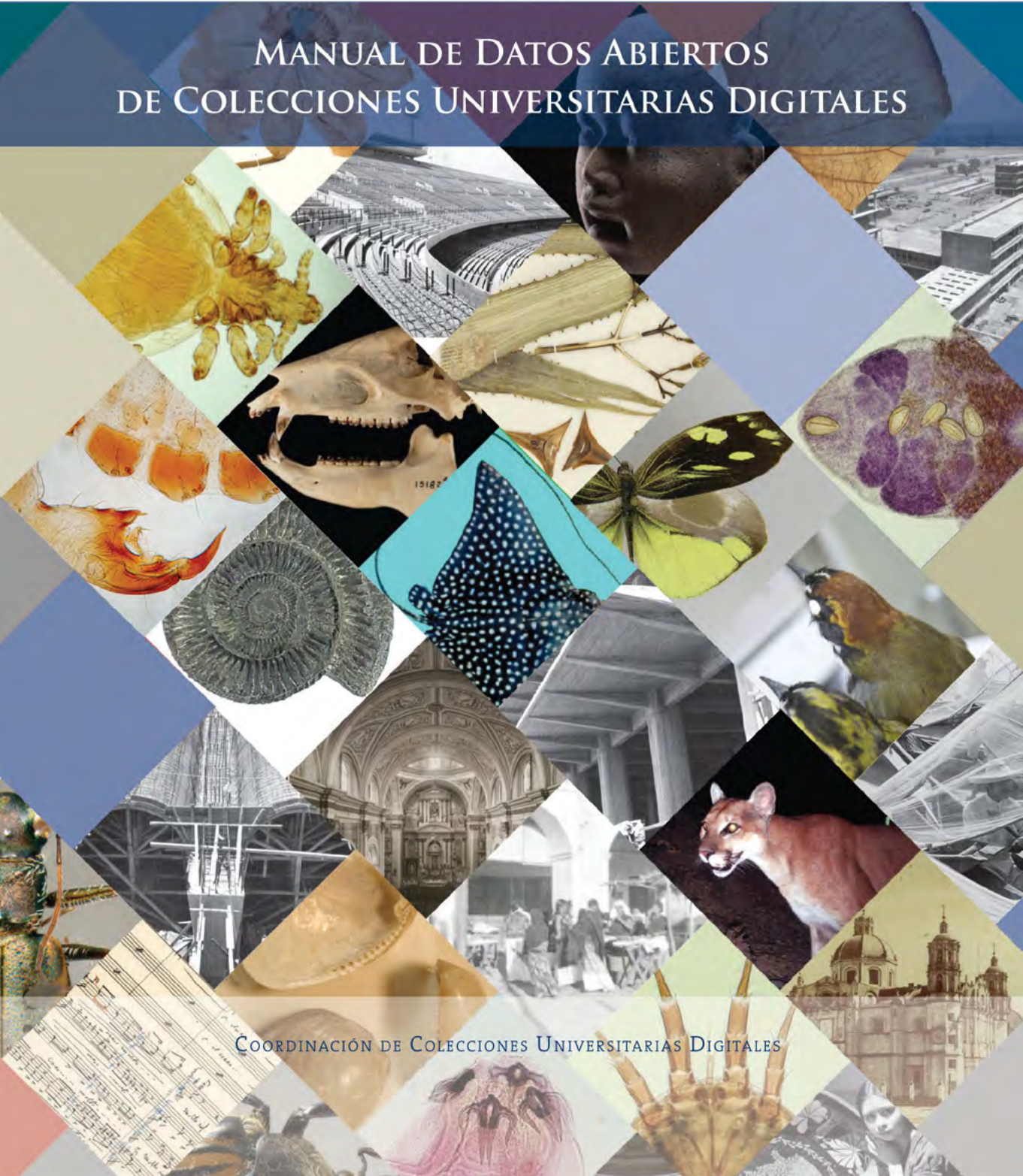




MANUAL DE DATOS ABIERTOS DE COLECCIONES UNIVERSITARIAS DIGITALES



COORDINACIÓN DE COLECCIONES UNIVERSITARIAS DIGITALES

1. *Datos Abiertos - México*; 2. *Investigación*;
3. *Ciencia Abierta*; 4. *Colecciones Universitarias Digitales – UNAM*.

COORDINACIÓN

Tila María Pérez Ortiz
Joaquín Giménez Héau

EDICIÓN Y REDACCIÓN

Eduarne Uriarte Santillán

CONTENIDOS

Ariana Chávez Méndez
Joaquín Giménez Héau
Daniel Pérez Castillo
Tila María Pérez Ortiz
Rubén Sáenz González
Eduarne Uriarte Santillán

DISEÑO

Maricarmen Miranda Diosdado

Las imágenes de esta publicación se realizaron con el apoyo y autorización de las entidades académicas proveedoras de datos del *Portal de Datos Abiertos UNAM*, *Colecciones Universitarias* (datosabiertos.unam.mx).

© D.R. Derechos reservados, 2017
Universidad Nacional Autónoma de México
Ciudad Universitaria, 04510, Ciudad de México

Coordinación de Colecciones Universitarias Digitales
www.ccuud.unam.mx

ISBN: 978-607-02-9086-2
Impreso y hecho en México

Forma sugerida de citar:

CCUD. 2017. *Manual de Datos Abiertos de Colecciones Universitarias Digitales*.
Coordinación de Colecciones Universitarias Digitales. SDI-UNAM. México.



MANUAL DE DATOS ABIERTOS DE COLECCIONES UNIVERSITARIAS DIGITALES

Coordinación de Colecciones Universitarias Digitales

Primera edición

Universidad Nacional Autónoma de México

2017

ÍNDICE

Presentación	7
Ámbito de aplicación	11
Sobre la Coordinación de Colecciones Universitarias Digitales	15
Agradecimientos	19
Introducción	21
Capítulo 1. Colecciones Digitales y Datos Abiertos en la UNAM	29
1.1 Las Colecciones Universitarias Digitales	31
1.2 El <i>Portal de Datos Abiertos UNAM, Colecciones Universitarias</i>	32
1.3 La Plataforma Informática Aurora	38
Capítulo 2. Conceptos y políticas de apertura del conocimiento	41
2.1 Conceptos generales	41
2.1.1 Acceso Abierto (<i>Open Access</i>)	41
2.1.2 Datos Abiertos (<i>Open Data</i>)	42
2.1.3 Datos Masivos (<i>Big Data</i>)	46
2.1.4 Ciencia Abierta (<i>Open Science</i>)	47
2.1.5 Revolución de los Datos (<i>Data Revolution</i>)	48
2.2 Políticas de Datos Abiertos	49
2.2.1 Contexto internacional	49
2.2.2 Gobierno Abierto y Datos Abiertos en México	51
Capítulo 3. Integración y publicación de Colecciones Universitarias Digitales	57
3.1 Etapa inicial: seleccionar datos y tipos de colecciones	59
3.1.1 Identificar las colecciones	59
3.1.2 Identificar los Datos Abiertos	61
3.1.3 Identificar las prioridades	62
3.1.4 Clasificar los datos	63
3.2 Protocolo de integración y publicación de colecciones en el <i>Portal de Datos Abiertos UNAM</i>	66
Definición del proyecto	66
Paso 1. Reunión exploratoria	66

Paso 2. Entrega de la muestra de datos	67
Paso 3. Análisis de la muestra de datos	67
Paso 4. Reunión de inicio de proyecto	68
Paso 5. Formalización del proyecto	69
Paso 6. Registro del censo de colecciones	69
Arquitectura y gestión de datos	72
Paso 7. Transferencia de datos y objetos digitales	72
Paso 8. Definición del estándar	73
Paso 9. Migración de datos al estándar	77
Paso 10. Control de calidad estructural	77
Paso 11. Filtrado de datos	79
Paso 12. Presentación del diagnóstico de control de calidad estructural	80
Paso 13. Entrega de la validación de datos modificados	81
Publicación de colecciones	82
Paso 14. Diseño de la ficha	82
Paso 15. Versión final de la base de datos a publicar	83
Paso 16. Permisos de publicación	83
Paso 17. Publicación de la colección	84
Paso 18. Entrega de proyecto	84
3.3 Actualizaciones	85
Referencias	87
Glosario	93
Anexos	99
1. Lineamientos para la integración y publicación de las Colecciones Universitarias Digitales en <i>el Portal de Datos Abiertos UNAM,</i> <i>Colecciones Universitarias</i>	101
2. Términos de Libre Uso de Datos Abiertos en la UNAM	111
3. Desplegable del Protocolo de integración	115

PRESENTACIÓN

Los datos son la materia prima en la construcción del conocimiento. Es fundamental clasificarlos, conservarlos, analizarlos, publicarlos, actualizarlos y reutilizarlos para la comprensión del mundo que nos rodea, descubrir su historia y vislumbrar escenarios futuros. Los datos generados y almacenados por los académicos de esta universidad tienen el potencial de contribuir en la investigación y la docencia universitaria y, al mismo tiempo, agregar valor al conocimiento en los diferentes sectores de la sociedad.

En las colecciones y proyectos de investigación de la universidad se reúne un acervo indispensable para la investigación de frontera. Los datos producidos pueden contribuir a la definición de políticas públicas que mejoren el bienestar social y la capacidad de respuesta a los nuevos problemas complejos que enfrenta el país, de escala global y de gran escala temporal, como son la seguridad alimentaria, la salud pública, la conservación de la biodiversidad y los ecosistemas.

Para lograr que un mayor número de personas de distintos ámbitos profesionales accedan a los datos de las colecciones universitarias, es necesario promover su apertura como un elemento indispensable para la innovación y el desarrollo de la sociedad. Sin embargo, existen muchas barreras que dificultan su disseminación, como son su dispersión, la heterogeneidad de sus formatos y la diversidad de fuentes de origen. Una manera de hacer frente a estos retos consiste en que los sistemas tradicionales de publicación y de evaluación académica se actualicen, que tomen ventaja de las nuevas capacidades para procesar, almacenar y transmitir conocimiento, asumiendo que la utilidad de la información aumenta en la medida en que es más accesible.

La Universidad Nacional Autónoma de México (UNAM) ha sido pionera en la aplicación y desarrollo de las tecnologías de la información (TIC) en México; fue a través de

esta universidad como nuestro país se conectó por primera vez a internet en 1982. Este liderazgo se ha mantenido y, en los últimos años, ha fomentado el aumento de su visibilidad en línea y el acceso abierto a sus contenidos académicos. Las TIC nos proveen una capacidad sin precedentes para adquirir, almacenar, operar y transmitir grandes volúmenes de datos, oportunidad que aprovechamos en la UNAM para compartir los acervos universitarios —científicos, humanísticos y culturales— para generar nuevo conocimiento.

Actualmente, los esfuerzos se han enfocado en lograr que el conocimiento creado en las facultades, escuelas, centros e institutos de esta universidad, esté disponible para ser reutilizado por la sociedad con la adopción de una política de datos abiertos y, en particular, con el lanzamiento del *Portal de Datos Abiertos UNAM, Colecciones Universitarias*. De esta manera se marca el camino hacia futuras formas de hacer ciencia y compartirla con la sociedad.

