

Coordinación de Colecciones Universitarias Digitales (CCUD)

Dra. Tila María Pérez Ortiz

Coordinadora

Febrero de 2013

Con el propósito de contribuir al fortalecimiento y desarrollo de la investigación que se realiza en la Universidad y en cumplimiento de las acciones del Plan de Desarrollo 2011-2015, el doctor José Narro Robles emitió, el 14 de enero de 2013, el acuerdo por el que se crea la Coordinación de Colecciones Universitarias Digitales, adscrita a la Secretaría de Desarrollo Institucional.

Las funciones de la CCUD están orientadas a desarrollar y administrar una plataforma tecnológica transdisciplinaria que dé visibilidad y contexto y, sobre todo, que integre de manera interoperativa los datos existentes de las distintas colecciones universitarias, con el fin de difundir entre los académicos y alumnos, así como instituciones de gobierno y la sociedad en general, los acervos universitarios y la información primaria que resulta de la investigación científica, humanística y artística.

Se trata de un programa universitario clave para la vinculación de la investigación con los diferentes sectores de la sociedad mexicana, beneficiándolos de manera directa con la capacidad que la Universidad cuenta en materia de investigación.

En ese sentido, la CCUD contribuye a que la investigación que se realiza en la UNAM tenga un mayor impacto en el análisis y la solución de los principales problemas de nuestro país.

ORGANIZACIÓN DE LA CCUD

La Coordinación de Colecciones Universitarias Digitales desarrolla sus funciones y organiza sus actividades a través de tres áreas de trabajo:

I. Área de Desarrollo Tecnológico

El objetivo de esta área es generar y configurar la ciberinfraestructura en su modalidad de hardware y software para proveer medios de publicación de datos, imágenes y documentos, optimizar

las operaciones, así como generar documentación técnica orientada al uso de la plataforma y a la transferencia tecnológica.

II. Área de Estadística e Informática

Esta área tiene la función de evaluar la información digital de las colecciones científicas, humanísticas y artísticas, así como de generar criterios de integración que permitan la interoperabilidad de los diversos datos; mediante la creación, evaluación y optimización de metodologías de control de calidad de la información, para permitir el registro de las fuentes de información, modificaciones y análisis del progreso de integración y depuración.

Por otra parte, da seguimiento y mantiene una constante retroalimentación de las observaciones realizadas por especialistas o usuarios de los servicios de la CCUD.

Con el fin de garantizar la coherencia y calidad de los datos de la información publicada se definen cinco líneas de trabajo: a) Bases de datos, que integra a nivel informático los datos de las colecciones; b) Análisis, la cual integra, revisa y depura la información, tomando criterios académicos; c) Estadística, encargada de analizar de forma continua la información; d) Control de calidad, la cual revisa y califica el contenido de la información, y e) Documentación, responsable de generar manuales, informes y documentar metodologías.

III. Área de Vinculación y Comunicación

El área de Vinculación y Comunicación es responsable de coordinar el enlace institucional, en lo referente a la continuidad de los procesos de trabajo y la relación entre la coordinación y las unidades de informática, así como de la comunicación y difusión de la dependencia, que incluye las publicaciones y el desarrollo de la página electrónica de la CCUD.

PRINCIPALES LOGROS

Para el cumplimiento de los objetivos de sistematizar y dar acceso en línea de forma interoperativa a las colecciones digitales universitarias, la CCUD se ha enfocado en cinco rubros principales que integran y reflejan el trabajo cotidiano de los académicos asociados a los acervos de las diversas entidades universitarias: 1) Vinculación de colecciones universitarias; 2) Integración y publicación de datos; 3) Desarrollo de software; 4) Desarrollo de infraestructura tecnológica (centro de datos); y 5) Administración y gestión de proyectos.

A continuación se mencionan los logros obtenidos en 2013, en cada uno de estos ámbitos de trabajo.

Vinculación de colecciones universitarias

Como una forma de poner en práctica nuevas formas de organización de la investigación, para transitar de un sistema basado en los esfuerzos individuales a uno de investigación colectiva, interdisciplinaria e interinstitucional, se inició el desarrollo de una red que integra tanto a las colecciones de las entidades universitarias como al personal académico vinculado a ellas.

