

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 realiza la Universidad y en cumplimiento de las acciones del Plan de Desarrollo de la Universidad 2011-2015, el 14 de enero de 2013 se publicó en **Gaceta UNAM** el Acuerdo del Rector por el que se crea la Coordinación de Colecciones Universitarias Digitales (CCUD), adscrita a la Secretaría de Desarrollo Institucional de la UNAM.

De esta manera se consolidó el Sistema de Informática de la Biodiversidad y el Ambiente (SIBA), ampliando su ámbito de acción a los tres subsistemas de la UNAM, así como a las facultades y escuelas.

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

La CCUD desarrolla un programa clave para la vinculación de la investigación con los diferentes sectores de la sociedad mexicana, beneficiándolos de manera directa con la capacidad con la que cuenta la Universidad 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.

Los objetivos de la CCUD son los siguientes:

- Implementar la Plataforma de Datos de la CCUD para facilitar la integración, interacción y aprovechamiento de las colecciones existentes en la UNAM, que se encuentran bajo resguardo de las entidades que forman los subsistemas de la Investigación Científica, de Humanidades, de Difusión Cultural, así como las facultades y escuelas.

- Recibir y sistematizar la información sobre las nuevas colecciones científicas, humanísticas y artísticas que de manera cotidiana genera el trabajo académico de la UNAM.
- Definir estándares, protocolos y lineamientos para el manejo digital de las colecciones de la UNAM y garantizar así su interoperabilidad.
- Proveer herramientas para la investigación transdisciplinaria mediante el acceso a las colecciones y su información asociada, a través de un portal de datos abiertos y otras herramientas de difusión de contenidos digitales.
- Difundir, a través de los medios adecuados a cada caso, las colecciones existentes en la UNAM.

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 área de Desarrollo Tecnológico es responsable del diseño, planeación, coordinación y administración del Centro de Datos de la CCUD, compuesto por equipos de alto rendimiento, equipos de telecomunicaciones y equipos de almacenamiento, con la finalidad de garantizar la integridad, persistencia, confidencialidad y respaldo de la información de acuerdo con un plan de administración de los activos informáticos de la CCUD. Asimismo, se encarga de diseñar, planear y administrar la infraestructura de tecnologías de información y comunicación de la CCUD.

Por otro lado, se encarga de las especificaciones, diseño, documentación y planeación del desarrollo técnico de la Plataforma de Datos de la CCUD, compuesta por manejadores de bases de datos, servidores web, sistemas de digitalización de colecciones, fichas digitales de registros en colecciones, repositorio de objetos digitales, sistemas de información geográfica en línea, portales de consulta y descarga de datos, el Portal de Datos Abiertos UNAM y la página institucional de la CCUD.

II. ÁREA DE ESTADÍSTICA E INFORMÁTICA

El área de Estadística e Informática tiene la función de organizar, coordinar y documentar la integración de nuevos datos, acervos y objetos digitales a la Plataforma de Datos CCUD, cuantificando y calificando la cantidad y calidad de los datos provenientes de las distintas colecciones, reportando las estadísticas de los datos integrados en la CCUD y dando seguimiento a las observaciones realizadas por especialistas o usuarios de la Plataforma de Datos CCUD.

Por otra parte, es responsable de la generación de los criterios que permitan la interoperabilidad de los datos, a través de ubicar e integrar en la plataforma los distintos catálogos para validar los datos según las áreas correspondientes a las colecciones.

También se encarga de la creación, evaluación y optimización de metodologías de control de calidad de la información y limpieza de datos.

III. ÁREA DE VINCULACIÓN Y COMUNICACIÓN

El área de Vinculación y Comunicación tiene la responsabilidad de mantener una relación constante entre las colecciones universitarias y sus versiones digitales en la CCUD, así como dar seguimiento a todas las tareas necesarias para que las labores de la Coordinación sean relevantes dentro del contexto universitario. Es a través de las bases de colaboración como se asegura la continuidad de los procesos de trabajo y la relación entre la CCUD y las entidades colaboradoras.

