmente puede verse la conexión al TapirLink en modo prueba en la URL <http://cdc.lamolina.edu.pe/tapirlink/admin/configurator.php>

Se han realizado pruebas desde diferentes puntos de conexión a la base de datos para garantizar su funcionamiento, con resultados favorables.

2. Indicador: Migración de los datos de la base de datos antigua a la nueva base de datos.

La nueva base de datos del CDC-UNALM cuenta con un diseño de más de 70 tablas relacionadas, en las cuales se puede almacenar diferentes tipos de información como las de localizaciones de especies del Perú, taxonomía, fuentes bibliográficas, tipos de evidencia, nombres de sitios, especímenes, datos básicos de las áreas naturales protegidas, entre otros.

Siguiendo el cuadro de indicadores de cumplimiento enviado en nuestra propuesta, una de las actividades a cumplir en el primer trimestre es migrar más de 30 000 registros a la nueva base de datos. Para poder cumplir con esta actividad, ya se encuentra migrada la información de las tablas fuentes bibliográficas, taxones, temas, subtemas, editorial, tipos de fuentes, colecciones y usuarios: (ver cuadro Nro. 1).

Cuadro Nro. 1 Registros migrados por tablas

Tablas	Nro. de Registros
Temas	24
Subtemas	115
Editorial	701
Tipos de fuentes	27
Colecciones	27
Usuarios	44
Fuentes bibliográficas	5117
Taxones	27808
Total	33863

Temas: Almacena los diferentes temas de las fuentes bibliográficas. (Por ejemplo: Biodiversidad, Ecología, Conservación, Fauna, Flora, entre otros.)

Subtemas: Almacena los diferentes subtemas de las fuentes bibliográficas. (Por ejemplo: Agricultura, Ganadería, Monitoreo, Censos, entre otros)

Editorial: Contiene los nombres de la editorial de la fuente bibliográfica.

Tipo de Fuentes: Contiene los diferentes tipos de fuentes bibliográficas.

Colecciones: Almacena los datos de la colecciones bibliográficas.

Usuarios: Contiene la información de los usuarios del sistema.

Fuentes Bibliográficas: Almacena toda la información referente a las fuentes bibliográficas, como autor, título, número de páginas, año de publicación, entre otros.

Taxones: Contiene información de la toda la taxonomía desde reino hasta subespecie.

LECCIONES APRENDIDAS, PROBLEMAS Y POSIBLES SOLUCIONES

Hasta el momento, se ha podido avanzar en el proceso sin mayores contratiempos. El principal problema encontrado hasta ahora es la forma de publicar la información bibliográfica y la información taxonómica, puesto que el TapirLink no reconoce los esquemas DublinCore ni PlinianCore. Esperamos abordar este problema en los próximos meses.

CONTRIBUCIÓN EN ESPECIES

Fondos de contrapartida					
Información de la institución (FUENTE)					
Beneficiario de fondos IABIN <input checked="" type="checkbox"/>	Academia <input checked="" type="checkbox"/>	ONG	Agencias gubernamentales	Agencias multilaterales	Otros
Nombre de la institución	Fundación para el Desarrollo Agrario				
Departamento/ oficina/otros	Centro de Datos para la Conservación de la Universidad Nacional Agraria La Molina (CDC-UNALM)				
Nombre de la persona autorizada	Edith Melgar Romo				
País	Perú				
Período	22-09-2008 al 22-03-2009				
Fecha	22-12-2008				
Favor seleccionar el componente al que su institución contribuye a los objetivos de IABIN (favor no incluir fondos contribuidos por IABIN)					
02 – Creación del contenido de datos					
Unidades	Costo del semestre (US\$)	Descripción			
Personal técnico	1800	Coordinación			
Personal administrativo	1018	Apoyo de secretaría y oficina, y costos administrativos			
Equipo y suministros	560	Servidor en línea, cuatro computadoras empleadas para el desarrollo del sistema, suministros de oficinas (papel y tintas)			
Infraestructura física					
Documentos					
Asistencia de personal en talleres					
Otros					

REPORTE FINANCIERO

ORGANIZACIÓN DE LOS ESTADOS AMERICANOS REPORTE DE GASTOS

PO #: 193361
 Date: 22/12/2008
 Application #: 1
 Page number: 1

INSTITUCION: FUNDACION PARA EL DESARROLLO AGRARIO DICIEMBRE DEL 2008

Item N°	Nombre del proveedor/Descripción	Número de Contrato	Ubicación	Monto original del contrato (equivalente en dolares)	% Financiado por SG/OEA	Pagos acumulados ya reportados	Fecha de Pago	Factura	Monto
1.	Luis Enrique Gushiken Okuhama	PO#193361	Lima	1 150.50			Noviembre	RH:001-0233	1 150.50
2.	Luis Enrique Gushiken Okuhama	PO#193361	Lima	2 400.00			Diciembre	RH:001-0237	2 400.00
3.	ITF	PO#193361	Lima	2.48			Diciembre	BCP	2.48
4.	Gastos Bancarios	PO#193361	Lima	53.00			Diciembre	BCP	53.00
5.									

TOTAL \$ 3 605.98