

Preservación digital

El caso del Portal de Datos Abiertos UNAM, Colecciones Universitarias

EDURNE URIARTE SANTILLÁN
Universidad Nacional Autónoma de México

INTRODUCCIÓN

El registro y resguardo digital de las colecciones contribuye a su preservación. Durante los procesos de captura de datos y catalogación de objetos digitales o digitalizados, especialistas, técnicos y académicos en bibliotecas y acervos históricos científicos y culturales dedicados a la gestión de los datos enfrentan desafíos similares. Son frecuentes las interrogantes relacionadas con cómo elaborar sistemas adecuados de catalogación y registro de datos; de qué manera es posible asegurar su vigencia y actualización; cómo evitar que se pierdan en el tiempo; cómo lograr su consulta, y quién debe validarlos (Salvador Benítez y Ruiz Rodríguez 2005). Estas inquietudes, entre muchas otras, apuntan a la necesidad de generar prácticas adecuadas, desarrollos tecnológicos y metodologías para documentar la captura y actualización.¹

1 En esta presentación se omite la distinción entre “datos” y “metadatos” ya que durante su desarrollo no se mencionan metodologías específicas de su gestión.

Además del beneficio que implica para la administración de acervos y colecciones, estas bases de datos pueden publicarse para que la información contenida se utilice por otras entidades y dependencias universitarias, así como por instituciones científicas y culturales externas. Para que este uso extendido de los datos sea posible, es necesaria una gestión adicional para integrar los con otras fuentes de información. Ésta es una de las funciones más importantes de la Coordinación de Colecciones Universitarias Digitales (CCUD), una dependencia de la UNAM dedicada a integrar y publicar, como datos abiertos, acervos, proyectos de investigación, publicaciones y repositorios a través del Portal de Datos Abiertos UNAM, Colecciones Universitarias (datosabiertos.unam.mx).²

La UNAM es depositaria de acervos que son de gran valor para el patrimonio universitario y nacional; éstos son una fuente inagotable de conocimiento científico, humanístico y cultural. Las colecciones que preserva se incrementan y enriquecen con el trabajo de sus académicos, quienes llevan a cabo tareas de administración y catalogación, las cuales dan origen a importantes bases de datos. Una manera de aprovechar su potencial es usar tecnologías de la información, que permiten administrar, gestionar y consultar las colecciones. Bajo la filosofía de acceso abierto, impulsada por la universidad, la CCUD desarrolló la plataforma informática Aurora, así como un protocolo de integración para que los contenidos puedan publicarse y ser reutilizados por diferentes usuarios. El procedimiento se enfoca en lograr una mayor homogeneidad y calidad de los datos, así como identificar aquéllos que se pueden compartir y publicar como datos abiertos. Además de esta revisión, una vez que éstos se publican, la consulta de los usuarios permite identificar inconsistencias que contribuyen a su conservación y actualización (CCUD 2017).

Para ejemplificar de qué manera se reutilizan los datos en la publicación de colecciones, tomaré un ejemplar del Herbario Nacional de México (resguardado en el Instituto de Biología) y lo

2 En adelante Portal de Datos Abiertos UNAM.

describiré en su relación con el Portal de Datos Abiertos UNAM. De igual manera, mencionaré brevemente lo que implica publicar colecciones en formatos abiertos y cómo trabaja la CCUD para hacer esto posible. La experiencia de gestión de datos de la CCUD no escapa a las inquietudes mencionadas en párrafos previos. Integrar y publicar datos es otra vertiente del trabajo con las colecciones que también requiere generar prácticas adecuadas y desarrollos tecnológicos efectivos. Espero que desde la Coordinación sea posible aportar elementos de reflexión en el contexto de la jornada de preservación de objetos sonoros.

LAS COLECCIONES UNIVERSITARIAS DIGITALES EN ACCESO ABIERTO

Los acervos, proyectos de investigación, publicaciones y repositorios que se encuentran en las diferentes entidades y dependencias universitarias generan diversos contenidos digitales de gran valor. Muchos de éstos pueden consultarse en acceso abierto ya que la UNAM ha sido una institución promotora de la apertura del conocimiento y generadora de diferentes plataformas con fines académicos, científicos y culturales. Sin embargo, estos recursos digitales regularmente se encuentran aislados, situación que dificulta su consulta.