Otra actividad de vinculación, para la integración de colecciones a la Plataforma de Datos de la CCUD, es el levantamiento de un censo, con el objetivo principal de contextualizar la publicación de los datos; derivado de esta actividad se genera un inventario de colecciones universitarias.

En cuanto a las actividades de comunicación y difusión de la Coordinación, principalmente consiste en lo relacionado con publicaciones impresas y digitales, incluyendo la elaboración de folletos y otros materiales informativos de la CCUD.

ADMINISTRACIÓN DE PROYECTOS

El área de Administración de Proyectos se encarga de dar seguimiento a los procesos administrativos, así como de la gestión de la información, expedientes y bases de datos de los proyectos de la CCUD.

En el año 2014 es creado el Sistema Documental y de Calidad de la CCUD (SDC-CCUD), conformado por documentación técnica, operativa y estratégica. En este sistema se elaboran, clasifican y actualizan expedientes de los procesos administrativos de presupuesto y personal, instrumentos jurídicos, seguimiento de reuniones de trabajo, entre otros, basado en un procedimiento de codificación y trazabilidad de documentos y procesos de la CCUD.

El SDC-CCUD contiene más de 20 procedimientos de operación de las áreas operativas del CCUD, cuatro protocolos específicos de trabajo, así como Lineamientos de Publicación de la CCUD y los Lineamientos de Condiciones de Uso de los Datos y Objetos Digitales del Portal de Datos Abiertos UNAM, entre otros.

Con este sistema se busca facilitar la comunicación y colaboración dentro de la CCUD, preservar información, optimizar el manejo de la documentación y mantener medidas de seguridad para el acceso y almacenamiento de los datos.

La conformación del SDC-CCUD, que contiene el soporte documental robusto de las actividades de la CCUD, deriva en la emisión del Reglamento Interno y la Normatividad Operativa de la CCUD.

PRINCIPALES LOGROS

INFRAESTRUCTURA

CENTRO DE DATOS

El Centro de Datos de la CCU alberga una red de servidores, *switches* y unidades de almacenamiento que almacenan y publican los contenidos de las colecciones universitarias y hospedan los diferentes servicios en línea que provee la CCUD.

Para asegurar que estos contenidos y servicios se encuentren constantemente disponibles, se instaló una acometida eléctrica que garantiza autonomía eléctrica al equipo de cómputo que conforma esta red.

Con este mismo fin, en 2014 se instaló un nuevo UPS (capacidad de 15 kVa) que protege estos equipos de cortes o irregularidades en el suministro eléctrico.

Se adquirió y configuró una nueva unidad de almacenamiento Dell MD3200i y un Switch Dell Force 10 S55, para robustecer la capacidad de almacenamiento y operación del Centro de Datos.

Como parte de la adecuación de las áreas operativas de la CCUD se pusieron en marcha 13 líneas telefónicas, de las cuales tres son digitales y 10 análogas.

Este Centro de Datos es estratégico para el buen funcionamiento de la CCUD, ya que en él se ubica la Plataforma Digital de la CCUD.

PLATAFORMA DIGITAL DE LA CCUD (PD-CCUD)

Se trata de un sistema de bases de datos estandarizadas que integra datos tabulares, geoespaciales y objetos digitales. Éste permite generar, entre otros productos, fichas digitales que despliegan, a través de Internet, toda la información concerniente a cada registro de las colecciones que se suman a esta plataforma. Cada una de estas fichas mantiene una URL estática, lo que les permite ser indexadas por motores de búsqueda de Internet como Google, aumentando así la presencia de las colecciones universitarias en línea y la visibilidad digital de la UNAM.

Esta plataforma cuenta con dos tipos de manejadores de bases de datos, PostgreSQL y Apache CouchDB, así como una plataforma Lucene-Solr para realizar la indexación de datos y texto, servidores de mapas GeoServer, servidores de *linked data* Jena, servidores de imágenes de alta resolución IIImage y servidores de aplicaciones Glassfish.