Actualmente 10 entidades universitarias integran esta red, a través de 19 solicitudes de ingreso:

1. Instituto de Biología (UNIBIO) - 11 colecciones nacionales
 - » Herbario Nacional (MEXU)
 - » Colección Nacional de Helminetos (CNHE)
 - » Colección Nacional de Moluscos (CNMO)
 - » Colección Nacional de Ácaros (CNAC)
 - » Colección Nacional de Arácnidos (CNAR)
 - » Colección Nacional de Crustáceos (CNCR)
 - » Colección Nacional de Insectos (CNIN)
 - » Colección Nacional de Peces (CNPE)
 - » Colección Nacional de Anfibios y Reptiles (CNAR)
 - » Colección Nacional de Aves (CNAV)
 - » Colección Nacional de Mamíferos (CNMA)
2. Facultad de Ciencias (UNICIENCIAS) - 2 colecciones
 - » Colección Ornitológica del Museo de Zoología de la Facultad de Ciencias "Alfonso L. Herrera"
 - » Colección de Mamíferos del Museo de Zoología de la Facultad de Ciencias "Alfonso L. Herrera"
3. Instituto de Ciencias del Mar y Limnología (UNINMAR) - 4 colecciones
 - » Colección Nacional de Equinodermos
 - » Colección Ictiológica
 - » Colección Nacional del Phylum Porifera
 - » Colección Malacológica
4. Instituto de Investigaciones Estéticas (UNIARTE), Archivo Fotográfico Manuel Toussaint - 29 colecciones
 - » Colección Abelardo C y G
 - » Colección Artistas de teatro
 - » Colección Beatriz de la Fuente
 - » Colección Blanco y Negro
 - » Colección Construcción de CU
 - » Colección Diapositivas sobre vidrio
 - » Colección Elisa Vargaslugo
 - » Colección Enrique Bordes Mangel
 - » Colección Fotografía antigua AF
 - » Colección Francisco de la Maza
 - » Colección Grabado Europeo. EB
 - » Colección Grabados de Manuel T
 - » Colección Guillermo Kahlo
 - » Colección Helen Escobedo, PG
 - » Colección José María Lupercio
 - » Colección José Verde Orive
 - » Colección Juan Guzmán
 - » Colección Julio Michaud
 - » Colección Justino Fernández

- » Colección Luis Márquez Romay
 - » Colección Manuel González Galván
 - » Colección Manuel Toussaint
 - » Colección Michael Zabé
 - » Colección Paul Gendrop
 - » Colección Proyecto de Pintura Mural Prehispánica
 - » Colección Raúl Flores Guerrero
 - » Colección Salvador Toscano
 - » Colección Tina Modotti
 - » Colección Vicente Cortés Sotelo
5. Instituto de Ecología - 2 colecciones
- » Especies invasoras
 - » Datos geospaciales del Laboratorio Nacional de Ciencias de la Sostenibilidad
6. Facultad de Ingeniería - 1 colección
- » Colección Paleontológica de la Facultad de Ingeniería
7. Instituto de Química (UNIIQUIM) - 1 repositorio de artículos
- » Información sobre la biodiversidad química de los organismos del país
8. Centro de Ciencias de la Atmósfera (UNIATMOS) - 1 repositorio de capas geospaciales
- » Atlas Climático Digital de México
9. Dirección General de Asuntos del Personal Académico (DGAPA) - 1 repositorio de proyectos
- » Proyectos apoyados por el PAPIIT durante el 2012
10. Red de Acervos Digitales, RADUNAM - 7 repositorios de objetos digitales
- » Repositorio CELE
 - » Repositorio de la Facultad de Filosofía y Letras
 - » Repositorio del Instituto de Investigaciones Sociales
 - » Repositorio del Instituto de Biología (Irekani)
 - » Repositorio de la Facultad de Economía
 - » Repositorio del Instituto de Ingeniería
 - » Instituto de Investigaciones Económicas

Con el fin de consolidar esta red se han impartido 12 asesorías para el desarrollo de sistemas y transferencia de tecnología. Sobre la plataforma en desarrollo de la CCUD se impartieron asesorías a la Facultad de Ciencias, al Instituto de Biología y al Instituto de Ciencias del Mar y Limnología. Acerca de la indexación con SOLR, se asesoró al Instituto Nacional de Antropología e Historia. Para el sistema de computarización de colecciones biológicas fueron asesoradas la Facultad de Ciencias y el Instituto de Biología.

Por otra parte, se realizaron 13 talleres orientados a los responsables de las unidades de informática de las dependencias sobre la transferencia de tecnología y el uso de los sistemas; talleres para los capturistas,

usuarios de los sistemas y analistas sobre la unificación de criterios de captura de datos y directrices para el manejo de información relativa a colectores, taxónomos y datos generales de las colectas.

Asimismo, se impartieron dos talleres externos sobre la estructuración de las bases de datos, redes de bases de datos, interconexiones de éstas y usos prácticos, para alumnos de posgrado y personal académico en general.