Para integrar, utilizar y difundir de forma eficiente el conocimiento generado en la UNAM, tenemos que aceptar y promover su intercambio abierto a través de publicaciones académicas formales, así como con la apertura de los datos que respaldan la investigación. Esto implica reforzar los esfuerzos ya realizados en las plataformas digitales estratégicas de la universidad que contemplan el Acceso Abierto (*Open Access*), los Datos Abiertos (*Open Data*) y los Datos Enlazados (*Linked Data*) e impulsar formas novedosas de promover también la investigación abierta, la ciencia ciudadana, mapas abiertos, entre otros.

La publicación de este *Manual* es uno de los resultados de los avances en materia de Acceso Abierto a la información que se genera en esta universidad, específicamente en su política de datos abiertos de colecciones digitales. En éste, los académicos universitarios encontrarán la información necesaria sobre esta iniciativa, así como la tecnología desarrollada, los pasos y buenas prácticas para publicar los acervos, proyectos de investigación, publicaciones y repositorios. De esta manera, las instituciones académicas pueden construir una gran red de conocimiento abierto para la investigación, la docencia y el interés de la sociedad en su conjunto.

Adicionalmente, el *Manual de Datos Abiertos de Colecciones Universitarias Digitales*, que me complace presentar, puede ser de utilidad para académicos y funcionarios

de otras universidades o instituciones para que compartan su conocimiento a través de este tipo de iniciativas. Mientras más instituciones de investigación se sumen a estos principios de apertura del conocimiento, contaremos con mejores instrumentos y herramientas para dotar a la sociedad de una mayor capacidad de participar activamente en la evaluación y elaboración de las políticas públicas.

ALBERTO KEN OYAMA NAKAGAWA
Secretario de Desarrollo Institucional, UNAM

ÁMBITO DE APLICACIÓN

De conformidad con el “Acuerdo que reorganiza las funciones y estructura de la Secretaría de Desarrollo Institucional”, publicado en *Gaceta UNAM* el 30 de noviembre de 2015, la Coordinación de Colecciones Universitarias Digitales (CCUD) tiene la función de expedir el Manual de Implementación de Datos Abiertos. Como resultado de este compromiso, se publica el *Manual de Datos Abiertos de Colecciones Universitarias Digitales*, que reúne las normas y directrices aplicables a todas las entidades y dependencias de la UNAM que integren conjuntos de datos a su portal (datosabiertos.unam.mx).

Las normas y procedimientos del presente *Manual* abarcan los temas mencionados en el capítulo II, numeral 9 de los “Lineamientos para la integración y publicación de las Colecciones Universitarias Digitales en el *Portal de Datos Abiertos UNAM, Colecciones Universitarias*”, publicados en *Gaceta UNAM*, el 24 de septiembre de 2015,¹ que incluyen:

- a) Las normas de operación del *Portal de Datos Abiertos UNAM, Colecciones Universitarias*.
- b) Las especificaciones técnicas que las entidades y dependencias universitarias deben observar para garantizar la interoperabilidad con el *Portal de Datos Abiertos UNAM, Colecciones Universitarias*.
- c) Las características para la integración de los datos y conjuntos de datos a través de la Plataforma de la CCUD y su publicación en el *Portal de Datos Abiertos UNAM, Colecciones Universitarias*.

1. LINEAMIENTOS (México: *Gaceta UNAM*, 24 de septiembre de 2015). Disponible en el Anexo 1.

- d) Las características y mecanismos de identidad que permitan la identificación digital de las colecciones pertenecientes a cada una de las entidades y dependencias universitarias.
- e) Los criterios para definir los estándares bajo los cuales deben estructurarse las bases de datos y conjuntos de datos, así como el procedimiento para el control de calidad que se aplicará a éstos.

La CCUD colabora con las entidades y dependencias proveedoras de datos por medio de proyectos de integración y publicación de colecciones. Éstos son de carácter institucional y siempre cuentan con el respaldo de sus directores. El *Manual* establece los conceptos y procedimientos necesarios para su realización.

SOBRE LA COORDINACIÓN DE COLECCIONES UNIVERSITARIAS DIGITALES

La Coordinación de Colecciones Universitarias Digitales (CCUD) es una dependencia de la UNAM que integra y publica las colecciones universitarias para su consulta como datos abiertos. Aquéllos que cuentan con la autorización para ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, se publican a través del *Portal de Datos Abiertos UNAM, Colecciones Universitarias* (datosabiertos.unam.mx).

La CCUD recupera la experiencia obtenida desde el 2004 en el proyecto universitario Impulsa 5, donde se creó el Sistema de Informática para la Biodiversidad y el Ambiente (SIBA),¹ cuyo objetivo fue publicar en línea diferentes acervos digitales sobre estas temáticas. Para ello, estructuró y puso en marcha una red de información para la organización e integración de datos biológicos y geográficos al fomentar la creación de unidades informáticas en las entidades participantes del Subsistema de la Investigación Científica. En una etapa posterior, SIBA replicó el modelo de trabajo con el acervo fotográfico del Instituto de Investigaciones Estéticas en 2010, integrando así datos de otras áreas del conocimiento.

La CCUD es resultado de este impulso institucional por expandir la red informática y adaptarla a un sistema que integrara, junto con las bases de datos y colecciones de biodiversidad, los acervos universitarios artísticos y humanísticos, así como los de otras áreas científicas.

1. SIBA se creó en 2005 con el objetivo de conformar un sistema modular de información de biodiversidad. Más información disponible en: <http://www.cic-ctic.unam.mx/cic/mas_cic/megaproyectos/impulsa_5.cfm>; fecha de la última consulta: 7 de octubre de 2016.

Dentro de los principales objetivos de la coordinación se encuentran los siguientes:

- a) Implementar una plataforma que facilite la integración, interacción y aprovechamiento de las colecciones existentes en la UNAM, bajo resguardo de las entidades y dependencias universitarias.
- b) Recibir y sistematizar la información sobre las colecciones científicas, humanísticas y artísticas de la universidad.
- c) Definir estándares y protocolos para el manejo digital de las colecciones con el fin de garantizar su interoperabilidad.
- d) Proveer con herramientas y asesoría a las entidades y dependencias para la integración de las colecciones digitales en formatos abiertos, a fin de promover la investigación transdisciplinaria² mediante el acceso a las colecciones y su información asociada.
- e) Difundir como Datos Abiertos las colecciones de la universidad a través del *Portal de Datos Abiertos UNAM, Colecciones Universitarias* (datosabiertos.unam.mx).

Creada por acuerdo del rector — *Gaceta UNAM*, 14 de enero de 2013 —, la CCUD es una dependencia adscrita a la Secretaría de Desarrollo Institucional. Las funciones de la coordinación se modificaron y ampliaron para cumplir con la difusión del acervo digital como Datos Abiertos a través del *Portal de Datos Abiertos UNAM, Colecciones Universitarias*, mismas que quedaron asentadas en el “Acuerdo que Reorganiza las Funciones y Estructura de la Secretaría de Desarrollo Institucional de la UNAM” — *Gaceta UNAM*, 30 de noviembre de 2015 —, del rector, el doctor Enrique Graue Wiechers.

2. La investigación transdisciplinaria se refiere al estudio realizado por facultades, escuelas, institutos, centros o programas de diversas disciplinas. De forma articulada e integrada, los proyectos desarrollan y fortalecen esfuerzos institucionales para el estudio y solución de algún tema, construyendo propuestas a partir de diferentes puntos de interés. Además, buscan trascender las fronteras entre diversas disciplinas.

AGRADECIMIENTOS

La experiencia de trabajo que se expresa en este *Manual* es resultado de la colaboración con curadores, académicos y autoridades de las diferentes entidades y dependencias proveedoras de datos. Sólo a través de una estrecha vinculación con ellos es posible integrar diversas fuentes de conocimiento derivadas de las Colecciones Universitarias Digitales y los proyectos de investigación académica.

Su trabajo es invaluable para alimentar, conservar y actualizar estos acervos. Igualmente valiosa es su disposición para publicar en línea los datos y objetos digitales que los componen, y así compartirlos abiertamente con otras comunidades académicas y con la sociedad.

Un especial agradecimiento a todos los académicos de las colecciones por su entusiasmo y sus aportaciones al proceso, así como a las autoridades de las distintas entidades y dependencias. Sin su respaldo institucional, la publicación del *Portal de Datos Abiertos UNAM* no hubiera sido posible.

INTRODUCCIÓN

En los últimos años han ocurrido grandes cambios en diferentes áreas del conocimiento, como la ciencia, la política, la economía e incluso en nuestra vida cotidiana, motivados por el crecimiento de las capacidades para producir y analizar información. Se estima que tan sólo en los últimos dos años se ha generado 90 por ciento de los datos con los que contamos en la actualidad. Junto con esta producción masiva, han surgido herramientas de almacenamiento, procesamiento y análisis de Datos Masivos (*Big Data*), constituyendo una transformación de gran impacto en las tecnologías de la información y comunicación (TIC). El aumento sin precedentes, en volumen y en capacidades de manejo, motiva nuevas perspectivas de comprensión y monitoreo de los acontecimientos naturales y de la sociedad; esto es lo que se llama la Revolución de los Datos (*Data Revolution*).

A pesar del incremento en el volumen de los datos, aún existe una brecha entre su producción y su uso por parte de los distintos sectores de la sociedad, lo que constituye uno de los principales retos a resolver para lograr una equidad en el acceso al conocimiento público. Los Datos Abiertos (*Open Data*) son una poderosa herramienta para zanjar esta brecha ya que garantizan su aprovechamiento de manera integral y gratuita a través de internet.

Los datos abiertos reducen los obstáculos para acercarse a la información, tomando en cuenta los rápidos, y a menudo impredecibles, cambios tecnológicos. Su publicación permite su uso y distribución por terceros, en su totalidad o formando parte de un conjunto construido a partir de diferentes fuentes, ya que se pueden hacer modificaciones y obras derivadas con los datos originales. Su potencial es muy amplio pues su uso no está restringido a ninguna persona o grupo por motivos de nacionalidad,

ámbito de trabajo y otros. Incluso, deben estructurarse de manera que puedan ser procesados automáticamente por computadoras, sin la intervención de personas.

La principal condición para el uso, reutilización y redistribución de los datos abiertos es el reconocimiento del trabajo intelectual y patrimonial de los proveedores —la autoría, los créditos de producción, la propiedad, la difusión, las restricciones de uso, los acuerdos financieros, las normas éticas y las condiciones de concesión de licencias—. Estas responsabilidades deben estar disponibles de manera clara y explícita como términos o licencias de uso.