Estos sistemas se comunican tanto entre sí como con sistemas externos e interfaces de usuario (Portal de Datos Abiertos UNAM), a través de una red de servicios web RESTful.

La integración de bases de datos, objetos digitales y capas geoespaciales es una tarea compleja que debe ser abordada con sistemas especializados, por lo que la PD-CCUD ha derivado en la concepción de otras dos nuevas plataformas: la Plataforma de Objetos Digitales de la CCUD (POD-CCUD) y la Plataforma de Capas Geoespaciales de la CCUD (PCG-CCUD).

PLATAFORMA DE OBJETOS DIGITALES DE LA CCUD (POD-CCUD)

Es común encontrar los registros de las colecciones universitarias acompañados de objetos digitales (como imágenes, artículos, videos, etc.), o bien como descripciones de éstos (i.e. descripción de acervos fotográficos). Los objetos digitales son un conjunto de datos que conforman una representación, misma que se pierde cuando el conjunto es dividido o modificado en su estructura. Por este motivo es fundamental contar con las herramientas correctas para su almacenamiento, gestión y visualización.

En esta línea se ha desarrollado un repositorio de objetos digitales y sus metadatos, equipado con las herramientas necesarias para realizar una vinculación eficiente, tanto entre las colecciones y sus objetos como con otros repositorios universitarios, a través de estándares de metadatos como el Dublin Core.

PLATAFORMA DE CAPAS GEOESPACIALES DE LA CCUD (PCG-CCUD)

Otro de los productos derivados de la PD-CCUD es la especializada en capas geoespaciales, es decir, datos que sólo tienen sentido en un contexto geográfico.

Esta plataforma es una herramienta que permite visualizar de manera geoespacial e interactiva los datos de las colecciones universitarias que se encuentran en la PD-CCUD. Muchos de estos datos, así como documentos asociados, tienen una referencia geoespacial para los cuales es posible dar una representación en un mapa. Esta plataforma permite afinar la visualización de manera interactiva, con el fin de dar una representación geográfica óptima, tomando en cuenta el inmenso volumen de datos y documentos a los que tiene acceso la PD-CCUD.

MARCO NORMATIVO

La gestión y publicación de los datos recopilados en distintas disciplinas del ámbito académico de la Universidad y almacenados en sus colecciones universitarias, requiere de un sólido marco normativo para garantizar su disponibilidad, visibilidad y utilidad sin afectar ninguna figura legal vigente. El marco normativo de la CCUD se compone de un reglamento interno, protocolos, lineamientos y procedimientos.

REGLAMENTO INTERNO DE LA CCUD

Se elaboró un documento para regular la organización y funcionamiento de la Coordinación de Colecciones Universitarias Digitales, estableciendo los criterios, procedimientos, lineamientos, entre otros, para lograr una eficiente integración de las colecciones universitarias que colaboren con la CCUD.

El contenido de este reglamento abarca los rubros que describen las actividades, alcances, responsabilidades y limitaciones de la CCUD para dar cumplimiento a los objetivos asignados. Este reglamento entró en vigencia el 28 de noviembre de 2014, a partir de la autorización de la Secretaría de Desarrollo Institucional de la UNAM.

PROTOCOLOS

Para cumplir los objetivos específicos de la CCUD es necesario contar con los planes escritos y detallados de los procesos a seguir, esto se ha realizado a través de la emisión de los protocolos correspondientes a cada tipo de proyecto que se realiza en la CCUD. Actualmente se cuenta con un protocolo general de proyectos y tres protocolos específicos para la digitalización de colecciones, para la integración de colecciones y para el desarrollo de software.

LINEAMIENTOS

Los objetivos de los lineamientos establecidos por la CCUD consisten en describir la ruta y los criterios que se utilizan para homogenizar los procesos de la CCUD, entre ellos, establecer las responsabilidades para el uso de la información y la publicación de datos en el Portal de Datos Abiertos UNAM (PDAUNAM). Estos lineamientos son de alcance a todos los usuarios internos y externos a la UNAM que accedan, descarguen, utilicen y/o redistribuyan información contenida en el PDAUNAM.