El avance tecnológico y la oferta actual de recursos digitales en acceso abierto, datos abiertos y ciencia abierta, nos permite acercar estos productos académicos. El Portal de Datos Abiertos UNAM es una de las herramientas institucionales de la universidad diseñada para este fin; es un punto de acceso en línea que permite que los datos

- Se usen, compartan y distribuyan con las menores limitaciones
- Se utilicen bajo términos específicos de libre uso de datos abiertos, para generar nuevo conocimiento
- Sean interoperables con otras colecciones y fuentes de datos, de objetos y de sistemas de información geográfica, al interior de la UNAM y con fuentes externas.

Preservación de documentos...

La filosofía de publicación de datos abiertos del Portal tiene como referencia el contexto de Acceso Abierto de la UNAM, que promueve el uso de recursos digitales provenientes de las ciencias, las artes y las humanidades, así como la ciencia abierta (*Open Science*), una forma de definir la práctica de apertura del conocimiento científico y sus procesos de creación y diseminación, con la ayuda de las tecnologías de la información (Kulczycki 2016).

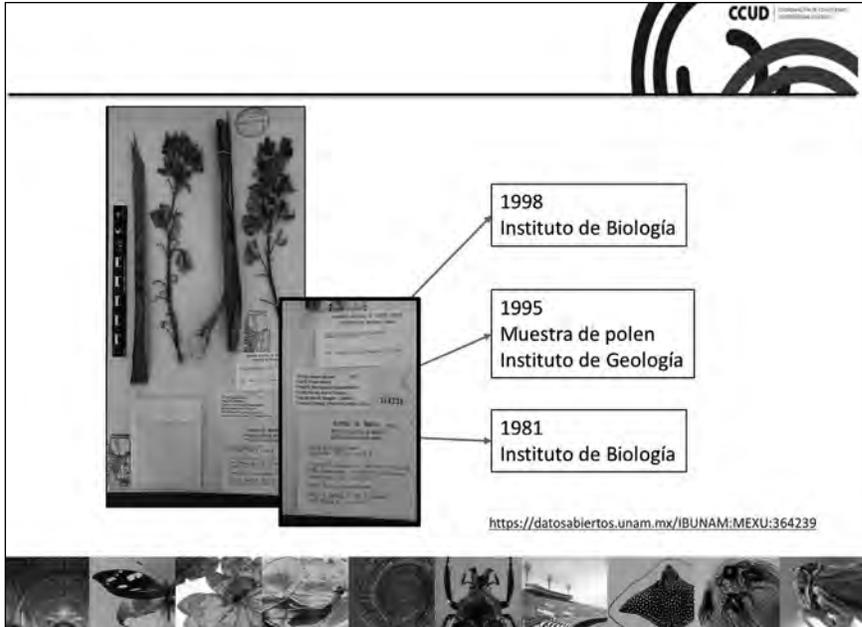
El Portal de Datos Abiertos UNAM hace públicos los datos de las colecciones para su uso como datos abiertos. Esto significa que cualquier persona los puede consultar y descargar bajo los Términos de Libre Uso de Datos Abiertos. Los datos abiertos son digitales, de carácter público y, en términos de las disposiciones aplicables, no tienen naturaleza reservada o confidencial; son accesibles en línea para ser utilizados, reutilizados y redistribuidos libremente por cualquier interesado. Los datos abiertos permiten:

- Ser descargados libremente en Internet en formatos accesibles y modificables.
- Ser reutilizados y redistribuidos bajo términos y condiciones que permitan su uso, reúso e integración con otros conjuntos de datos que pueden combinarse de maneras novedosas.
- Ser compartidos con instituciones de los sectores civil, gubernamental y educativo, en un marco de colaboración nacional e internacional, dado que no existen restricciones sobre su uso (CCUD).