El uso de los datos abiertos es una política que se asienta cada vez más en la gestión pública, aunque tiende a implementarse en nuevos ámbitos como en datos académicos de investigación. Dentro de este contexto, es fundamental reconocer la importancia de los datos contenidos en el vasto acervo que conforman las colecciones científicas, humanísticas y culturales custodiadas por las entidades y dependencias de la UNAM.

Con el fin de poner a disposición de la sociedad estos datos, en enero de 2013 se creó la Coordinación de Colecciones Universitarias Digitales, y el 9 de marzo de 2016, el rector de la Universidad Nacional Autónoma de México, el doctor Enrique Graue Wiechers, presentó el *Portal de Datos Abiertos UNAM, Colecciones Universitarias* como “una puerta al acervo cultural, humanístico y científico de esta casa de estudios.”

Para la fecha en que se publica este *Manual*, el *Portal de Datos Abiertos UNAM* cuenta con más de un millón y medio de registros provenientes de 29 colecciones biológicas, de obra artística, proyectos universitarios y otros objetos digitales alojados en distintas entidades y dependencias de la universidad. Con su apertura se espera promover nuevas perspectivas de conocimiento, así como estimular la creatividad y el desarrollo de áreas emergentes de investigación y docencia.

La publicación de datos abiertos fomenta la investigación transdisciplinaria al facilitar la coproducción de conocimiento entre actores académicos y no académicos. De igual manera, aumenta la capacidad de respuesta de la universidad y de la sociedad para enfrentar problemáticas coyunturales.

Los datos abiertos son indispensables también para fortalecer el acceso y la rendición de cuentas de la universidad a la sociedad, transparenta los procesos de investigación junto con la publicación de los resultados finales, adoptando así la política de Ciencia Abierta (*Open Science*).

El *Manual* responde al compromiso de la CCUD, en el marco de las políticas de la universidad para el Acceso Abierto (*Open Access*), de documentar de manera clara y eficiente el proceso para integrar un conjunto de datos al *Portal de Datos Abiertos UNAM*. Indica los pasos a seguir para la identificación de datos abiertos potenciales en una colección académica —acervos, proyectos de investigación, publicaciones y repositorios—, su arquitectura, gestión y publicación en línea. Así mismo, puede ser usado como una referencia directa sobre cada una de las etapas del proceso.

Además de cumplir con los compromisos asignados a la CCUD dentro de las políticas de acceso abierto, la realización del *Manual de Datos Abiertos de Colecciones Universitarias Digitales* tiene otros objetivos importantes como ofrecer información general acerca de los datos abiertos e iniciativas de acceso abierto; documentar la experiencia universitaria en la integración, gestión, publicación y la creación de los marcos tecnológicos, metodológicos y jurídicos necesarios para el funcionamiento de la política de datos abiertos de colecciones universitarias. Con esto se espera que también sea de utilidad para otras instituciones académicas interesadas en adoptar prácticas similares.

La estructura que se le ha dado al *Manual* busca cumplir con estos objetivos. El primer capítulo es una perspectiva general acerca de los avances de la UNAM en materia de datos abiertos: define las Colecciones Universitarias Digitales, el *Portal de Datos Abiertos UNAM* y la plataforma informática que lo alimenta. Todos los desarrollos mostrados en este capítulo han sido elaborados por la universidad, por lo que es importante su documentación y difusión para compartir las experiencias adquiridas.

El segundo capítulo está concebido como una guía de referencia rápida sobre datos abiertos y su aplicación en el contexto internacional y nacional. Aquí se define el concepto de datos abiertos, así como otros relacionados con la apertura, la generación y el análisis de datos. Debido a lo novedoso del tema, estos conceptos constantemente se actualizan o surgen nuevos; sin embargo, se ha cuidado que en este apartado se muestren en un contexto actualizado.

El tercer capítulo funciona como un manual de integración y publicación de colecciones universitarias digitales en la UNAM, menciona las características para definir, seleccionar y priorizar un conjunto de datos y el protocolo desarrollado por la CCUD para datos abiertos.

Finalmente, el *Manual* cuenta con un glosario y documentos anexos que complementan la información ofrecida.

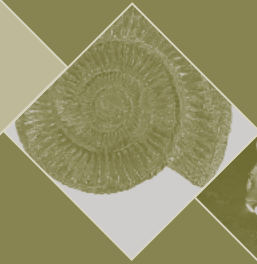
El *Manual* se suma a la publicación en *Gaceta UNAM* de los Lineamientos para la integración y publicación de colecciones en el *Portal de Datos Abiertos UNAM*, *Colecciones Universitarias*, así como a los Términos de Libre Uso de Datos Abiertos, completando la documentación acerca de datos abiertos en la universidad.

El *Manual* se dirige principalmente a investigadores, curadores, técnicos y desarrolladores responsables del manejo de bases de datos de colecciones universitarias. Además, constituye una herramienta valiosa para la comunidad académica —investigadores, profesores, técnicos o estudiantes— cuya producción intelectual o formación involucre la elaboración y gestión de bases de datos que cumplan con las características aquí mencionadas y, por lo tanto, tengan el potencial de publicación en el *Portal*.

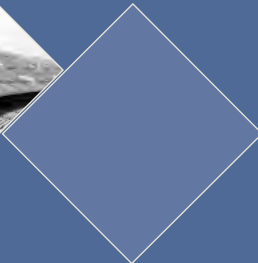
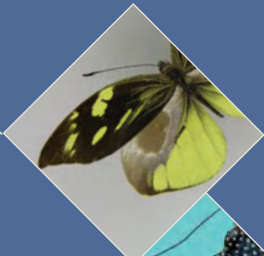
Esperamos que el *Manual* contribuya a las buenas prácticas de elaboración, conservación, actualización y publicación de las colecciones universitarias. Al hacer esto, queda constancia de la importancia de difundir la producción intelectual de los universitarios en un mundo que demanda mayor volumen y flujo de información para generar nuevo conocimiento.

TILA MARÍA PÉREZ ORTIZ
Coordinadora General, CCUD

JOAQUÍN GIMÉNEZ HÉAU
Coordinador Técnico, CCUD



CAPÍTULO 1



COLECCIONES DIGITALES Y DATOS ABIERTOS EN LA UNAM

1.1 Las Colecciones Universitarias Digitales

La Universidad Nacional Autónoma de México resguarda bienes culturales e intelectuales entre los que se encuentran las colecciones científicas, artísticas y humanísticas, además del conocimiento que produce y acumula en materia académica. Éstos no han quedado al margen del avance de las tecnologías de la información y comunicación, por el contrario, han derivado en versiones electrónicas a las que nos referimos como *Colecciones Universitarias Digitales*, las cuales son el conjunto ordenado de datos, de hechos tangibles e intangibles, reunidos y clasificados por su valor académico, cultural, histórico y de identidad universitaria, propiedad de la UNAM.¹ Sus datos se encuentran codificados en formatos electrónicos que se pueden leer y compartir mediante computadoras y otros dispositivos. Proviene de tres fuentes principales:

a) Acervos: fuentes de consulta, referencia o preservación, compuestas principalmente de datos primarios; por ejemplo, colecciones biológicas (Herbario Nacional, Paleontológica Nacional y Aves); fotográficas (Archivo Fotográfico Manuel Toussaint); mapas y capas geoespaciales (acervos de mapas históricos), entre otras.

b) Proyectos de investigación académica: datos generados en el desarrollo de proyectos de investigación (como experimentos o análisis estadístico) que sustentan los resultados de las publicaciones.

c) Publicaciones y repositorios: son datos que describen y clasifican los productos académicos terminados, como documentos, imágenes y otros objetos multimedia.

1. LINEAMIENTOS, *op.cit.*

El patrimonio intelectual de la UNAM presenta diversos obstáculos para su consulta debido a lo dispersa que se encuentra la información: distribuida en los acervos, proyectos de investigación, publicaciones y repositorios que se generan y conservan en diferentes facultades, escuelas, centros, institutos de investigación y museos de la universidad. El desarrollo de las tecnologías de la información facilita el acceso digital a este invaluable patrimonio histórico e intelectual, que se encuentra en las diferentes entidades y dependencias. Uno de estos esfuerzos universitarios para su integración es el *Portal de Datos Abiertos UNAM, Colecciones Universitarias* que se suma al camino abierto con proyectos como Toda la UNAM en Línea, la Red de Acervos Digitales UNAM y varios más apoyados por la Dirección General de Cómputo y de Tecnologías de la Información y Comunicación (DGTIC), entre otras instancias universitarias.

El *Portal de Datos Abiertos UNAM* nos permite acceder a bases de datos, objetos digitales (imágenes, documentos, audios y videos) y capas geoespaciales desde un solo punto de acceso para consultarlas como una colección mayor.

1.2 El Portal de Datos Abiertos UNAM, Colecciones Universitarias²

Para entender el *Portal de Datos Abiertos UNAM* es necesario definir brevemente el concepto de datos abiertos.³ De acuerdo con el manual de *Open Data* de la organización Open Knowledge International,⁴ éstos pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, sujetos al requerimiento mínimo de citar la fuente de origen.

El *Portal* es un punto de acceso en línea para la consulta de datos de colecciones universitarias que cuentan con autorización para su uso, sin restricciones de derechos de autor, patentes u otros mecanismos de control, siempre y cuando se respeten los Términos de Libre Uso de Datos Abiertos de la UNAM.⁵

2. En adelante, *Portal de Datos Abiertos UNAM*.

3. En el capítulo 2, apartado 2.2 se da una definición más amplia.

4. El manual de *Open Data*, en inglés, *The Open Data Handbook*, del Open Knowledge International. Disponibles en línea: <<http://opendatahandbook.org/guide/es/what-is-open-data/>>; fecha de última consulta: 7 de octubre de 2016.

5. Términos de Libre Uso de Datos Abiertos en la UNAM (México: Universidad Nacional Autónoma de México, 2016) Disponibles en el Anexo 2 y en línea: <<https://datosabiertos.unam.mx/informacion/terminosdeuso.html>>; fecha de última consulta: 7 de octubre de 2016.

Figura 1. Página de inicio del *Portal de Datos Abiertos UNAM*.



- 1 Consulta y combina los distintos acervos de datos.
- 2 Integra miles de registros de colecciones en fracciones de segundo.
- 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.

El *Portal de Datos Abiertos UNAM* tiene un diseño que permite la integración de distintos acervos digitales para su consulta. Su página de inicio (fig. 1) muestra en fracciones de segundo el contenido total de registros de las colecciones, así como las opciones para hacer más específica la búsqueda. En él se agrupan conjuntos de información que son comunes a todas las colecciones para que el usuario pueda seleccionar temas, colecciones, países o proveedores de datos (fig. 2). Al elegir uno de los conjuntos, el sistema reacomoda los datos y muestra nuevos subconjuntos, especializándose cada vez más en el tipo de colección a la que accedemos.

Figura 2. Ejemplo de filtros de consulta.