PROCEDIMIENTOS

Parte fundamental en la ejecución de proyectos es el aseguramiento de que las actividades se lleven a cabo de manera reproducible y trazable; esto se logra con la descripción de las instrucciones claras a través de procedimientos de operación. Actualmente se cuenta con los procedimientos para obtener los permisos de publicación por parte de los responsables de los acervos, la aplicación de la metodología de control de calidad, el levantamiento de censos de colecciones universitarias, la implementación de bases de datos y generación de fichas así como aquellos para la operación cotidiana administrativa de la CCUD.

VINCULACIÓN CON ENTIDADES

Constantemente se reciben solicitudes para la elaboración de nuevos proyectos, ya sean de digitalización, integración o desarrollo de software. Para dar seguimiento a estas solicitudes, se elaboró un protocolo general de proyectos. Dicho protocolo consta de las siguientes etapas:

1. La entidad interesada debe realizar una solicitud formal.
2. Se definen las metas y los objetivos del proyecto. En esta etapa se debe determinar si los alcances del proyecto están alineados con los objetivos de la CCUD.
3. Se realiza un análisis de factibilidad por parte de la CCUD.
4. Se desarrolla un plan de trabajo que integre las distintas etapas según el tipo de proyecto, incluyendo tiempos definidos por las cargas de trabajo de ambas partes.
5. Se genera y firma un documento o instrumento jurídico que formalice la relación entre la CCUD y la entidad colaboradora.
6. Se ejecuta el plan de trabajo acordado según los tiempos definidos con la entidad colaboradora.

BASES DE COLABORACIÓN

Se celebraron cuatro bases de colaboración con entidades universitarias para la integración de colecciones: el Instituto de Investigaciones Estéticas, el Instituto de Geología, la Dirección General de Asuntos del Personal Académico y la Facultad de Ciencias UNAM

Asimismo, se firmó un adendo al convenio tripartita para el desarrollo de una plataforma de visualización de obra artística con la Secretaría de Hacienda y Crédito Público y el Instituto de Investigaciones Estéticas -UNAM.

DIGITALIZACIÓN DE COLECCIONES

Del Instituto de Biología se digitalizaron, en 2014, las siguientes colecciones:

- Herbario Nacional MEXU; 656 451 ejemplares
- Colección Nacional de Arácnidos; 13 022 ejemplares
- Colección Nacional de Insectos; 60 301 ejemplares
- Colección Nacional de Mamíferos; 421 ejemplares

De la Facultad de Ciencias se digitalizaron 32 601 ejemplares de 20 colecciones.

INTEGRACIÓN DE NUEVAS COLECCIONES

Están en proceso de integración nuevas colecciones de cinco entidades universitarias.

Instituto de Biología:

- Fotocolectas biológicas
- Acervo histórico del Herbario Nacional MEXU
- Colección de Ejemplares Tipo del Herbario Nacional MEXU

Dirección General de Actividades Cinematográficas:

- Diversas colecciones de la Filmoteca (en procesos de ser censadas)

Escuela Nacional de Música:

- Archivo Silvestre Revueltas

Facultad de Medicina Veterinaria y Zootecnia:

- Imágenes de Parasitología
- Imágenes de Patología
- Microscopio Virtual

Instituto de Geología:

- Ejemplares Tipo de la Colección Nacional de Paleontología

Se dio seguimiento al proceso de integración a las 49 colecciones que se sumaron el año pasado.

Se realizaron 19 censos a distintas colecciones del Instituto de Biología, de la Facultad de Medicina Veterinaria y Zootecnia, de la Facultad de Ciencias y de la Red de Acervos Digitales de la UNAM (RADUNAM).

Asimismo, se definieron y documentaron estándares para homogeneizar e integrar datos de objetos digitales y proyectos universitarios. Este último se definió en colaboración con la Dirección General de Asuntos del Personal Académico (DGAPA).