Integración y publicación de datos

Con el fin de apoyar la divulgación de la ciencia, tanto la dirigida a universitarios como a la población general, se ha trabajado en la integración y publicación de datos resultantes de la investigación universitaria. Para ello se han seguido las siguientes etapas:

1) Detección y descripción de colecciones

Se elaboró un cuestionario con más de 60 preguntas basado en el estándar internacional Natural Collections Descriptions, a través del cual se realizó el censo descriptivo de 49 colecciones de la UNAM, detectando un aproximado de 8 497 606 de registros digitales distribuidos en 12 entidades.

Este censo ofrece un panorama general inicial de las colecciones a cargo de cada dependencia y el estado en que se encuentran. En el cuestionario se consideran los siguientes puntos:

- Las colecciones digitales disponibles
- El estado del arte
- El personal asociado
- Una descripción de las colecciones, así como sus derechos de publicación y reproducción.

El inventario es una herramienta fundamental para desarrollar el plan de trabajo con cada colección y es realizado tanto por personal de la CCUD como por los responsables de cada colección.

2) Definición de estándares

La integración de los datos de diversas fuentes requiere que se sigan estándares para lograr un acceso unificado a la información. Éstos se definen principalmente por los responsables de las entidades, con el apoyo del personal de la CCUD.

Con el fin de poder integrar los datos de las colecciones de distintas entidades, se definieron y documentaron tres estándares para el intercambio de tres tipos de datos:

- **Colecciones biológicas.** En colaboración con el Instituto de Biología, se definió el estándar internacional Darwin Core para el manejo e integración de datos de las colecciones biológicas de las distintas entidades de la UNAM.
- **Obra artística y cultural.** En colaboración con el Instituto de Investigaciones Estéticas se diseñó un estándar basado en el estándar internacional LIDO, para el manejo e integración de los datos de las colecciones artísticas y culturales de las distintas entidades de la UNAM.

- **Datos fisicoquímicos.** Con la participación del Instituto de Ciencias del Mar y Limnología se diseñó y estandarizó una base de datos para que sea utilizada como estándar de manejo e integración de datos fisicoquímicos.

3) Control de calidad

En este primer año se integraron a la preplataforma 1.3 millones de registros detectados a través del censo, mismos que fueron depurados por medio de un proceso de control de calidad y homogenización. Este proceso se divide en:

- **Lógico.** Los datos se califican de manera automática por medio de la comparación con catálogos de administración, de clasificación, geográficos y curatoriales.
Los errores de asignación, omisión, ortográficos y sistemáticos se corrigen y documentan. Los registros con errores que no pueden ser solucionados de manera lógica se etiquetan para el siguiente proceso.
- **Por especialistas.** Se revisan los errores del punto anterior con un especialista de la colección.

4) Publicación

Alrededor de un millón de registros provenientes de la Facultad de Ciencias, el Instituto de Biología, el Instituto de Ciencias del Mar y Limnología y el Instituto de Investigaciones Estéticas, aprobaron el control de calidad y están listos para su publicación una vez que la plataforma de datos de la CCUD se libere.

Por otro lado se han publicado dos folletos informativos.

5) Computarización de colecciones

En conjunto con el Instituto de Biología se diseñó y administró un sistema de digitalización para las colecciones biológicas, a través del cual se han logrado digitalizar los datos de 576 mil ejemplares biológicos pertenecientes a cinco colecciones; el Herbario Nacional (MEXU), la Colección Nacional de Crustáceos (CNCR), la Colección Nacional de Insectos (CNIN), la Colección Nacional de Moluscos (CNMO) y la Colección Nacional de Arácnidos (CNAN). Estas cinco colecciones se dividen, a su vez, en 18 subcolecciones.

Como apoyo para este proyecto se abrió un programa de servicio social para alumnos de la Facultad de Ciencias.

Un total de 335 mil ejemplares digitalizados han pasado por los procesos de control de calidad y están listos para su publicación en la plataforma de datos de la CCUD.

Este mismo sistema de computarización de colecciones se instaló en la Facultad de Ciencias, para computarizar de manera masiva, a partir del 2014, 17 colecciones biológicas custodiadas por esa facultad.