Temas

Colecciones Biológicas	1,359,780
Colecciones de Obra Artística	119,336
Objetos Digitales	56,693
Proyectos Universitarios	727

Colecciones

Herbario Nacional, Instituto de Biología	977,678
Insectos, Instituto de Biología	198,791
Archivo Fotográfico, Investigaciones Estéticas	119,336
Mamíferos, Instituto de Biología	30,554

Entidades y Dependencias

Instituto de Biología	1,340,552
Instituto de Investigaciones Estéticas	119,336
Facultad de Ciencias	24,914
Facultad de Música	18,765

Países

México	1,270,056
Estados Unidos	43,725
Guatemala	19,020
Costa Rica	15,965

Una segunda característica es su capacidad para generar gráficas y mapas. Las gráficas muestran los porcentajes de los datos contenidos en el sistema de consulta y a través de los mapas tenemos representaciones de distribución de los objetos y ejemplares, que se pueden comparar con información adicional como datos climáticos, división política, escenarios de cambio global, entre otros. Esta herramienta es muy valiosa para registros que cuentan con coordenadas geográficas.

En tercer lugar, cada registro de las bases de datos se publica como una ficha única (fig. 3), con una dirección propia y de fácil recuperación. Al seleccionar uno de los registros se abre una ficha en línea que cuenta con:

- a) Un URL⁶ propio y estático para su recuperación en internet.
- b) Herramienta de consulta de imágenes en alta resolución.
- c) Información de un registro dentro de la base de datos de la colección que describe a un objeto o ejemplar. Incluye datos de curación que muestran su origen como el nombre de la colección, la entidad o dependencia que la resguarda, el curador o responsable y el departamento al que pertenece; referencias de clasificación que describen las características del objeto o ejemplar. Por ejemplo, los datos de técnica fotográfica o taxonomía y colecta de ejemplares biológicos.
- d) Visualización geográfica: en caso de contar con información georreferida, la ficha muestra un mapa con su ubicación, indicando país, estado, municipio y localidad.
- e) Información de la entidad o dependencia proveedora de datos: el nombre de la colección e información de contacto.
- f) Información general: incluye cómo citar la ficha publicada en internet y el dato de contacto de la colección.
- g) Datos Enlazados (*Linked Data*): las colecciones comparten procesos de estandarización y criterios homogéneos para los datos, lo que permite que cada una de ellas se enriquezca con otras fuentes de información.

6. Los localizadores de recursos, o Uniform Resource Locator (URL), son direcciones únicas que sirven para encontrar contenidos en internet.

Figura 3. Características de las fichas de publicación de registros del Portal de Datos Abiertos UNAM.

1 <https://datosabiertos.unam.mx/IBUNAM:MEXU:557785>

The screenshot displays the following elements:

- 1:** The URL of the record: <https://datosabiertos.unam.mx/IBUNAM:MEXU:557785>
- 2:** A high-resolution image of a plant specimen, labeled 'IMÁGENES DE ALTA RESOLUCIÓN'.
- 3:** 'Datos curatoriales' section containing:
 - Nombre de la colección: Herbario Nacional de México (MEXU)
 - Responsables de la colección: Dr. David Sebastian Geramdi, Jefe del Herbario Nacional de México
 - Colectión asociada: Plantas Vasculares
 - Responsables de la colección asociada: Dr. David Sebastian Geramdi, Curador de Plantas Vasculares
 - Dependencia: Departamento de Botánica, Instituto de Biología (IBUNAM)
 - Institución: Universidad Nacional Autónoma de México (UNAM)
 - Identificador único (URN): IBUNAM:MEXU:557785
 - Número de catálogo o folio: 557785
 - Tipo de registro: Objeto físico
 - Tipo de preservación: Herborizado
- 4:** 'INFORMACIÓN RELACIONADA EN FUENTES EXTERNAS' section showing:
 - 146 Ejemplares de la misma especie en la colección
 - 140 Ejemplares de la misma especie en algunas colecciones
 - Number of external sources: (Fuente: Capital Natural de México 2016)
- 5:** 'Otros datos' section including:
 - Colectores: M. en C. Martha Oveera García, Responsable de la Sula Theresia
 - Fecha de colecta: 09/11/07
 - País: México
 - Localidad: La Paz, Baja California Sur
- 6:** 'Información general' section with a map and a 'Data citation' section.
- 7:** 'Data citation' section providing citation information for the record.

- 1 Cada ficha tiene un URL propio y estático para su recuperación en internet.
- 2 Una herramienta de visualización de imágenes en alta resolución.
- 3 Cada registro está acompañado de los datos que le dan contexto y lo remiten a su colección.
- 4 Los registros se vinculan como datos enlazados (*Linked Data*) con información proveniente de otras colecciones de la UNAM y de bases de datos externas.
- 5 La información de las fichas es localizable desde los motores de búsqueda (como *Google*).
- 6 Cada ficha es una publicación única en línea y va acompañada de su forma de citar.
- 7 Cada registro cuenta con el contacto de la colección.

El *Portal* aumenta la visibilidad de los datos, objetos digitales y capas geoespaciales de las colecciones para permitir su consulta y descarga en formatos abiertos y estructurados con el propósito de:

- a) Generar nuevo conocimiento en los procesos de investigación.
- b) Estimular la creatividad y el desarrollo de áreas emergentes de investigación y docencia.
- c) Ayudar en la mejora constante de la calidad y excelencia académica de la universidad.
- d) Mejorar la capacidad de respuesta a los nuevos problemas globales.
- e) Contribuir a la innovación de herramientas, servicios y productos.
- f) Transparentar los procesos de investigación utilizados para la publicación formal de resultados.

El *Portal* es la interfaz de consulta de la Plataforma Informática de la CCUD (Aurora), diseñada por la coordinación para el resguardo, manejo, integración y publicación de la información de las colecciones universitarias digitales.

1.3 La Plataforma Informática Aurora

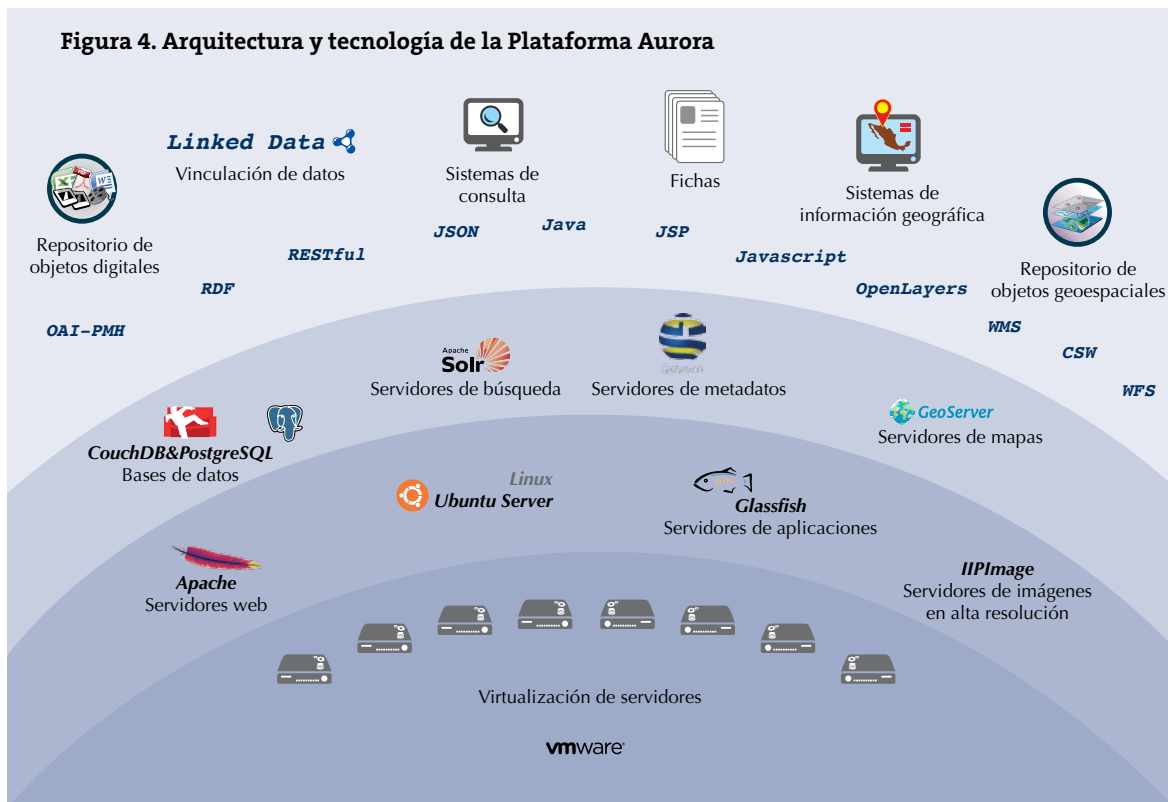
Aurora es un sistema modular que cuenta con herramientas para la exploración en línea de las colecciones universitarias digitales, despliegue de fichas y visualización de mapas geográficos, imágenes de alta resolución y otros objetos digitales. Es interoperable con otras plataformas tecnológicas ya que permite el uso y distribución de datos a través de servicios web, interfaces de programación de aplicaciones (API)⁷ y consultas a bases de datos no relacionales.⁸ Entre sus diferentes módulos podemos destacar dos:

- a) Repositorio de Objetos Digitales (DOR por sus siglas en inglés): su principal función es el resguardo y administración 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. El DOR es capaz de dar tratamiento especial a imágenes de alta resolución para su visualización en navegadores web sin detrimento por su tamaño o el ancho de banda del usuario.
- b) Repositorio de capas geoespaciales: resguarda y administra capas y datos geoespaciales, así como sus metadatos. Es capaz de almacenar hasta 50 mil capas y cuenta con mecanismos de acceso rápido a los contenidos. Utiliza protocolos estandarizados para compartir metadatos y capas con otras plataformas. Cuenta con una herramienta de visualización de mapas en línea con capas que pueden provenir de los datos de las colecciones o de servidores ubicados en otras dependencias y entidades universitarias o externas.

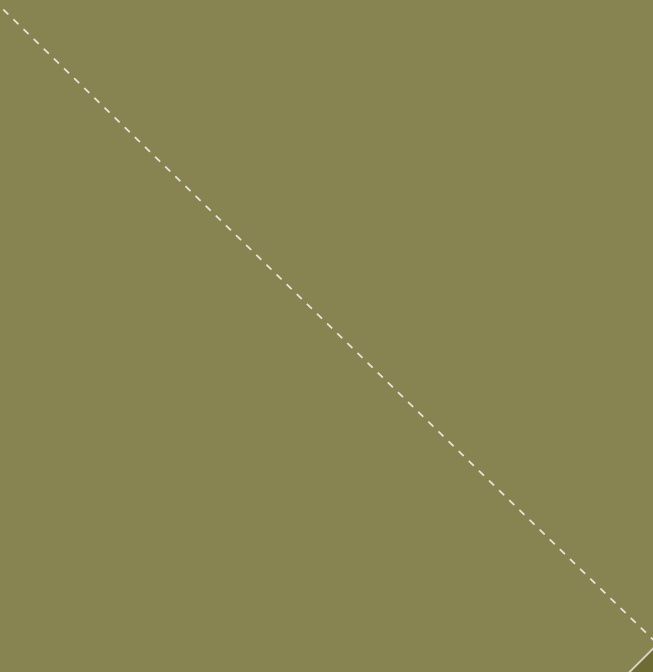
7. Por sus siglas en inglés (Application Programming Interface), se refiere a la interfaz de programación de aplicaciones o conjunto de métodos, funciones y procedimientos de uso general, que pueden ser utilizados por otro *software* sin necesidad de programar todo desde un inicio.