Al proceso de control de calidad estructural ingresaron poco más de dos millones de registros, de los cuales 1 304 184 fueron calificados como aptos para su publicación al cumplir con los criterios de calidad principalmente en los campos de clasificación (técnica artística, materiales, taxonomía, etc.) y de ubicación geográfica: 989 080 del Instituto de Biología, 262 545 del Instituto de Investigaciones Estéticas, 25 454 de la Facultad de Ciencias, 13 759 del Instituto de Ciencias del Mar y Limnología, 9 974 de RADUNAM, 1 882 del Instituto de Geología, 763 de la Facultad de Medicina Veterinaria y Zootecnia, y 727 de DGAPA.

Los registros que no aprobaron el proceso de control de calidad están en revisión por los responsables de las colecciones.

DESARROLLO DE SOFTWARE

La experiencia adquirida dentro de la CCUD para integrar, estandarizar y contextualizar datos, es apreciada por otras entidades, tanto universitarias como extrauniversitarias, que requieren integrar distintas bases de

datos. Tal es el caso de la Dirección General de Promoción Cultural y Acervo Patrimonial (DGPCAP) de la Secretaría de Hacienda y Crédito Público, con quienes se lleva a cabo un proyecto de desarrollo de software para el manejo de la obra artística en resguardo de la DGPCAP, aprovechando la experiencia obtenida con el Instituto de Investigaciones Estéticas de la UNAM.

Actualmente se ha desarrollado un diagnóstico detallado de requerimientos a través de entrevistas con los usuarios, análisis de procedimientos y volumetrías con el fin de dimensionar y delimitar la futura elaboración de un nuevo sistema de gestión de obra artística. Este desarrollo cuenta con financiamiento externo a través de la figura de ingresos extraordinarios.

DIFUSIÓN

Como parte de la difusión de las actividades de la CCUD se realizaron presentaciones en reuniones académicas, se publicó un artículo y se distribuyó un folleto para dar a conocer los lineamientos de la CCUD:

- Cartel "Digitalization of the National Herbarium of Mexico: data validation and error correction". Congreso de The Botanical Society of America (julio 2014)
- Ponencia "Repositorio de datos abiertos: el caso de la UNAM", Coloquio de Repositorios Digitales y Acceso Abierto. Universidad Autónoma del Estado de México (agosto 2014).
- Ponencia "Retos y logros en la digitalización y publicación de las colecciones biológicas de la UNAM". Simposio Internacional de Preservación Audiovisual y Digital: Archivos Contemporáneos. Instituto Nacional de Antropología e Historia (octubre 2014).
- Publicación digital. Gernardt, D., Sánchez-Cordero, V., Melo Samper Palacios, U., Giménez Héau, J. y G. Salazar. 2014. "Digitalización del Herbario Nacional de México: Avances y retos del futuro". **Revista Digital Universitaria** 15(4): 1-13
- Folleto descriptivo del Reglamento y los lineamientos generales para el uso y publicación de los datos. Tiraje para distribución a responsables de colecciones y directivos de las entidades universitarias colaboradoras de la CCUD.

PORTAL DE DATOS ABIERTOS UNAM

Se ha concluido el desarrollo del Portal de Datos Abiertos UNAM (PDAUNAM) en su primera versión (versión alfa) para consultar, filtrar, visualizar y descargar los datos, objetos digitales y capas geoespaciales que han sido publicados a través de la Plataforma de Datos CCUD.

El Portal de Datos Abiertos UNAM, es un punto de acceso en línea que publica los datos de las colecciones universitarias que cuentan con autorización para su uso, reutilización y redistribución libre por cualquier persona, sin restricciones de *copyright*, patentes u otros mecanismos de control, siempre y cuando se respeten los términos de uso en su versión vigente.

El PDAUNAM permite que nuevos usuarios generen nuevo conocimiento a partir de la combinación de distintos acervos de datos o el análisis de patrones en grandes volúmenes de información.