La forma ideal de publicar datos abiertos es que éstos estén estructurados y disponibles para su descarga en formatos libres de restricciones. Tim Berners-Lee, inventor de la red informática mundial (*World Wide Web*), utiliza un criterio de cinco estrellas para mostrar cinco posibilidades de publicación de datos abiertos. Inicia con la primera estrella que se refiere a datos no estructurados, como tablas o información en un PDF que se puede compartir y redistribuir en acceso abierto y concluye con cinco, que corresponde a un formato óptimo, con datos estructurados, en formatos

no propietarios (esto es, que se pueden abrir sin software propietario como Excel), que cuentan con una URL³ única y, además, se pueden conectar con otros datos que le aportan valor a su contenido (*Linked Data* o “datos enlazados”).⁴

Imagen 1: Datos abiertos



La publicación como datos abiertos implica una revisión cuidadosa de la información ya que, al ser de acceso abierto, es importante vigilar que cumplan con los criterios de publicación de datos

3 *Uniform Resource Locator* (URL) por sus siglas en inglés, es un identificador de recursos digitales uniforme y único.

4 El esquema de cinco estrellas es utilizado con frecuencia para la identificar el grado de apertura de los datos abiertos. Se puede consultar más información sobre las características de las cinco estrellas en: <http://5stardata.info/es/>.

personales y que no se pongan en riesgo personas u objetos. La Coordinación de Colecciones Universitarias Digitales (CCUD)⁵ es la dependencia universitaria que integra y publica, de manera transdisciplinaria e interoperativa, las bases de datos de la información primaria obtenida de los acervos universitarios y de las investigaciones científicas, humanísticas y artísticas de la UNAM para su consulta en el Portal de Datos Abiertos UNAM.

La CCUD surge de la experiencia obtenida en el Proyecto Impulsa 5 de la UNAM, donde se creó el Sistema de Informática para la Biodiversidad y el Ambiente (SIBA).⁶ Su objetivo consistió en la integración de diferentes acervos universitarios vinculados a la biodiversidad y el ambiente. Para ello, una de las principales tareas fue la estructuración y puesta en marcha de una red de información para la organización, integración y vinculación de datos nacionales sobre dichas temáticas.

Los objetos y conjuntos de datos se integran con la finalidad de publicarse y difundirse a través del *Portal de Datos Abiertos UNAM, Colecciones Universitarias*. Para lograrlo, la CCUD desarrolla cinco ejes de trabajo:

1. Mantener de un centro de datos, que es el espacio destinado al sistema informático para mantener las bases de datos, los objetos y los servicios disponibles en línea.
2. Establecer una relación institucional cercana con autoridades y curadores ya que el éxito de una publicación recae en una estrecha colaboración con los proveedores de datos como los dueños y expertos en sus contenidos.

5 La CCUD es una entidad adscrita a la Secretaría de Desarrollo Institucional (SDI) creada por acuerdo del rector en 2013 y ratificada por la SDI en el Acuerdo que Reorganiza las Funciones y Estructura de la Secretaría de Desarrollo Institucional de la UNAM, publicado en *Gaceta UNAM* el 30 de noviembre de 2015.

6 Para más información sobre el Proyecto Impulsa 5, véase: http://www.cic-ctic.unam.mx/cic/mas_cic/megaproyectos/impulsa_5.cfm.

3. Elaborar metodologías para optimizar la publicación de datos de calidad. Es una de las tareas más importantes y se divide en dos etapas. La primera consiste en un proceso de arquitectura de información para estandarizar y normalizar las colecciones (arquitectura). El segundo se refiere a la gestión de los datos a partir de procesos de control de calidad. Este proceso incluye un análisis lógico y automatizado que después es revisado por los curadores. De esta manera, se detectan los registros que requieren de validación o revisión a detalle por parte de los especialistas.
4. Generar normas que garanticen que los datos se puedan publicar en formatos abiertos. Esto incluye generar lineamientos que eviten la publicación de información sensible o confidencial, así como datos personales que pudieran encontrarse en las bases de datos, o bien, condiciones específicas de las colecciones que pudieran afectar la integridad de personas u objetos.
5. Finalmente, el desarrollo tecnológico que posibilita la publicación e integración en línea de datos, capas geo espaciales y objetos digitales.

