

–CCUD– Coordinación de Colecciones Universitarias Digitales

Dra. Tila María Pérez Ortiz

Coordinadora ~ desde febrero de 2013

Con el propósito de contribuir al fortalecimiento y desarrollo de la investigación que realiza la Universidad, el 14 de enero de 2013, por Acuerdo del Rector, fue creada la Coordinación de Colecciones Universitarias Digitales (CCUD), quedando adscrita a la Secretaría de Desarrollo Institucional.

Las funciones de la Coordinación se modificaron y ampliaron para cumplir con la difusión del acervo universitario digital a través del Portal de Datos Abiertos UNAM, Colecciones Universitarias, mediante un nuevo Acuerdo del Rector publicado el 14 de septiembre de 2015, y posteriormente se ratificaron en el Acuerdo del rector Enrique Graue Wiechers que Reorganiza las Funciones y Estructura de la Secretaría de Desarrollo Institucional de la UNAM, publicado el 30 de noviembre del mismo año.

Las funciones de la CCUD están orientadas a desarrollar y administrar una plataforma digital transdisciplinaria, que dé visibilidad e integre de manera interoperativa los datos existentes de las distintas colecciones universitarias, con el fin de difundir entre los académicos y los alumnos, así como hacia 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.

Estos datos son integrados y publicados como Datos Abiertos de las colecciones universitarias digitales, con el fin de optimizar su uso para generar nuevo conocimiento y propiciar que la investigación de la UNAM tenga un mayor impacto para el análisis y solución de los principales problemas de nuestro país.

En ese sentido, la CCUD desarrolla un programa clave para la vinculación 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.

De acuerdo a lo anterior, los objetivos de la CCUD son:

- Implementar la Plataforma Informática de la CCUD (Aurora) para facilitar la integración, interacción y aprovechamiento de las colecciones existentes en la UNAM, bajo resguardo de las entidades y dependencias universitarias.
- Recibir y sistematizar la información sobre las colecciones científicas, humanísticas y artísticas de la Universidad.
- Definir estándares y protocolos para el manejo digital de las colecciones, con el fin de garantizar su interoperabilidad.
- Proveer de herramientas y asesoría a las entidades y dependencias para la integración de las colecciones digitales en formatos abiertos, con el propósito de promover la investigación transdisciplinaria, mediante el acceso a las colecciones y su información asociada.
- Difundir como Datos Abiertos las colecciones de la UNAM a través del Portal de Datos Abiertos UNAM, Colecciones Universitarias.

PRINCIPALES LOGROS

Dentro de los principales logros alcanzados por la CCUD durante 2015, destaca el desarrollo de la versión 2.0 del Portal de Datos Abiertos UNAM, Colecciones Digitales, que integra más de un millón y medio de registros provenientes de 29 colecciones, custodiadas por 10 entidades académicas y una dependencia universitaria. Hasta el momento este acervo proviene de colecciones biológicas, de obra artística, proyectos universitarios y diversos objetos digitales.

Este Portal es un punto de acceso en línea para la consulta de los datos, que cuentan con autorización para su uso y redistribución libre por cualquier persona, sin restricciones de *copyright*, patentes u otros mecanismos de control, siempre y cuando se respeten los términos de libre uso de Datos Abiertos de la UNAM.

El Portal de Datos Abiertos UNAM da visibilidad a los datos, objetos digitales y capas geoespaciales; además, permite la consulta y descarga de bases de datos en formatos abiertos y estructurados, con el objeto de:

- ✓ Generar nuevos conocimientos en los procesos de investigación
- ✓ Estimular la creatividad y el desarrollo de áreas emergentes de investigación y docencia
- ✓ Coadyuvar en la mejora constante de la calidad y excelencia académica de la Universidad
- ✓ Mejorar la capacidad de respuesta a los nuevos problemas globales
- ✓ Contribuir a la innovación y mejora de herramientas, productos y servicios
- ✓ Transparentar los procesos de investigación utilizados para la publicación formal de resultados
- ✓ Fortalecer la apertura y rendición de cuentas de la Universidad a la sociedad

Cada uno de los registros de este Portal se presenta como una ficha digital con una dirección de Internet única, donde se integran datos, objetos digitales (imágenes, audio y documentos) capas geoespaciales, así como datos entrelazados para cumplir con los estándares más altos de Datos Abiertos.

A través de este Portal es posible consultar y descargar en formatos abiertos y en fracciones de segundo millones de registros de las colecciones universitarias, así como generar mapas con información de distintas entidades y visualizar imágenes de muy alta resolución.