8. Es un modelo de bases de datos que se orienta al manejo de estructuras de datos, independientes y atomizados, para agilizar el almacenamiento y recuperación de grandes cantidades de información.

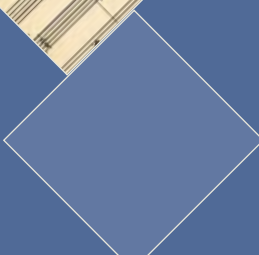
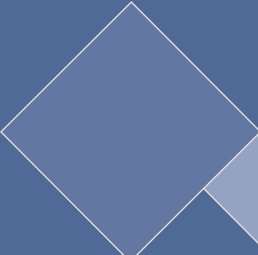
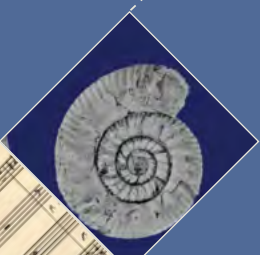
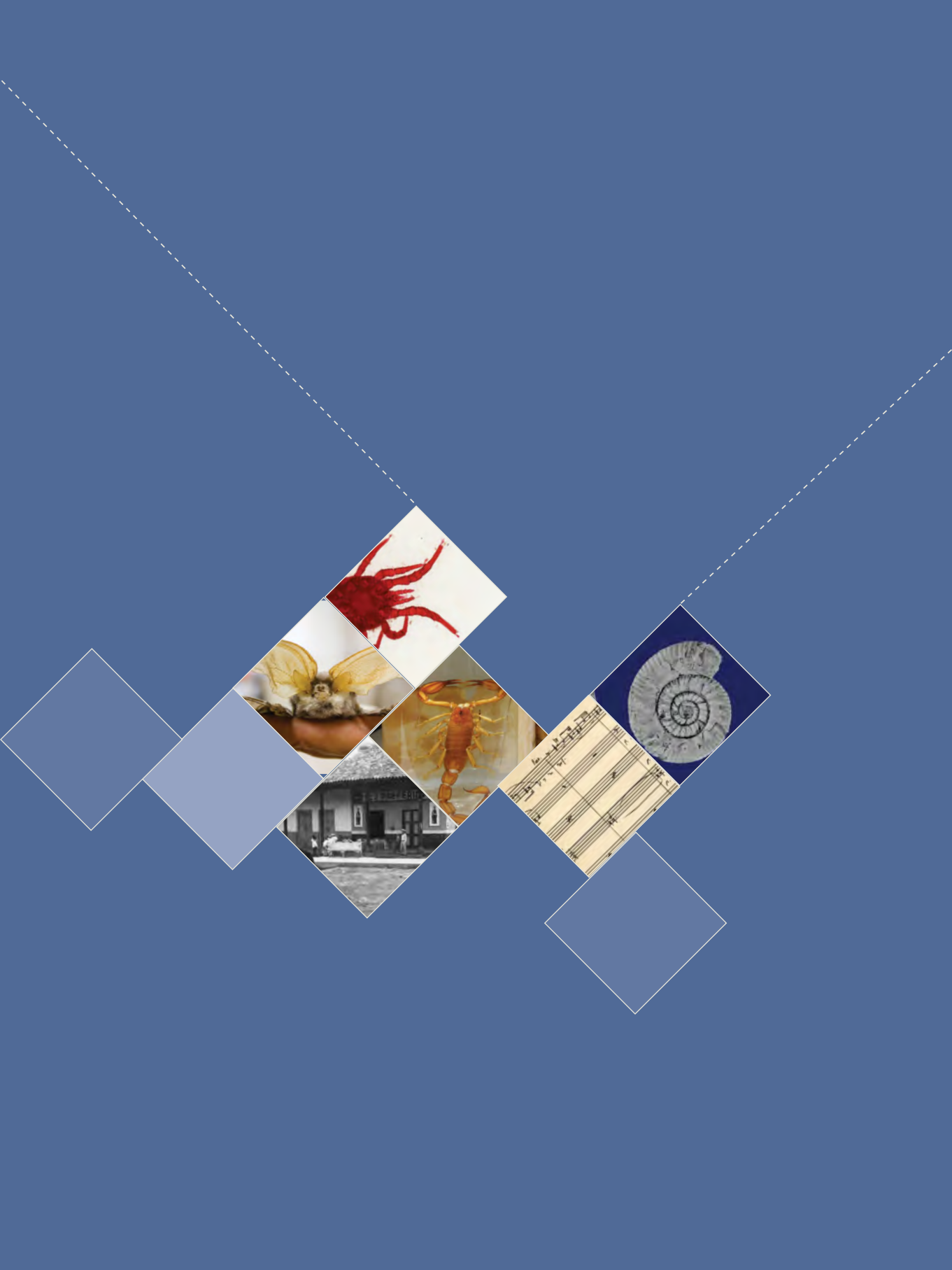
En la figura 4, se muestra un esquema general de la arquitectura y las tecnologías de Aurora:



La arquitectura y el diseño de Aurora aportan un marco tecnológico de frontera para ejecutar los procesos de integración, control de calidad, análisis y publicación de las colecciones en internet. *El Portal de Datos Abiertos UNAM* es el principal punto de acceso a la información que integra la plataforma Aurora, gracias a éste se fortalece la apertura y rendición de cuentas del trabajo académico de la universidad a la sociedad.



CAPÍTULO 2



CONCEPTOS Y POLÍTICAS DE APERTURA DEL CONOCIMIENTO

2.1 Conceptos generales

Vivimos un mundo digital e interconectado. El crecimiento en las tecnologías de la información tiene implicaciones en las actividades cotidianas, en el diseño de políticas públicas, en el acceso a la información y la participación ciudadana, así como en la ciencia y la forma en que se lleva a cabo la investigación a nivel global.

Los datos abiertos forman parte de movimientos internacionales y una política que promueve el acceso abierto como un derecho de interés general y una herramienta útil para la difusión del conocimiento a públicos más extensos. En este contexto, las universidades tienen un alto potencial para desarrollar tecnologías de uso de la información que contribuyan con la publicación de contenidos estratégicos.

En los últimos años, la UNAM ha fomentado el aumento de su visibilidad en línea a través de la apertura de resultados de investigación. Para comprender mejor el entorno digital en el cual se insertan los datos abiertos y su relevancia, es necesario familiarizarse con algunos conceptos.

2.1.1 Acceso Abierto (*Open Access*)

Es una política que busca extender el uso de contenidos digitales con la menor cantidad de restricciones posibles. En la UNAM, las iniciativas de Acceso Abierto tienen como finalidad la apertura de resultados de investigaciones, materiales educativos, académicos, científicos, tecnológicos, de innovación y culturales, de acuerdo con las licencias que se otorgan a los contenidos.¹ Por medio de la apertura de la información se procura eliminar, en la medida de lo posible, fronteras legales, económicas y técnicas para dar

1. LINEAMIENTOS, *op. cit.*

mayor visibilidad a estos recursos y que, a su vez, un mayor número de usuarios se beneficien de su uso y aplicación.

Algunos resultados centrales para lograr su difusión de manera global son la Iniciativa de Budapest para el Acceso Abierto (2002), la Declaración de Bethesda sobre Publicación de Acceso Abierto (2003) y la Berlin Declaration on Open Access to Knowledge in Sciences and Humanities (2003), a la cual la UNAM se adhirió el 22 de agosto de 2006.² Desde entonces, la universidad ha establecido un compromiso para el uso universal a los recursos en línea y ha mantenido una visión de apertura por medio de diferentes portales institucionales, que a su vez están registrados en el portal Toda la UNAM en Línea.³

2.1.2 Datos Abiertos (*Open Data*)

Los datos cobraron tal relevancia en años recientes que incluso generaron un movimiento propio, los datos abiertos; mejor conocidos por su término en inglés como *Open Data*. Éste es un término que se utiliza en distintos ámbitos gubernamentales, académicos y sociales, para referirse al acceso a datos digitales de carácter público y que, en términos de las disposiciones aplicables,⁴ no tienen naturaleza reservada o confidencial (se encuentran en línea para ser utilizados, reutilizados y redistribuidos libremente por cualquier interesado). Su uso permite:

- a) Descargarlos libremente de internet en formatos accesibles, estructurados y modificables.
- b) Reutilizarlos y redistribuirlos bajo términos y condiciones que permitan su uso, reúso e integración con otros conjuntos de datos.

2. Acuerdo por el que se establecen los Lineamientos Generales para la Política de Acceso Abierto de la Universidad Nacional Autónoma de México (México: *Gaceta UNAM*, 10 de septiembre de 2015). Documento en línea: <http://www.unamenlinea.unam.mx/files/TUL_AcuerdoLineamientosGeneralesPoliticaAccesoAbierto.pdf>; fecha de última consulta: 16 de octubre de 2016.

3. Página oficial: <<http://www.unamenlinea.unam.mx/>>.

4. Se refiere a un marco jurídico de normas de Acceso Abierto, Datos Abiertos, transparencia y derechos de autor.

Como se menciona en los Lineamientos,⁵ para considerarse como datos abiertos, éstos deben cumplir con las siguientes características:

- a) Accesibles: disponibles para la gama más amplia de usuarios, para cualquier propósito.
- b) De libre uso: citan la fuente de origen como único requerimiento para ser utilizados libremente.
- c) Gratuitos: se obtienen sin entregar a cambio contraprestación alguna.
- d) Integrales: contienen el tema que describe a detalle y con los metadatos necesarios.
- e) Legibles por máquinas: estar estructurados, total o parcialmente, para ser procesados e interpretados por equipos electrónicos de manera automática.
- f) No discriminatorios: disponibles para cualquier persona, sin necesidad de registro.
- g) Oportunos: actualizados periódicamente conforme se generan.
- h) Permanentes: se conservan en el tiempo; las versiones históricas relevantes para uso público se mantendrán disponibles con identificadores adecuados al efecto.
- i) Primarios: provienen de la fuente de origen con el máximo nivel de desagregación posible.

Los datos generados en distintos campos del conocimiento deben ser concebidos como abiertos desde su origen dentro de un marco normativo para su uso público. El resultado será una posición proactiva donde éstos se abrirán y serán publicados, anticipándose a cualquier necesidad o solicitud de información.

La calidad y utilidad de la publicación de datos abiertos se califica a partir de un esquema de cinco estrellas propuesto por Tim Berners-Lee, inventor de la red de informática mundial (World Wide Web). Este se utiliza mundialmente y define cinco conceptos generales, como se muestra en la figura 5:

5. Anexo 1, capítulo III, inciso 10.

Figura 5. Esquema de clasificación para la publicación de Datos Abiertos.