Por ello, la CCUD publicó los “Lineamientos para la integración y publicación de las colecciones universitarias digitales en el Portal de Datos Abiertos UNAM, Colecciones Universitarias”, en 2015.⁷ Los “Lineamientos” atienden a condiciones generales de publicación de contenidos. Para dar un mayor respaldo a la publicación, colaboramos con las entidades y dependencias para que generen condiciones específicas en caso de ser necesario.

7 Disponibles en http://ccud.unam.mx/docs/Lineamientos_PortalDatosAbiertosUNAM.pdf.

EL PORTAL DE DATOS ABIERTOS UNAM, COLECCIONES UNIVERSITARIAS

Para ilustrar algunos de los principales componentes del Portal de Datos Abiertos UNAM, muestro como ejemplo la visualización de un registro en el sistema de consulta. El ejemplar de la siguiente imagen es de la especie *Yucca mixtecana* (imagen 2). Pertenecer a la familia de los agaves, de la misma familia con la que se elabora el mezcal. Se preserva en el Herbario Nacional de México, dentro del departamento de Botánica, Instituto de Biología. Este ejemplar oaxaqueño se colectó en 1981 por tres investigadores entre quienes se encuentra el determinador (especialista que describe sus características taxonómicas), el doctor Abisaí García Mendoza, también especialista de la región. Su colecta es resultado del proyecto de investigación “La flora de Tehuacán”, apoyado por el Instituto de Biología.

Imagen 2: ejemplar

3. Portal de Datos Abiertos UNAM, Colecciones Universitarias
(datosabiertos.unam.mx)

Un punto de acceso en línea para consultar datos públicos de las colecciones digitales de la UNAM, puestas a disposición como datos abiertos para su uso, reuso y redistribución.



- 1 Consulta y combina los distintos acervos de datos
- 2 Integra miles de registros de colecciones en fracciones de segundos
- 3 Genera estadísticas y gráficas de visualización de resultados
- 4 Cuenta con los Términos de libre uso de Datos Abiertos en la UNAM
- 5 Retroalimentación de errores y comentarios



Parte de la historia taxonómica y del lugar en el que fue colectado el ejemplar se documenta en sus etiquetas. Este ejemplar muestra tres de ellas; la primera corresponde a la fecha de colecta; la segunda, de 1995, describe la toma de una muestra de polen realizada por otro investigador del Instituto de Geología de la UNAM para un proyecto diferente y, en 1998, el mismo curador que lo colectó y determinó, el doctor García, regresó al ejemplar y lo revisó, diecisiete años después. Este segundo análisis sirvió para una nueva determinación al renombrarse la especie (García 1998).

El ejemplar de *Yucca mixtecana* representa un punto en el tiempo, pues contiene información de una colecta hecha por un investigador de la UNAM y un trabajo de determinación y, posteriormente, se utilizó para nuevos estudios de esta especie y sus etiquetas lo documentan, así como el cambio del nombre científico de la planta. Estos acontecimientos narran un proceso científico de varios personajes e investigaciones asociadas a su estudio y los cambios ocurridos en el conocimiento taxonómico; nos muestra un proceso dinámico de la ciencia. De aquí es posible agregar que el conocimiento científico se construye a partir de la participación activa y continua de diferentes actores e instituciones que le imprimen ese dinamismo bajo referentes teóricos que también se circunscriben a un contexto histórico y del propio desarrollo del conocimiento (Merton 2002). Así como el conocimiento es dinámico, las comunidades científicas también lo son. Por lo tanto, las herramientas y los desarrollos tecnológicos que se construyen para capturar y consultar los datos obtenidos también deben ser flexibles.

El ejemplar de *Yucca mixtecana* (imagen 1) se encuentra en las colecciones de biodiversidad. En la siguiente imagen, podemos ver la página de inicio del Portal en su versión actual.

Imagen 3: Portal de Datos Abiertos UNAM.



Portal de Datos Abiertos UNAM
Colecciones Universitarias

Acceso a usuarios

TRÁNSFERA DE ALTA RESOLUCIÓN



Yucca mixtecana García-Mend.