Los objetivos principales de este portal son:

- Generar nuevas perspectivas de conocimiento en los procesos de investigación

- Estimular la creatividad y el desarrollo de áreas emergentes de investigación y docencia
- Ayudar en la mejora constante de la calidad y excelencia académica de la Universidad
- Mejorar la capacidad de respuesta a los nuevos problemas globales
- Contribuir en la innovación y mejora de herramientas, productos y servicios
- Transparentar los procesos de investigación que son usados para la publicación de artículos
- Fortalecer la apertura y rendición de cuentas de la Universidad a la sociedad

El PDAUNAM es un punto de acceso a los datos de la UNAM a través de la Plataforma de Datos CCUD y está conformado por un motor que ha sido denominado Minero, el cual, a través de una configuración *ad hoc* de Solr-Lucene para cada tipo de colección y una interfaz de usuario, analiza, estructura e indexa los datos de las colecciones universitarias.

El Minero permite consultar, de forma eficiente y en fracciones de segundo, millones de registros de las bases de datos, éstos se presentan como fichas digitales que integran la información del registro con objetos digitales, capas geoespaciales y con otras bases de datos universitarias y extrauniversitarias, a través de la tecnología de datos entrelazados (Linked data). Cada una de estas fichas tiene una URL fija, lo que permite que se acceda directamente a ella a través de un motor de búsqueda como Google.

La selección y descarga de los datos provistos por el Minero se hacen a través de una interfaz web (de acceso restringido actualmente). Esta interfaz permite al usuario de manera amigable realizar consultas por palabras, filtrar y agrupar los resultados, visualizar los datos de diversas maneras (gráficas, mapas, etc.), explorar las fichas de los registros y descargar en un formato universal (csv) los datos seleccionados. En esta misma interfaz se encuentra la documentación asociada a los datos, a las entidades que los proveen, así como a la CCUD.

PERSPECTIVAS

Una vez presentadas las herramientas desarrolladas en la CCUD a la comunidad académica y sociedad civil en general, deben desarrollarse las Unidades de Documentación, que tienen la misión de generar conjuntos de datos seleccionados de la Plataforma Digital de la CCUD (PD-CCUD), con el fin de ser utilizados para cumplir objetivos o desarrollar proyectos particulares. De esta manera se reforzará a la UNAM en su capacidad de responder de manera más eficiente a los problemas nacionales que requieran del conocimiento inter y transdisciplinario que ha sido generado y custodiado por las colecciones académicas universitarias a lo largo del tiempo y que ahora puede ser usado de manera óptima a través de la PD-CCUD.

Es fundamental, también, definir lineamientos para el manejo y publicación de datos dentro de la UNAM, con el fin de asegurar su permanencia y pertinencia a partir del uso de buenas prácticas que permitan su integración, uso y reuso por distintos tipos de usuarios. Esto debe promover nuevas políticas universitarias de manejo de información que permitan que nuevos usuarios generen nuevo conocimiento a partir del acceso a los acervos de datos de la UNAM.

Además de continuar cumpliendo con las metas y objetivos por los que fue creada la CCUD, deben alcanzarse nuevas metas que han surgido a partir de la experiencia del trabajo con las colecciones durante estos dos años y de la aplicación de nuevas políticas de manejo de información a nivel nacional e internacional.

En nuestro país, en los últimos meses, se han vivido cambios radicales en lo que refiere al acceso a la información gubernamental y académica. Prueba de esto son la Estrategia Digital Nacional presentada por el

gobierno federal el 25 de noviembre de 2013 y el decreto por el que se reforma la Ley de Ciencia y Tecnología del 20 de mayo de 2014, donde se anuncia la implementación de un Repositorio Nacional de Acceso Abierto para las publicaciones académicas financiadas con fondos del Conacyt.

La CCUD busca dotar a la UNAM con las herramientas y experiencia de vanguardia para participar como líder en estos cambios hacia nuevas tecnologías de Acceso Abierto (Open Access), principalmente en lo referente a Datos Entrelazados y Abiertos (Open Linked Data).