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

Este esquema nos ayuda a ubicar diferentes alternativas de publicación y permite evaluar su capacidad para relacionarse con otras plataformas y portales de información. Los datos abiertos idealmente deben presentarse de manera estructurada y ser descargables en formatos no propietarios, es decir, libres de restricciones económicas, identificados de manera única a través de un URL estático y con términos para su uso.

Los datos abiertos tienen un alto impacto en el progreso de los diferentes campos del conocimiento y en diversos sectores de la sociedad. Contribuyen a generar soluciones a problemáticas en temas tales como el desarrollo sustentable, la educación, la transparencia, la rendición de cuentas y la participación ciudadana. De esta manera, se colocan en el centro de la toma de decisiones de gobiernos, organismos civiles y no gubernamentales, iniciativas privadas e instituciones científicas y educativas para diseñar estrategias nacionales e internacionales de desarrollo.

2.1.3 Datos Masivos (*Big Data*)

La publicación masiva de información requiere de *software* y *hardware* que permita extraer su valor y generar contenidos para los diferentes usuarios. De esta necesidad

surge lo que ahora conocemos como Datos Masivos o *Big Data*. Este concepto define y engloba un conjunto de herramientas y tecnologías para el análisis, gestión, almacenamiento, interoperabilidad y consulta de grandes volúmenes de datos de diferente tipo y origen para su análisis.

Las herramientas y tecnologías de datos masivos poseen cuatro características definidas como las “4V”:

- 1) Volumen de datos: se refiere a la gestión de cantidades masivas de datos generadas diariamente alrededor del mundo.
- 2) Variedad de datos: capacidad de combinar información diversa para su uso.
- 3) Velocidad: capacidad para almacenarlos y utilizarlos además de reducir el tiempo de trabajo frente a herramientas de análisis tradicionales.
- 4) Veracidad: deben producir información veraz y de utilidad para la toma de decisiones.

Estas cuatro características son fundamentales para la obtención, uso y análisis de grandes volúmenes de información de alto valor social, económico y político, derivadas de una pluralidad de fuentes, idiomas y temáticas. Para trabajar con estos conjuntos de datos, se requiere que cuenten con estándares y protocolos que hagan posible su interacción a partir de una estructura homogénea.

2.1.4 Ciencia Abierta (*Open Science*)

Open Science, o Ciencia Abierta, es un movimiento internacional que tiene como objetivo dar libre acceso a las distintas etapas de la investigación científica, en particular a los datos que produce. Incluye la publicación de contenidos en acceso abierto y datos abiertos para su uso y redistribución. Entre los principales objetivos de este movimiento, podemos mencionar:

- a) Transparentar la investigación, su experimentación, metodologías, observaciones y procesos de colecta de datos.
- b) Publicar datos disponibles y reutilizables.

- c) Acceder de forma pública y transparente a la investigación científica y sus medios de transmisión.
- d) Facilitar la colaboración a través de diferentes herramientas digitales.

Este es un movimiento que crece para dar mayor acceso al conocimiento científico, entendido como una actividad que debe realizarse con procedimientos transparentes y compartidos, con la capacidad de reproducirse si todos los componentes son accesibles. Los datos abiertos son un detonante para la validación científica y el intercambio internacional, una exigencia para la apertura de los resultados del trabajo de investigación y la recopilación del proceso.

2.1.5 Revolución de los Datos (*Data Revolution*)

Tradicionalmente, los datos se han considerado de alto valor y, en general, se resguardan con mucho recelo o suelen intercambiarse por alguna contraprestación. Esto es así porque toma mucho tiempo generarlos, analizarlos e interpretarlos. De hecho, la mayoría de las veces no los conocemos en su nivel más primario y desagregado, sino procesados y publicados como parte de los resultados de investigación. Las políticas y movimientos de acceso abierto y ciencia abierta han permitido que esta actitud cambie sin que los datos pierdan su valor y ha derivado en una nueva manera de intercambiarlos.

Al mismo tiempo, el desarrollo de tecnologías de información y comunicación hace posible una mayor creación, distribución y análisis de conjuntos de datos en grandes cantidades. Muchas de las plataformas digitales se diseñan específicamente para hacer frente a la riqueza de contenidos y sus posibilidades de vinculación con otras fuentes. Esto se acompaña del diseño de metodologías para la estandarización de datos y la publicación de políticas de apertura.

El conjunto de filosofías, iniciativas, políticas y desarrollos tecnológicos y metodológicos para utilizar grandes volúmenes de datos (*Big Data*), imprime a la realidad una nueva manera de trabajar y de intercambiar información. Esta transformación e impacto de los datos en nuestra vida se le conoce como la Revolución de los Datos (*Data Revolution*) y se refiere a acciones realizadas por diferentes actores de la sociedad, encaminadas al uso y difusión de datos con valor y calidad.

2.2 Políticas de datos abiertos

2.2.1 Contexto internacional

Desde 2009 existen iniciativas notables en la apertura de datos de investigación y de gobierno. Los principales promotores provienen de diversas agencias y organizaciones gubernamentales y no gubernamentales como la Organización para la Cooperación y el Desarrollo Económico (OCDE), que promueve políticas públicas para un marco normativo de intercambio internacional de información. Los aportes de la OCDE aceleran el crecimiento en la escala, alcance y complejidad de los datos al generar guías y principios para el acceso a datos que provienen de las investigaciones realizadas con fondos públicos. Estos deben ser accesibles y utilizables para que contribuyan al conocimiento y al desarrollo de iniciativas sociales y científicas innovadoras.⁶

Por otro lado, el G8 — grupo constituido por Estados Unidos, Japón, Alemania, Italia, Francia, Reino Unido, Canadá y Rusia —, elaboró la Carta Internacional de Datos Abiertos, firmada en julio de 2013, que recientemente se actualizó para integrar a más países.⁷ La Carta es una iniciativa para la colaboración, desarrollo y aplicación de políticas y principios para la publicación de datos abiertos en los países miembro.

De manera complementaria a las políticas de datos abiertos de la OCDE y el G8, se encuentra la Organización de Naciones Unidas (ONU) que promueve el uso de información pública a través de diferentes portales de transparencia e investigación bajo el principio de dar acceso a las actividades de la organización y de los países miembro.⁸ La ONU generó, en colaboración con diferentes organizaciones, el documento *A world that counts*,⁹ para referirse a las posibilidades que actualmente existen para compartir y generar datos de sustentabilidad a partir del desarrollo de tecnologías y buenas prácticas.

6. Organización para la Cooperación y el Desarrollo Económico, *Principles and Guidelines for Access to Research Data from Public Funding*, documento en línea: <<http://www.oecd.org/sti/sci-tech/38500813.pdf>>; fecha de última consulta: 3 de octubre de 2016.

7. International Open Data Charter. <<http://opendatacharter.net>>; fecha de última consulta: 9 de mayo de 2016.

8. Algunos ejemplos serían el portal *UNdata, a world of information*, <<http://data.un.org/>> y el proyecto de datos abiertos *United Nations Development Program*, <<http://open.undp.org/#2016>>; fecha de última consulta: 7 de octubre de 2016.

9. Documento disponible en línea: <<http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>>; fecha de última consulta: 17 de octubre de 2016.

La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO por sus siglas en inglés) habilita datos útiles para el desarrollo de investigaciones e iniciativas llevadas a cabo en los diferentes sectores de la sociedad.¹⁰ Por su parte, el Banco Interamericano de Desarrollo (BID) cuenta con un Portal de Datos Abiertos¹¹ y una política propia de publicación, que incluye una serie de documentos sobre cómo integrar datos a través de su portal y algunas recomendaciones para promover su difusión. Además, por medio de diferentes oficinas, el BID realiza actividades para impulsar la publicación de Datos Abiertos en la ciencia con encuentros tipo *hackatón*.¹² Adicionalmente, a través de la Iniciativa Latinoamericana por los Datos Abiertos (ILDA), se apoyan diferentes proyectos y políticas sobre el tema para esta región.¹³

Dentro de las organizaciones no gubernamentales relevantes para la ciencia, encontramos diversos ejemplos que han sumado conocimientos para el desarrollo de políticas y prácticas de datos abiertos como el International Council for Science (ICSU), el comité interdisciplinario de ICSU, Committee on Data for Science and Technology (Codata) y The World Academy of Sciences for the advancement of sciences in developing countries (TWAS).¹⁴

Uno de los resultados recientes y más significativos de estas organizaciones es el acuerdo internacional “Datos abiertos en un mundo de grandes datos”,¹⁵ diseñado para el intercambio de la comunidad científica en el marco de las políticas públicas internacionales para la ciencia. El documento propone principios fundamentales y sugiere su adopción para aprovechar las oportunidades y retos de la revolución de los datos

10. El Portal de Transparencia de la UNESCO se encuentra disponible en: <<http://opendata.unesco.org/>>; así mismo, cuenta con la Iniciativa de Transparencia de la Ayuda Internacional, disponible en: <<http://www.aidtransparency.net/>>; fecha de última consulta: 21 de octubre de 2016.

11. Banco Interamericano de Desarrollo, *Numbers for development*, disponible en línea <<https://data.iadb.org/>>; fecha de última consulta: 2 de febrero de 2016.

12. Término que se refiere a los conceptos de maratón y *hacker*, refiriéndose a una experiencia colectiva que tiene como meta la colaboración para diversos fines asociados a la programación como el desarrollo de *software* y otros propósitos educativos y sociales.

13. Página oficial: <<http://idatosabiertos.org/>>.

14. Paul F. Uhlir y Peter Schröder, “Open data for global science” en *Data Science Journal*, Volumen 6, 17 de junio de 2007, disponible en línea: <<http://doi.org/10.2481/dsj.6.OD36>>; fecha de última consulta: 2 de febrero de 2016.

15. International Council for Science, *Open Data in a Big Data World*, disponible en línea: <<http://www.icsu.org/science-international/accord>>; fecha de última consulta: 2 de febrero de 2016. La versión abreviada en español se puede consultar en: <<http://www.icsu.org/science-international/accord/datos-abiertos-en-un-mundo-de-grandes-datos>>; fecha de última consulta: 21 de octubre de 2016.

para, de esta manera, contribuir al rigor de la investigación y maximizarla en beneficio del desarrollo de las naciones.

Otras organizaciones dedicadas específicamente a la promoción y difusión de datos abiertos son la World Wide Web Foundation,¹⁶ dedicada a generar propuestas innovadoras para la expansión de internet; Open Knowledge International,¹⁷ que genera indicadores de apertura de datos por país y tipo de contenido, y el Open Data Institute¹⁸ que aporta asesorías y consultoría sobre este tema.