Herbario Nacional de México (MEXU)
Plantas Vasculares

Datos de identificación

Número de la colección: Herbario Nacional de México (MEXU)

Asesoramiento de la colección: Dr. David Sebastián Gamard, jefe del herbario Nacional de México

Colectores asociados: Ramón Valderrama

Responsables de la colección asociada: Dr. David Sebastián Gamard, Curador de Plantas Vasculares

Depositarios: Departamento de Botánica, Instituto de Biología (IBUNAM)

INSTITUCIÓN: Universidad Nacional Autónoma de México (UNAM)
Identificador único (ID): IBUNAM-MEXU:184229

Número de catálogo o folio: 184229

Tipo de registro: Oculto Falso

Tipo de preservación: medicinal

Códiigo: M. en C. Adolfo Reyes García
Asesoramiento de la Sala de Actos

Datos taxonómicos

Apodo: Yucca

Nombre científico: Yucca mixtecana

Clase: Plantas Vasculares

Orden: Asparagales

Familia: Yuccaceae

Género: Yucca

Especie: Yucca mixtecana

Nombre científico: Yucca mixtecana García-Mend.

Detalles de la colección (ID): Yucca mixtecana García-Mend.

Apodo: Yucca mixtecana

Fecha de deposición: 1998-05-25

Datos de colecta

Colector (s): Gamard (M), A. Gamard T., B. Gamard & A. Gamard M.

Número de evento: 712

Fecha de recolección: 1987-05-11

Número de individuos (epítetos): 1

Apodo de tipo de espacio: Actos

Datos geográficos

País: México (MEX)

Estado: Oaxaca

Municipio: Ocotlán

Localidad: Ruta 195 ca. 15 km al NW de Tehuacan, a Rio de Oro

Altitud (m): 2200

OTR: 020

Información general

Cómo citar esta página: Departamento de Botánica, Instituto de Biología (IBUNAM), Yucca mixtecana García-Mend., depositario del herbario Nacional de México (MEXU), Plantas Vasculares, del Portal de Datos Abiertos UNAM (en línea), México, Universidad Nacional Autónoma de México. Dirección en: <http://datosabiertos.unam.mx/IBUNAM/MEXU/184229> Fecha de actualización: 22/05/2014, 12:30:26 p.m. Fecha de consulta: 04/10/2016, 9:04:59 a.m.

Copyright de la información: Para consultas relacionadas con proyectos de investigación, favor de enviar una carta solicitando al Dr. David Gamard. Correo: mgamard@ib.unam.mx

Para visitas de grupos con fines de educación o difusión, favor de dirigirse a René C. Moya de la Paz, Coordinador Educativo, encargado de Vinculación del Instituto de Biología, Correo: vinculacion@ib.unam.mx Tel: 53M 5148 y 5622 1076 Ext. 47839

INFORMACIÓN RELACIONADA EN BASES DE DATOS DE LA UNAM

17

Ejemplares de esta especie en la colección

16

Ejemplares de la misma especie con etiquetas digitales



SDI
Sistema de Datos Integrados

Únete al Portal de Datos Abiertos UNAM

¿Qué más ofrece?

Enlace Afiliados en la UNAM

Terminos de Uso y Condiciones de Acceso en la UNAM

Lineamientos para la Integración y publicación de los contenidos informáticos digitales

Contacto

Directorio

Mapa de sitio



106

Algunas de las herramientas disponibles en la versión actual son:

1. Mostrar la totalidad de los datos: la página muestra el total de registros que la integran y los datos que hay para cada tipo de colección. Los datos generales se presentan con la finalidad de que los usuarios accedan a un sistema que desde el inicio puede mostrar su contenido. Esta misma información se representa gráficamente.
2. Índice lateral: muestra los temas, colecciones, entidades y dependencias participantes y la referencia geográfica de los datos a nivel de país.
3. Sistema de información geográfica: es una herramienta de visualización de datos para representar los registros que cuentan con coordenadas.
4. Descarga de datos: las bases de datos consultadas se pueden descargar para su libre uso y reutilización como datos abiertos, con la restricción de citar siempre la fuente original.
5. Publicación de fichas de registros y objetos: cada registro de la una colección se publica con una dirección única que le permite ser citada y recuperada en Internet.
6. Botón de ayuda: los usuarios contribuyen al mejoramiento de los datos con la posibilidad de detectar errores, o bien, compartiendo su experiencia para mejorar los servicios de consulta.