El Portal de Datos Abiertos constituye la parte visible de una infraestructura que reúne recursos, servicios y herramientas alrededor de cinco ejes principales:

1. Centro de Datos
2. Red de entidades proveedoras de datos
3. Desarrollo metodológico
4. Marco legal
5. Plataforma Informática Aurora

1. Centro de Datos

El Centro de Datos de la CCUD alberga una red de servidores, switches y unidades de almacenamiento mediante los que resguarda, procesa y publica los contenidos de las colecciones universitarias; asimismo, hospeda los diferentes servicios en línea que provee la CCUD.

Para cumplir con el objetivo de contar con infraestructura moderna, actualizada y segura, así como para responder al crecimiento de las colecciones, en 2015 se adquirió una nueva unidad de almacenamiento con capacidad para alojar hasta 72 Terabytes de datos.

En la misma línea de crecimiento, se añadieron dos servidores más al clúster de la plataforma de virtualización, con lo que se incrementó la capacidad y redundancia de la nube de la CCUD.

En términos de seguridad, se dotó al Centro de Datos con un sistema antincendios automatizado, que monitorea el espacio físico del centro para detectar y confirmar la presencia de fuego y, en caso de existir, genera las alarmas correspondientes y libera un gas que sofoca el fuego sin dañar ni interferir en el funcionamiento de los equipos. Finalmente, se instaló un sistema de cámaras de circuito cerrado para monitorear las 24 horas la actividad en el área.

2. Red de entidades proveedoras de datos

Uno de los principales ejes de trabajo de la CCUD es contar con una red de entidades proveedoras de datos, para lo cual son fundamentales las labores de vinculación con entidades y dependencias universitarias.

Durante el 2015 se realizaron 21 presentaciones de la versión 1.0 del Portal de Datos Abiertos UNAM, para pruebas y retroalimentación con los responsables de las colecciones integradas, así como con los titulares de las entidades y dependencias participantes o involucradas.

Estas fueron: Centro de Ciencias de la Atmósfera; Coordinación de Planeación, Presupuestación y Evaluación; Dirección General de Cómputo y de Tecnologías de Información y Comunicación, así como el programa Toda la UNAM en Línea; Dirección General de Asuntos del Personal Académico; Dirección General de Asuntos Jurídicos; Dirección General de Legislación Universitaria; Dirección General del Patrimonio Universitario; Facultad de Ciencias; Facultad de Medicina Veterinaria y Zootecnia; Instituto de Biología - Dirección, Departamento de Zoología, Departamento de Botánica, Jardín Botánico y Estaciones; Instituto de Ciencias del Mar y Limnología; Instituto de Geología; Instituto de Investigaciones Estéticas - Dirección y Colegio de Investigadores; Oficina del Abogado General; Red de Acervos Digitales Universitarios (RAD UNAM).

En estas mismas reuniones se recabaron ocho autorizaciones de publicación de Datos Abiertos firmadas por las siguientes entidades:

- Dirección General de Asuntos del Personal Académico
- Facultad de Ciencias
- Facultad de Medicina Veterinaria y Zootecnia
- Instituto de Biología
- Instituto de Ciencias del Mar y Limnología
- Instituto de Geología
- Instituto de Investigaciones Estéticas

Por otra parte, se realizaron doce asesorías a curadores de colecciones para el análisis y control de calidad de los datos de sus respectivas colecciones.

3. Desarrollo metodológico

Uno de los principales ejes de trabajo de la CCUD es el desarrollo de una metodología que permita optimizar los resultados de los proyectos y que garantice la calidad de los datos. Esta metodología incluye el levantamiento de un censo de cada colección, así como procedimientos y protocolos para los proyectos de digitalización, integración de colecciones y desarrollo de software.

CENSO DE COLECCIONES

A través de los censos de colecciones se reúne la información necesaria para que cada registro cuente con una referencia a su origen. El censo incluye datos descriptivos de la colección, así como información acerca del acervo al que pertenece, la entidad o dependencia que la resguarda, los responsables o curadores, entre otros. Esta información se integra en un catálogo de colecciones universitarias digitales de la CCUD.

A lo largo del año 2015 se realizaron tres nuevos censos, correspondientes a la Colección de Tipos del Herbario Nacional, a la colección de Fotocolectas (ambas del Instituto de Biología) y a la Biblioteca Digital *Silvestre Revueltas* de la Facultad de Música. Asimismo, se actualizaron los censos correspondientes a 112 colecciones asociadas.