Los datos de la UNAM se encuentran en una frontera de oportunidad en este proceso, al ser datos académicos es posible incrementar sus posibilidades de uso para el desarrollo de la ciencia y la difusión de la cultura. Además, la universidad cuenta con información valiosa para la toma de decisiones y propuestas de políticas públicas que contribuyan al desarrollo de los países y del bienestar social.

2.2.2 Gobierno Abierto y Datos Abiertos en México

Algunos gobiernos alrededor del mundo promueven la apertura de información para transparentar sus gestiones. De esta manera, colaboran en la difusión de los datos abiertos y la generación de marcos jurídicos para favorecer su distribución libre y gratuita. Existen iniciativas ejemplares como las ejercidas por el Reino Unido, Dinamarca, Finlandia, Estados Unidos y Canadá, y en América Latina de Colombia, Uruguay, Chile, Brasil y México.¹⁹

En septiembre de 2011, el Consejo General de la ONU lanzó la Alianza para el Gobierno Abierto (Open Government Partnership) con la misión de intercambiar experiencias y tecnologías para el manejo y publicación de datos abiertos. México fue uno de los ocho países fundadores de la Alianza, actualmente integrada por 64 naciones. Adicionalmente, en el periodo 2014-2015, asumió la presidencia del grupo.

16. Página oficial: <<http://webfoundation.org/>>.

17. Página oficial: <<https://okfn.org/about/>>.

18. Página oficial: <<http://theodi.org/about>>.

19. Open Knowledge International. *Global Open Data Index*, disponible en línea: <<http://global.census.okfn.org/>>, fecha de consulta: 1 de febrero de 2016.

Desde entonces, en el país se han decretado una serie de reformas y leyes para impulsar la apertura de datos de gobierno. Las más relevantes son:

- 1) Decreto por el que se reforman y adicionan diversas disposiciones de los artículos 6º, 7º, 27, 28, 73, 78, 94 y 105 de la *Constitución Política de los Estados Unidos Mexicanos*, en materia de Telecomunicaciones: en el artículo décimo cuarto transitorio indica que el Ejecutivo Federal tendrá a su cargo la política de inclusión digital universal que incluirá, entre otras cosas, el programa de gobierno y datos abiertos.²⁰
- 2) En 2013, el gobierno federal presentó la Estrategia Digital Nacional, que es parte del Plan Nacional de Desarrollo (2013-2018), y que incluye la Política Nacional de Datos Abiertos. El documento indica que las bases de datos que se generan y administran desde el propio gobierno podrán ser publicadas con ciertas características para su uso, reúso y redistribución, en beneficio de los diferentes sectores de la sociedad (iniciativa privada, comunicación, academia y sociedad civil). Para ello, se creó la licencia de Libre Uso MX.²¹
- 3) El 20 de mayo de 2014 se publicó en el *Diario Oficial de la Federación* el Decreto por el que se reforman y adicionan diversas disposiciones de la Ley de Ciencia y Tecnología, de la Ley General de Educación y de la Ley Orgánica del Consejo Nacional de Ciencia y Tecnología.²² Éste plantea lineamientos para la creación de un repositorio nacional de acceso abierto sobre recursos de investigación y ciencia. Parte del buen funcionamiento de este repositorio incluye la publicación y promoción de bases de datos y colecciones de información científica y tecnológica en formato digital, en libre acceso, interoperables con otras plataformas y repositorios.
- 4) El 4 de diciembre de 2014 se publicó el Acuerdo por el que se aprueba la

20. *Diario Oficial de la Federación*, (México: Secretaría de Gobernación, 11 de junio de 2013), documento en línea: <http://www.dof.gob.mx/nota_detalle.php?codigo=5301941&fecha=11/06/2013>; fecha de última consulta: 2 de febrero de 2016.

21. La información de la Política Nacional de Datos Abiertos se encuentra disponible en: <<http://mxabierto.github.io/iniciativa-datos-abiertos/documentos/politica/pnda/>>; fecha de última consulta: 2 de febrero de 2016.

22. *Diario Oficial de la Federación*, (México: Secretaría de Gobernación, 20 de mayo de 2014), documento en línea: <http://www.dof.gob.mx/nota_detalle.php?codigo=5345503&fecha=20/05/2014>; fecha de última consulta: 2 de febrero de 2016.

Norma Técnica para el acceso y publicación de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional.²³ En él se establecen las bases para la publicación de datos y conjuntos de datos del Sistema Nacional de Información Estadística y Geográfica en formato de datos abiertos y define sus características y restricciones. Posteriormente, en junio de 2015 publicaron su *Manual de Implementación*.

- 5) El 20 de febrero de 2015 se publicó el Decreto por el que se establece la regulación en Materia de Datos Abiertos²⁴ que norma la apertura de los datos generados por las dependencias y entidades de la Administración Pública Federal y de las empresas del Estado que sean de interés público.
- 6) El 4 de mayo de 2015 se publicó el Decreto por el que se expide la Ley General de Transparencia y Acceso a la Información Pública²⁵ que promueve la publicación de Datos Abiertos y accesibles por parte de los organismos públicos, como parte de los derechos de acceso a la información.
- 7) El 18 de junio de 2015, la Unidad de Gobierno Digital de la Secretaría de la Función Pública, en atención al Decreto por el que se establece la regulación en Materia de Datos Abiertos — 20 de febrero de 2015, mencionado en el punto 5 —, publicó la *Guía de Implementación de la Política de Datos Abiertos*.²⁶ Dicho documento describe los pasos técnicos para que las dependencias gubernamentales ingresen sus datos a la plataforma Adela, así como los formatos en que deben encontrarse los conjuntos de datos y su integración al Catálogo Único de Datos Abiertos del gobierno federal.

23. *Diario Oficial de la Federación*, (México: Secretaría de Gobernación, 4 de diciembre de 2014), documento en línea: <http://www.dof.gob.mx/nota_detalle.php?codigo=5374183&fecha=04/12/2014>; fecha de última consulta: 2 de febrero de 2016.

24. *Diario Oficial de la Federación*, (México: Secretaría de Gobernación, 20 de febrero de 2015), documento en línea: <http://dof.gob.mx/nota_detalle.php?codigo=5382838&fecha=20/02/2015>; fecha de última consulta: 2 de febrero de 2016.

25. *Diario Oficial de la Federación*, (México: Secretaría de Gobernación, 4 de mayo de 2015), documento en línea: <http://www.dof.gob.mx/nota_detalle.php?codigo=5391143&fecha=04/05/2015>; fecha de última consulta: 2 de febrero de 2016.

26. *Diario Oficial de la Federación*, (México: Secretaría de Gobernación, 18 de junio de 2015), documento en línea: <http://www.dof.gob.mx/nota_detalle.php?codigo=5397117&fecha=18/06/2015>; fecha de última consulta: 2 de febrero de 2016.

Cada vez son más estados e instituciones gubernamentales los que utilizan este marco normativo para el desarrollo de políticas específicas de publicación de datos abiertos. A partir de esto, es posible generar protocolos y lineamientos específicos para la ciencia y la educación.²⁷

Dentro de este contexto, con la creación de la CCUD desde 2013, la UNAM ha generado un marco normativo y metodológico para el manejo, publicación y distribución de los datos abiertos. El 24 de septiembre de 2015 publicó los “Lineamientos para la integración y publicación de colecciones universitarias digitales en el *Portal de Datos Abiertos UNAM, Colecciones Universitarias*”; también generó los “Términos de libre uso de datos abiertos de la UNAM”²⁸ y, finalmente, la publicación de este *Manual*.

Otras acciones que impulsan la apertura del conocimiento universitario se encuentran en el “Acuerdo por el que se establecen los Lineamientos Generales para la Política de Acceso Abierto de la Universidad Nacional Autónoma de México”²⁹ y el “Reglamento de transparencia, acceso a la información pública y protección de datos personales para la Universidad Nacional Autónoma de México.”³⁰ Sin embargo, el marco normativo para datos y acceso abierto es un área donde todavía queda mucho trabajo por realizar.

27. Se pueden mencionar otras guías para la publicación de datos abiertos, como la publicada en Chile *Datos científicos abiertos: la ciencia la hacemos entre todos*. Documento en línea: <http://datos.bcn.cl/es/obtienearchivo?id=documentos/10221.1/44377/1/datos_cientificos_abiertos_conicyt.pdf>; fecha de última consulta, 9 de octubre de 2016; así como la *Guía para la apertura de datos en Colombia*. Documento en línea: <http://programa.gobiernoonlinea.gov.co/apc-aa-files/da4567033d075590cd3050598756222c/Datos_Abiertos_Guia_v2_0.pdf>; fecha de última consulta: 9 de octubre de 2016.

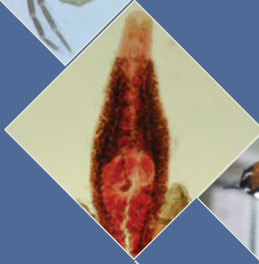
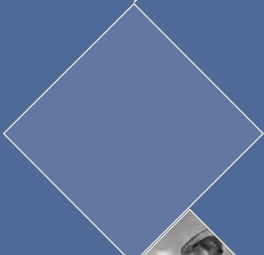
28. Ambos se pueden consultar en este *Manual* en los anexos 1 y 2, respectivamente.

29. *Gaceta UNAM*, (México, UNAM, 10 de septiembre de 2015), documento en línea: <http://www.unamenlinea.unam.mx/files/TUL_AcuerdoLineamientosGeneralesPoliticaAccesoAbierto.pdf>; fecha de última consulta: 17 de octubre de 2016.

30. *Gaceta UNAM*, (México, UNAM, 25 de agosto de 2016), documento en línea: <http://www.transparencia.unam.mx/files/documentos/reglamento_transparencia2016.pdf>; fecha de última de consulta: 7 de octubre de 2016.



CAPÍTULO 3



INTEGRACIÓN Y PUBLICACIÓN DE COLECCIONES UNIVERSITARIAS DIGITALES

El presente capítulo establece los pasos administrativos, normas técnicas y metodologías de publicación en línea a través del *Portal de Datos Abiertos UNAM, Colecciones Universitarias*. A diferencia de otras guías y manuales de integración de datos en plataformas digitales,¹ este *Manual* describe un flujo de trabajo que requiere de una estrecha colaboración con las entidades y dependencias de la UNAM ya que el proceso está enfocado en el cuidado de:

- a) La homogeneidad de los datos (estandarización).
- b) La calidad de la información por medio de metodologías de control de calidad.
- c) Los permisos de publicación.

De esta manera, se describe una etapa inicial para la identificación de las colecciones que serán publicadas en formatos abiertos y el proceso de arquitectura y gestión de datos al cual se incorporan, previo a su publicación.