Una ficha de registro, como la del ejemplar oaxaqueño, considera los siguientes datos:

1. Curación: es la información asociada al origen del ejemplar, es decir, a la entidad, la colección y los responsables de su curación, resguardo y catalogación. Es el contexto institucional del ejemplar.
2. Descripción del objeto o ejemplar: son los datos que describen al objeto propiamente; en este caso, son taxonómicos que provienen de la base de datos de la colección.

Preservación de documentos...

3. Ubicación geográfica: es el contenido relacionado con su ubicación espacial desde el país hasta la ubicación con coordenadas. En el ejemplo de las colecciones biológicas, se muestra el punto de colecta siempre que este dato sea público y no ponga en riesgo a la especie.
4. Información general: el Portal de Datos Abiertos UNAM es una vitrina para la consulta de los acervos universitarios y proyectos de investigación. Los datos pertenecen a las entidades y dependencias que los han generado y esta sección permite contactar a la colección directamente. Adicionalmente, al ser una publicación en línea, se incluye una manera sugerida de cómo citar esta publicación en línea.

Los elementos descritos se pueden ubicar en la ficha de publicación que se muestra en la imagen 3, que corresponde a la *Yucca mixtecana*:

Imagen 4: ejemplar en el Portal de Datos Abiertos UNAM.

The image shows a screenshot of the 'Portal de Datos Abiertos UNAM' (UNAM Open Data Portal) for the 'Colecciones Universitarias' (University Collections). The main record is for 'Yucca mixtecana García-Mend.' from the 'Herbario Nacional de México (MEXU), Plantas Vasculares'. The record includes a photograph of the plant, its name, the Herbario Nacional de México (MEXU), and curator information: 'Dr. David Sebastian Gernandt, Jefe del Herbario Nacional de México'. A callout box highlights related data counts: 17 for 'Información relacionada en bases de datos de la UNAM' and 16 for 'Datos relacionados'.

La información de los registros proviene directamente de los acervos y los datos de curación se obtienen a través de un censo

de colecciones elaborado por la CCUD.⁸ Para lograr la publicación integrada de diferentes acervos, la coordinación procesa las bases de datos originales por medio de una metodología que incluye su estandarización (arquitectura) y control de calidad (gestión de datos) por medio de catálogos. Cada etapa del proceso se documenta cuidadosamente.

La documentación es un paso fundamental en este tipo de procesos ya que en ella se registra la definición de los campos, las relaciones entre las tablas, las modificaciones realizadas y los filtros de publicación. Una adecuada documentación ayuda a prevenir la pérdida de información y a conservar el historial de los datos en el tiempo; registra las decisiones que se toman, esenciales en la gestión de los datos.

Los procesos de integración y publicación de acervos digitales son muy dinámicos, es decir, se actualizan con frecuencia. Este sistema de consulta se sostiene en la Plataforma Informática Aurora, un desarrollo de la CCUD conformado por bases de datos estandarizadas, servicios web y un conjunto de servidores que permiten la integración y publicación de datos de colecciones, capas geoespaciales y objetos digitales. Aurora almacena grandes cantidades de información y permite su consulta de forma masiva, en fracciones de segundo, a través del Portal de Datos Abiertos UNAM.

A través del Portal, es posible reagrupar los datos de la misma colección para construir nuevos contenidos, además de vincularlos con otras fuentes. Ésta es información adicional que los enriquece pues les aporta contenidos que no tenían de origen.

En el caso de ejemplares biológicos, por ejemplo, éstos pueden mostrar el número de ejemplares dentro de la colección, los que cuentan con imágenes, o bien, asociarse con bases de datos de nombres comunes, entre otras posibilidades. Esta información se puede tomar de otras colecciones dentro de la UNAM o de fuentes externas. A este método para relacionar y publicar bases

8 Ésta es la que aparece específicamente en la sección de contexto y en la información general.

Preservación de documentos...

de datos estructuradas se le conoce como *Linked Data* o “datos enlazados”.⁹ Si retomamos el ejemplo de la *Yucca mixteca*, podemos observar en la ficha nueva información ubicada del lado derecho, con información adicional.