DIGITALIZACIÓN

Se apoyó al Instituto de Biología para la digitalización de 342,926 ejemplares del Herbario Nacional y de las Colecciones Nacionales de Insectos, Moluscos y Mamíferos.

INTEGRACIÓN

Se integraron al área de Informática y Estadística de la CCUD 1'629,655 registros de colecciones (nuevos y actualizaciones) y tras aplicarse el protocolo de control de calidad, que consiste en el análisis y la validación lógica de los registros a través de catálogos, se clasificaron de la siguiente manera:

- 646,942 registros fueron consistentes con los catálogos y pudieron ser publicados inmediatamente.
- 853,982 registros fueron modificados por los curadores de las colecciones junto con personal de la CCUD para hacerlos consistentes con los catálogos.
- 128,303 registros fueron inconsistentes con los catálogos y fueron integrados a un proceso de revisión por el especialista.
- En total 1'500,924 registros fueron publicados después de aplicar el proceso de control de calidad.

Se definieron y documentaron cuatro estándares para la integración de estos datos provenientes de diversas fuentes, para lograr un acceso unificado a la información.

1. Estándar de Biodiversidad Darwin Core (DwC)
<http://www.ccud.unam.mx/docs/Est%C3%A1ndar%20de%20Biodiversidad%20Darwin%20Core%20DwC.pdf>
2. Estándar sobre objetos digitales Dublin Core cualificado (DC)
<http://www.ccud.unam.mx/docs/Est%C3%A1ndar%20sobre%20objetos%20digitales%20Dublin%20Core%20cualificado%20DC.pdf>
3. Estándar de Obra Artística de la UNAM
<http://www.ccud.unam.mx/docs/Est%C3%A1ndar%20de%20Obra%20Art%C3%ADstica%20de%20la%20UNAM.pdf>
4. Estándar de la UNAM de Proyectos Universitarios PAPIIT
<http://www.ccud.unam.mx/docs/Est%C3%A1ndar%20de%20la%20UNAM%20de%20Proyectos%20Universitarios%20PAPIIT.pdf>

DESARROLLO DE SOFTWARE

En 2015, la CCUD y la Dirección General de Acervo Patrimonial y Promoción Cultural de la Secretaría de Hacienda y Crédito Público (SHCP) celebraron el convenio de colaboración DGAJ-DPI-34-100415-175 para desarrollar un sistema de cómputo en línea para la administración del inventario de obra artística y la gestión de procesos curatoriales de las colecciones “Pago en especie” y “Acervo patrimonial”, cuya colección se compone de más de seis mil

obras (pinturas, esculturas, grabados y fotografías) de 700 artistas, entre los que se encuentran Diego Rivera, Rufino Tamayo, José Luis Cuevas y Francisco Toledo, entre otros.

Tomando como base su experiencia en el manejo de colecciones digitales y en el desarrollo de proyectos tecnológicos, la CCUD recabó las necesidades, modeló procesos, diseñó y desarrolló el nuevo Sistema de Gestión de Acervo Patrimonial (SIGAP) para la SHCP. Este sistema en línea es modular y está basado en servicios web para gestionar el inventario de obra artística y gestionar los procesos alrededor de éstas, además cuenta con la capacidad de generar y resguardar documentos administrativos (reportes, pases de salida, formatos para dictámenes, convenios, etcétera) y para visualizar imágenes de alta resolución en navegadores web.

4. Marco legal

La publicación de Datos Abiertos en línea requiere de un marco legal sólido que defina tanto las obligaciones de los usuarios, como los lineamientos que deben mantenerse para garantizar el apego a las leyes vigentes de derechos de autor.

En 2015 se llevaron a cabo 13 reuniones con la Oficina del Abogado General, la Dirección General de Asuntos Jurídicos y la Dirección General de Estudios sobre Legislación Universitaria con el fin de concluir e implementar los siguientes documentos que conforman el marco legal de la CCUD:

- ✓ Lineamientos para la Integración y Publicación de las Colecciones Universitarias Digitales en el Portal de Datos Abiertos UNAM, Colecciones Universitarias, aprobados por la CCUD, previa opinión favorable de la Dirección General de Estudios de Legislación Universitaria y publicados en *Gaceta UNAM* el 24 de septiembre de 2015.
- ✓ Términos de Libre Uso de Datos Abiertos de la UNAM, elaborados por la Dirección General de Asuntos Jurídicos y publicados 9 de octubre de 2015, en el Portal de Datos Abiertos UNAM, Colecciones Universitarias.