Desarrollo de software

Con el fin de poder capturar, administrar, consultar y analizar los datos de las colecciones de la UNAM, se han desarrollado los siguientes sistemas en línea:

- **Portal de especies invasoras del Instituto de Ecología.** Se desarrolló un portal (<http://www.unibio.unam.mx/invasoras/>) donde se pueden consultar, de manera individual e integral, 2 286 registros de opiniones de expertos del grado de invasividad de especies de plantas invasoras, recopilados por académicos del Instituto de Ecología.
- **Sistema de datos geospaciales para el Laboratorio Nacional de Ciencias de la Sostenibilidad con datos generados por ellos mismos.** Se generaron e integraron 160 capas geográficas, junto con 11 documentos PDF y 104 imágenes, en un sistema de información geográfica en línea que será utilizado por los distintos proyectos del Laboratorio Nacional de Ciencias de la Sostenibilidad del Instituto de Ecología de la UNAM.
- **Portal del índice de géneros del Herbario Nacional.** Se desarrolló también un portal (<http://unibio.unam.mx/generosmexu/>) que enlista los 6 942 géneros de plantas vasculares del acervo del Herbario Nacional de México (MEXU), reunidos por el personal académico asociado a MEXU. Con este índice se pueden ubicar los ejemplares botánicos en las distintas salas del Herbario Nacional.
- **Sistema de reporte de errores en ejemplares digitalizados del Instituto de Biología.** Con el fin de optimizar el sistema de computarización de las colecciones nacionales del Instituto de Biología y de la Facultad de Ciencias, se desarrolló un sistema que almacena y prioriza los errores o problemas detectados por los capturistas durante el proceso de digitalización. Este sistema permite que diversos expertos revisen los ejemplares para solucionar los problemas y corregir los errores.

A su vez, se encuentran en etapa de pruebas el portal de la CCUD y el portal de Consulta de Datos de la UNAM, mismos que serán puestos en línea durante el primer semestre del 2014 y desde donde se podrán consultar, analizar y descargar los datos publicados.

Este portal permitirá la publicación de fichas en línea de cada registro de las distintas colecciones universitarias, aumentando enormemente la presencia de contenidos académicos de la UNAM en Internet, lo que se reflejará en un aumento en los ranqueos de universidades, a partir de sus contenidos en línea y su presencia en los motores de búsqueda.

También está en proceso el desarrollo un sistema de gestión de obra artística de la Dirección General de Promoción Cultural y Acervo Patrimonial de la SHCP, así como un Catálogo de autoridades de obra artística y cultural, que se ha desarrollado junto con el Instituto de Investigaciones Estéticas de la UNAM, el Instituto Nacional de Bellas Artes (INBA) y el Instituto Nacional de Antropología e Historia (INAH).

Desarrollo de infraestructura tecnológica (centro de datos)

Con el fin asegurar el almacenamiento y publicación de la plataforma de datos, se construyó un centro de datos con un área total de 10 metros cuadrados, ubicado en el módulo C del edificio principal del Instituto de Biología.

El espacio físico donde se ubica el centro de datos cuenta con piso elevado a 30 centímetros del suelo, plafón y perímetro de cristal templado.

El control térmico del área se lleva a cabo mediante un equipo de aire acondicionado de precisión con capacidad de 60 mil BTU/hr. La configuración actual del área corresponde a 20°C y 40 por ciento de humedad relativa.

La capacidad de almacenamiento total es de 40 terabytes con una configuración de redundancia en discos como primer nivel de respaldo.

La capacidad de procesamiento de datos está dividida en:

- 10 microprocesadores de alto rendimiento (CPU)
- 397 GB de memoria de acceso aleatorio (RAM)

Estos equipos soportan la plataforma de virtualización de la CCUD, cuyo objetivo es proveer un ambiente de redundancia de recursos y balanceo de carga de trabajo, con lo que se busca maximizar el tiempo de servicio en línea y minimizar el impacto de fallas en algún eslabón de la infraestructura.

La capacidad de transmisión de datos está consolidada por: enlace de fibra óptica con capacidad para 100 MB de ancho de banda provisto por la Dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC); tres conmutadores ethernet de 48 puertos para usuarios; un conmutador ethernet de 48 puertos de alto rendimiento para servidores.

Administración y gestión de proyectos

Dada la magnitud de las interacciones de trabajo de la CCUD con las diversas entidades universitarias y extrauniversitarias, se ha establecido una dinámica de trabajo a través de convenios o bases de colaboración.

Durante el año 2013 se firmó un convenio de colaboración para el desarrollo de un sistema de gestión de obra artística con el Instituto de Investigaciones Estéticas y la Secretaría de Hacienda y Crédito Público; así como unas bases de colaboración para la digitalización de colecciones de la Facultad de Ciencias.

También se han generado y documentado los protocolos y procedimientos de las cinco áreas que conforman la estructura de la CCUD (Coordinación General, Coordinación Técnica, área de Desarrollo Tecnológico, área de Informática y Estadística, y la área de Vinculación y Comunicación) y de su interacción con las Unidades de Informática de las distintas entidades.