Imagen 5: Linked data.

CALIFICACIÓN DE LOS DATOS ABIERTOS:		
★	No estructurados (por ejemplo, PDF)	 PDF
★ ★	Estructurados (por ejemplo, archivo Excel)	 XLS
★ ★ ★	Estructurados y en formatos no propietarios (como texto separado por comas “CSV”)	 CSV
★ ★ ★ ★	Estructurados, en formatos no propietarios y con una URL única y estática	 RDF
★ ★ ★ ★ ★	Estructurados, en formatos no propietarios, con URL única y estática e interconectados (<i>Linked Data</i>) con otras bases de datos.	 LOD

ccud.unam.mx

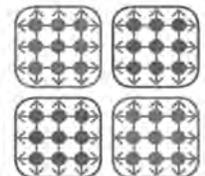
En síntesis, lo que podemos lograr con esta integración es una gran constelación de datos de colecciones, resultados de investigación y publicaciones, con información a largo plazo se pueda conectar con datos de otras fuentes y otros acervos, también en formatos abiertos. Bajo esta perspectiva de acceso y datos abiertos, el proceso de integración y publicación en el Portal de Datos

9 Para más información sobre este método de integración y publicación, consultar <http://linkeddata.org/>.

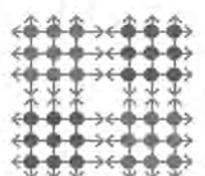
Abiertos UNAM enfrenta al menos tres retos importantes de las colecciones: visibilidad, contexto e integración.

Imagen 6: esquema CCUD.

PROBLEMAS COMUNES EN LAS COLECCIONES

VISIBILIDAD	CONTEXTO	INTEGRACIÓN
		
Los datos sólo pueden ser consultados a través de portales específicos. Forman parte de la web invisible y motores de búsqueda (como Google) no los pueden recuperar.	Las colecciones no siempre cuentan con información sobre su estructura, antecedentes, características, forma de uso, administración (autor, derechos, fechas), otros.	Las colecciones no comparten plataforma ni estructura de datos, es decir, no están estandarizadas y administrativamente están en diferentes entidades y dependencias.

VENTAJAS DE LA PLATAFORMA AURORA 

		
El contenido de las bases de datos se replica en formatos compatibles con motores de búsqueda de Internet para aprovechar los sistemas de búsqueda e indizado en la red.	Cada registro consultado desde Internet incluye los datos de la colección a la que pertenece, responsables, entidad o dependencia, y su relación con otras bases de datos.	Bases de datos estandarizadas y una plataforma que integra tecnologías de indexación, etiquetado de datos, recursos y una red de servicios web (Plataforma Aurora).

ccud.unam.mx

Preservación de documentos...

Estos retos son, a su vez, beneficios para la gestión de las colecciones. En primer lugar, la visibilidad se refiere al acceso y ubicación en línea de los productos y contenidos de las colecciones universitarias. La visibilidad en el Portal significa que cada ficha, es decir, cada registro de cada objeto o ejemplar de las colecciones tiene su propia dirección y representa una publicación digital en la red. Actualmente, más de un millón y medio de estos registros han sido publicados para su consulta y descarga. Los datos de las colecciones que se encuentran activas y en permanente uso tienen una alta capacidad de preservarse en el tiempo, actualizarse y depurarse.

En segundo lugar, el contexto es fundamental ya que cada registro cuenta con sus datos de origen. Esta información permite que el usuario identifique la información de los datos y los objetos y, con ello, se otorga el crédito a los autores o responsables del resguardo y, además, permite que más estudiantes y especialistas en el tema se puedan poner en contacto directo con los responsables de la colección.

Finalmente, la integración de las colecciones es clave para poder reunir datos de distinto origen y naturaleza. Lograrlo requiere de una adecuada metodología para estandarizar bases de datos (flexibles al cambio y a la naturaleza de los datos de las colecciones universitarias), documentar adecuadamente cada paso de la gestión y, sobre todo, la limpieza o el control de calidad de los mismos, realizada con el apoyo de los especialistas que validan las adecuaciones que realiza la CCUD en el proceso.