Esta documentación es relevante ya que sienta un precedente en el ámbito nacional al ser la UNAM una institución pionera en la elaboración de normas para la publicación de la información producto de la docencia e investigación. De igual manera, los Términos de Libre Uso constituyen una licencia universitaria para el uso de datos abiertos académicos.

Los lineamientos definen las normas de publicación en el Portal de Datos Abiertos de la UNAM para que las bases de datos que se generan como producto del quehacer académico universitario se pongan a disposición en formatos abiertos y para facilitar su uso, consulta, reutilización y redistribución. Estas normas se dividen en cinco capítulos cuyo contenido se refiere a:

Disposiciones generales. El primer capítulo menciona que los lineamientos son de observancia general y obligatoria para todas las entidades académicas y dependencias universitarias que generen, produzcan o administren información de los acervos digitales con posibilidades de publicarse como Datos Abiertos. El capítulo primero define los conceptos utilizados en el documento. Posteriormente, el segundo capítulo describe que la integración e

interoperabilidad de los datos abiertos de Colecciones Universitarias se realiza a través de la Plataforma Aurora de la CCUD, así como a los principios generales de la publicación en el Portal de Datos Abiertos UNAM. El tercer capítulo es fundamental para conocer las características mínimas de los Datos Abiertos, los criterios básicos para definir la prioridad de publicación de un acervo, los tipos de datos que se integran y los metadatos mínimos requeridos. De igual manera, el artículo 16 de este capítulo se refiere a la necesidad de que las entidades y dependencias cuenten con documentación que acredite que tienen los derechos de publicación y divulgación del material a publicar. Además, incluye un artículo que se refiere a la identificación única de cada registro en línea a través de un triplete que es necesario para la recuperación de cada uno de los objetos registrados en una base de datos. Finalmente, el cuarto capítulo describe las restricciones de publicación, es decir, a aquellos datos que por su naturaleza confidencial o reservada no son sujetos a publicación, tomando en cuenta que con su publicación se puede generar un daño a terceros o bien, a los objetos y ejemplares.

Los lineamientos establecen que la CCUD generará un Manual de Implementación para profundizar en todos los aspectos mencionados en la normatividad para la integración y publicación de datos en el Portal de Datos Abiertos UNAM.

Por su parte, los Términos de Libre Uso de Datos Abiertos son un esquema de licencia universitaria para que los usuarios conozcan las condiciones de uso y consulta de los contenidos del Portal de Datos Abiertos UNAM, Colecciones Universitarias. Estos términos se rigen por las leyes mexicanas y a ella se debe remitir para cualquier disputa relacionada con la interpretación y cumplimiento de los mismos.

Los términos descritos indican al usuario lo que debe entender por Datos Abiertos, así como el reconocimiento de los principios universitarios en la docencia e investigación, mismos que no acreditan ninguna postura que tome el usuario sobre la información. Finalmente, hay un punto dedicado a la obligación de citar siempre la fuente proveedora de la información, con ejemplos de cómo citar un registro o una consulta y el reconocimiento de que en caso de haber contenidos propiedad de terceros, el usuario tiene la obligación de solicitar autorización específica al titular de los derechos correspondientes. Las salvedades mencionadas en los términos se refieren a la protección de logos, marca y diseño del portal, así como la titularidad de las bases de datos, todas ellas propiedad de la UNAM.

5. Plataforma Informática Aurora

El Portal de Datos Abiertos UNAM de Colecciones Digitales se alimenta de la Plataforma Informática Aurora, realizada por el equipo de desarrollo de la CCUD. En ella, cada base de datos pasa por un proceso de migración, estandarización y de un estricto control de calidad.

La Plataforma Informática Aurora, a través de un conjunto de recursos, servicios y herramientas electrónicos específicos, da visibilidad, contexto e integración a las diferentes colecciones de la UNAM, permitiendo utilizar y compartir información de una gran variedad de fuentes especializadas de diversas procedencias, fomentado la colaboración, la integración y la transdisciplinariedad.

Durante 2015 se realizó el registro del código fuente de la Plataforma ante el Instituto Nacional del Derecho de Autor (Indautor), así como el registro de marca ante el Instituto Mexicano de la Propiedad Industrial (IMPI) como una marca propiedad de la UNAM.

La Plataforma informática Aurora, además del manejo de los datos, incluye un repositorio de objetos digitales y un geovisualizador.