LA CIENCIA COMO UNA FORMA DINÁMICA DE CONOCIMIENTO

A partir de la exposición de las principales características del Portal de Datos Abiertos UNAM y de la forma de trabajo de la CCUD, es posible concluir algunos elementos para la discusión de la preservación de objetos de origen digital:

1. El uso de los datos y la información de las colecciones digitales contribuye a su preservación ya que la mantiene vigente y actualizada. De esta manera, es conveniente generar tecnologías de la información que se puedan vincular con otras, creadas en la UNAM o externas, con la finalidad de enriquecer los acervos, en un esfuerzo colectivo a favor del conocimiento.
2. Desarrollar metodologías de gestión de los datos: la estandarización de las colecciones universitarias y el diseño de buenas prácticas para la gestión de los datos permite un seguimiento ordenado y sistemático de la información. Su gestión incluye una constante documentación para evitar la pérdida de contenidos en el tiempo y, adicionalmente, que se publiquen con otras colecciones universitarias y fuentes externas adicionales. Esta dinámica de publicación de acceso y datos abiertos enriquece y contribuye a cada colección pero también fomenta mayor comunicación entre especialistas y promueve la investigación transdisciplinaria.
3. Generar normas para la integración y publicación de los datos y los objetos, con atención a las leyes de propiedad intelectual, de protección de datos personales y de confidencialidad.
4. Mantener una comunicación abierta y permanente con los curadores o responsables de los acervos. La comunicación permite, desde un inicio, identificar las problemáticas que más aquejan a las colecciones; seleccionar los conjuntos más pertinentes para su publicación y conservación, y contribuye a la participación activa de la comunidad académica.

La preservación digital de los acervos universitarios, de sus datos en particular, con una mirada que incorpore su uso y reutilización, contribuye al desarrollo del conocimiento científico y cultural de nuestra sociedad. La UNAM, a través de sus acervos y resultados de investigación, puede ser un agente catalizador de formas innovadoras para la investigación y difusión de la ciencia y la cultura.

Preservación de documentos...

Las colecciones y los acervos históricos y de investigación universitaria resguardan información de alto valor histórico, científico y cultural. Los datos e información que generan son dinámicos en la mayoría de los casos; su preservación requiere del involucramiento activo de diferentes actores. Al identificar las características de sus contenidos, es posible diseñar bases de datos adecuadas para cada fin.

BIBLIOGRAFÍA:

CCUD (Coordinación de Colecciones Universitarias Digitales). *Manual de Datos Abiertos de Colecciones Universitarias Digitales*. México: Universidad Nacional Autónoma de México, 2017. Disponible en http://www.ccu.unam.mx/datos_abiertos_en_la_unam.

García Mendoza, Abisaí. “Una nueva especie de *Yucca* (Agavaceae) de Oaxaca y Puebla, México” en *Acta Botánica Mexicana*, 42, 1-5, 1998.

Departamento de Botánica, Instituto de Biología. *Yucca mixteca* García-Mend., ejemplar del Herbario Nacional de México, Plantas Vasculares, en Portal de Datos Abiertos UNAM (en línea). México: Universidad Nacional Autónoma de México. Disponible en: <http://datosabiertos.unam.mx/IBUNAM:MEXU:364239>.

Kulczycki, Emanuel. 2016. “Rethinking open science: the role of communication” en *Analele Universității din Craiova*, 37 (1). 2016). Disponible en https://repozytorium.amu.edu.pl/bitstream/10593/14864/1/Kulczycki_rethinking_open_science.pdf.

Merton, Robert K. “La ciencia y el orden social” en *Teoría y estructuras sociales*, México: Fondo de Cultura Económica, 2002.

Salvador Benítez, Antonia y Antonio Ángel Ruiz Rodríguez. “Metadatos para la preservación de colecciones digitales” en *Cuadernos de documentación multimedia*, 48-60, 2005. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=2048508>.

Términos de Libre Uso de Datos Abiertos en la UNAM emitidos por la Oficina del Abogado General de la Universidad Nacional Autónoma de México y publicados en el Portal de Datos Abiertos UNAM. Colecciones Universitarias, 2015. Disponible en <https://datosabiertos.unam.mx/informacion/terminosdeuso.html>.