REPOSITORIO DE OBJETOS DIGITALES

La Plataforma Aurora cuenta con un módulo para el resguardo y administración de objetos digitales. Este módulo es denominado Repositorio de Objetos Digitales (DOR por sus siglas en inglés). Su principal función es la de automatizar el proceso de organización de imágenes, documentos y audios, para ser procesados, indexados y expuestos para consulta y visualización en una sola plataforma. Cuenta con una API web donde se establecen accesos estandarizados utilizando un identificador de objeto digital (DOI en inglés) para obtener diferentes versiones de los objetos o metadatos asociados a estos. El DOR es capaz de dar tratamiento especial a imágenes de alta resolución, con el fin de ser visualizadas en navegadores web sin detrimento por su tamaño o del ancho de banda del usuario.

Durante el 2015 se integraron al DOR 804,000 objetos digitales, de los cuales 766,773 corresponden a imágenes de alta resolución y el resto a imágenes de baja resolución, documentos y audios.

Cada uno de los objetos digitales en el repositorio puede asociarse, en primera instancia, a un registro de metadatos para describirlo y, en una segunda instancia, a otros registros de las colecciones universitarias para poder ser visualizados en conjunto y enriquecer así el contexto de la información.

GEOVISUALIZADOR

El geovisualizador es una herramienta en línea que permite la visualización de las diferentes capas integradas a la CCUD, que se organizan en dos tipos principales: capas base y capas disponibles. Las capas base están conformadas por aquellas imágenes que permiten ubicar otros datos georreferidos en un contexto geográfico general, se incluyen las desarrolladas por Google Maps® (Google Satélite y Google Calles) y de iniciativas abiertas como Open Street Maps. Las capas disponibles son de diferentes temas que pueden integrarse con las primeras para conformar una visualización que ayude a la observación de un comportamiento o relación específica. Las capas disponibles en el geovisualizador pueden provenir de los datos de colecciones universitarias alojadas en la nube de la CCUD o de servidores de mapas ubicados en otras dependencias y entidades, dentro y fuera de la UNAM.

Adicionalmente, el geovisualizador es capaz de integrarse con otras aplicaciones del Portal de Datos Abiertos de la UNAM para incorporar representaciones visuales de datos georreferidos y filtrados por alguno de dichos sistemas, de tal forma que puede conseguirse una visualización de datos específicos combinados con capas de temas específicos y contextualizados geográficamente. El geovisualizador cuenta con herramientas para realizar traslaciones, acercamientos, manejo de opacidad y ubicación de las capas mostradas.

DIFUSIÓN

Se impartieron las siguientes conferencias por parte del personal académico de la CCUD en foros sobre preservación y publicación de acervos digitales:

- “La preservación de los documentos digitales: ventajas y retos” en la mesa de discusión teórica Principios, Criterios y Normatividad del INAH, 15 de abril 2015.
- “Buenas prácticas y consideraciones para repositorios de datos” en las Jornadas de Visualización Web UNAM 2015, 12 de junio 2015.
- “Datos abiertos en la investigación científica” en The International Wikimedia Conference, 17 de julio de 2015.
- “Publicación de Datos Abiertos de Colecciones Universitarias” en el marco del Segundo Encuentro Universitario de Ingeniería de Software y Bases de Datos realizado en el Centro de Exposiciones y Congresos de la UNAM, 29 de octubre 2015.

Se publicó una nueva versión de la página web institucional de la CCUD (www.ccud.unam.mx), donde se explica la metodología de trabajo y se publican los documentos tanto del marco legal, como los protocolos y lineamientos para la elaboración de proyectos.

Finalmente, como parte de los materiales de difusión elaborados para el Portal de Datos Abiertos UNAM, Colecciones Universitarias, se realizaron las siguientes publicaciones: folleto del Protocolo general para la ejecución de proyectos de la CCUD; folleto del Protocolo específico de integración de colecciones universitarias digitales; díptico informativo del Portal de Datos Abiertos UNAM, Colecciones Universitarias.

PERSPECTIVAS

El trabajo realizado durante 2015 se enfocó en el desarrollo de una plataforma de Datos Abiertos para la UNAM, la cual se concibe como un parte de una estrategia universitaria de plataformas abiertas digitales. Dentro de esta visión universitaria, la CCUD busca desarrollar una serie de plataformas complementarias e interoperativas que integren datos con publicaciones y capas geoespaciales a través de repositorios universitarios.

La CCUD, en colaboración con otras dependencias de la Secretaría de Desarrollo Institucional, enfrenta el reto de generar y documentar un marco conceptual con los lineamientos específicos que garanticen la interoperabilidad, la visibilidad y la permanencia de la información académica generada y puesta a disposición por la Universidad