3.1 Etapa inicial: seleccionar datos y tipos de colecciones

3.1.1 Identificar las colecciones

El primer paso para la integración consiste en seleccionar las colecciones que se desean publicar. En el primer capítulo ya se ha definido que una colección universitaria digital es el conjunto de datos reunidos y clasificados por su valor académico,

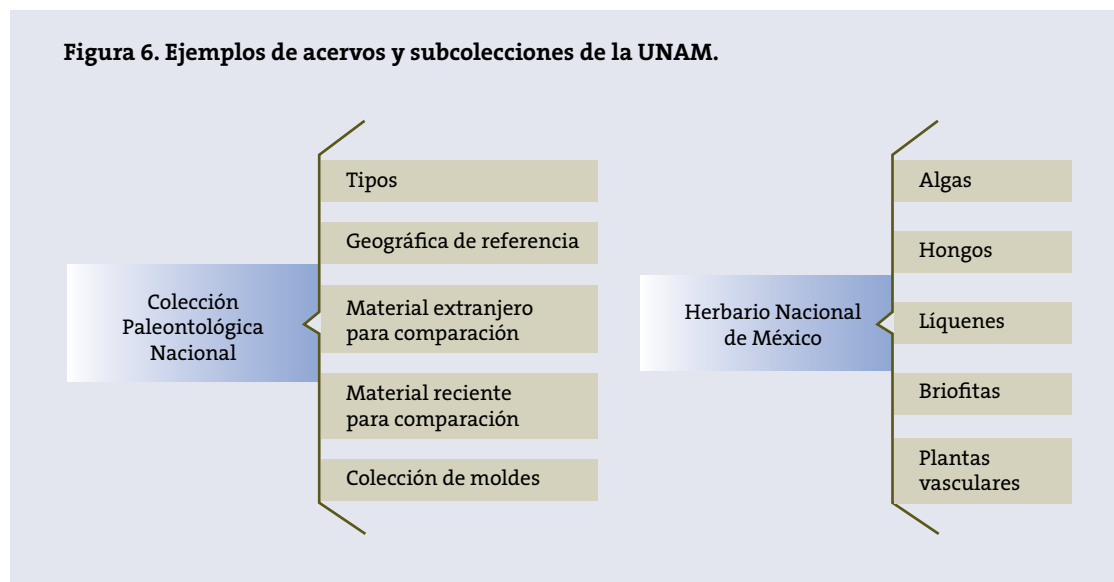
1. Como serían los casos del *Manual de implementación para el acceso y publicación de datos abiertos* del Inegi y la *Guía de implementación de datos abiertos* del gobierno federal (datos.gob.mx); ambos son procedimientos para el desarrollo de equipos internos que integran datos de manera autogestiva en atención a una serie de requisitos técnicos.

cultural, histórico y de identidad universitaria; codificados en formatos electrónicos, se pueden leer y compartir por medio de computadoras y otros dispositivos digitales.²

Para seleccionar una colección, es muy importante delimitarla con claridad. Para ello, la CCUD aplica dos criterios esenciales:

- a) Que exista una relación temática de los datos que la conforman como ejemplares de un herbario, fotografías de un archivo o datos de un proyecto de investigación.
- b) Que el conjunto cuente con un curador o responsable designado.

En algunos casos, una colección puede formar parte de un fondo o acervo más amplio. En este caso, cada una debe considerarse una colección asociada; por ejemplo, la Colección Paleontológica Nacional se compone de 5 colecciones ya que cada una guarda una relación temática y tiene un curador específico asociado. Otro ejemplo es el Herbario Nacional de México —un gran acervo botánico— que se subdivide en otras colecciones como se muestra en la figura 6:



2. Capítulo 1, apartado 1.1. “Las Colecciones Universitarias Digitales”.

Cada proveedor de datos define la estructura de sus colecciones. El siguiente paso consiste en evaluar si cuentan con las condiciones para ser publicadas como datos abiertos.

3.1.2 Identificar los Datos Abiertos

Poner a disposición las colecciones en formatos abiertos implica reducir, en gran medida, las restricciones de acceso a los usuarios, con el objetivo de optimizar el uso de la información. Para ello, el siguiente paso consiste en identificar los potenciales datos abiertos, que se definen como digitales, de carácter público y, en términos de las disposiciones aplicables por la universidad,³ no tienen naturaleza reservada o confidencial; son accesibles en línea para ser utilizados, reutilizados y redistribuidos por cualquier interesado.⁴ Antes de publicar una colección, debe hacerse un análisis cuidadoso para detectar que no exista algún impedimento para su uso público. Las principales causas son:

- a) Que esté clasificada como confidencial o reservada.
- b) Que contenga datos personales.
- c) Que afecte la integridad de personas, especies u objetos.
- d) Que afecte el desarrollo de los proyectos en curso.
- e) Que afecte acuerdos de confidencialidad.
- f) Que afecte procedimientos legales en curso.

Es conveniente identificar este tipo de contenidos para que la elección permita, desde un inicio, trabajar con la mayor certeza posible en la apertura de los datos.

3. Se refiere a un marco jurídico de normas de Acceso Abierto, Datos Abiertos, transparencia y derechos de autor.

4. Para saber más sobre los datos abiertos y las disposiciones en la UNAM, se puede consultar el capítulo 2 “Sobre los datos abiertos” de este *Manual*, o bien, el Anexo 1, LINEAMIENTOS, *op. cit.* capítulo III “Del Portal de Datos Abiertos UNAM Colecciones Universitarias”.

3.1.3 Identificar las prioridades

Para que los esfuerzos realizados generen el mayor impacto posible, se dará preferencia a las colecciones y conjuntos de datos⁵ que cumplan con la mayoría de los siguientes criterios:

- a) De carácter público o similar a los publicados por otras iniciativas nacionales e internacionales, por ejemplo: datos sobre genoma que ya forman parte de iniciativas como la organización GenBank o datos sobre biodiversidad que forman parte de proyectos como GBIF.⁶
- b) Con alta demanda de consulta por parte de entidades académicas, de gobierno o de la sociedad civil, por ejemplo: “Plantas de la Ciudad de México” o información de proyectos de investigación realizados en la UNAM.
- c) Relevantes para la comprensión y solución de temáticas nacionales para el diseño y aplicación de políticas públicas, por ejemplo: datos de contaminación o escenarios de cambio climático.
- d) Que se encuentran en formatos digitales y disponibles para publicación.
- e) Que se encuentran en formatos estructurados como tablas.
- f) Que estén atomizados: provienen de la fuente de origen con el máximo nivel de desagregación posible.
- g) Que sean integrales: datos completos que contienen la información que los describe a detalle.
- h) Que sean permanentes: una de las características más importantes de los datos abiertos es que se conservan en el tiempo y las versiones históricas relevantes se encuentran disponibles con los identificadores adecuados.
- i) Que cuentan con permisos de publicación: está implícito que las colecciones seleccionadas contarán en su momento con permisos de publicación⁷ y únicamente en casos excepcionales habrá registros que no serán datos públicos.

5. LINEAMIENTOS, *op. cit.*, capítulo III, apartado 12 (sobre la priorización de datos abiertos y conjuntos de datos).

6. Para más información, se puede consultar la página de Genbank: <<http://www.ncbi.nlm.nih.gov/genbank/>>. La página de GBIF se encuentra disponible en: <<http://www.gbif.org/>>.

7. Éstos se detallan en la sección 3.2, apartado sobre arquitectura y gestión de datos.

3.1.4 Clasificar los datos

La integración de Colecciones Universitarias Digitales se hace a través de la plataforma informática de la CCUD (Aurora),⁸ un sistema modular que es capaz de manejar cuatro tipos o categorías de datos con sus metadatos:⁹ registros de colecciones, bases de datos, capas geoespaciales y objetos digitales.

a) Registros de acervos digitales y proyectos universitarios:

Son conjuntos de campos correspondientes a una fila dentro de una tabla (fig. 7). Entre cada registro no hay una relación directa y se pueden publicar con independencia de los demás; por ejemplo, en el caso de una base de datos de un herbario, un registro corresponde a un ejemplar colectado, o bien, en una colección de obra artística, puede corresponder a una pieza (una fotografía o un objeto). La información del ejemplar de herbario y el objeto de obra artística se pueden publicar de forma independiente a los demás sin que su pierdan contexto.

Figura 7. Registro de un ejemplar en una base de datos publicado como una ficha. Los datos de este ejemplar se publican con independencia de los demás.

occurrenceID	family	scientificname	recordedby	recordnumber	eventdate	identifiedby
BUNAM:MEXU:112909	Moraceae	<i>Ficus velutina</i> Humb. & Bonpl. ex Willd.	M. Ramírez Cantú	1418	1961-09-10	Guillermo Barra Manríquez
BUNAM:MEXU:112910	Urticaceae	<i>Burchardia</i> sp.	Hubert Kruse	1418	1964-08-28	Esther Ledo
BUNAM:MEXU:112930	Malvaceae	<i>Sida acutifolia</i> Mill.	Heike Vibrans	0997	2000-11-24	E. Camona Chá
BUNAM:MEXU:112932	Onagraceae	<i>Epilobium ciliatum</i> Raf.	A. Bernalis		1993-12-03	A. Chimal
BUNAM:MEXU:112919	Malvaceae	<i>Ocotea</i> sp.	P. Fryxell	617	1977-10-18	
BUNAM:MEXU:112979	Hydrocotylaceae	<i>Pedicularis hirsuta</i> (O. Berg) Nied.	Octavio Martínez	175	1977-03	
BUNAM:MEXU:112973	Burseraceae	<i>Bursera excelsa</i> (Kunth) Engl.	Hubert Kruse	1510	1969-06-26	J. Rzedowski
BUNAM:MEXU:112954	Hydrocotylaceae	<i>Abronia maritima</i> Nutt. ex S. Watson	Ada Érika Zúñiga	9	1967-04-20	Esther Saucedo Monarou
BUNAM:MEXU:112937	Moraceae	<i>Bracteanthium</i> sp.	Bernagol	1155	1954-02-19	
BUNAM:MEXU:112957	Onagraceae	<i>Epilobium ciliatum</i> Raf.	S. Zamudio R.	7955	1990-07-18	Ela O. Guzmán Quintana
BUNAM:MEXU:112965	Ranunculaceae	<i>Clematis douglasii</i> L.	Rogberto Flores	142	1987-01-11	A. Chimal H.
BUNAM:MEXU:112937	Chenopodiaceae	<i>Amaranthus semibaccatus</i> R. Br.	Heike Vibrans	7190	2001-05-16	Heike Vibrans
BUNAM:MEXU:112917	Hydrocotylaceae	<i>Mirabilis viscosa</i> Cav.	H. Díaz Barriga	3535	1905-11-08	Patricia Hernández Ledes
BUNAM:MEXU:112916	Hydrocotylaceae	<i>Mirabilis viscosa</i> Cav.	A. Miranda y J. J.	2065	2002-04-04	Patricia Hernández Ledes
BUNAM:MEXU:112929			Guzmán & Guzmán	1117	1990-12-28	Angélica Ramírez Rúa
BUNAM:MEXU:112925		BUNAM:MEXU:112925	Mario Sousa	3371	1967-12-09	
BUNAM:MEXU:112920			W. G. D'Arcy	15682	1984-07-12	M. Nees
BUNAM:MEXU:112880	Hydrocotylaceae	<i>Pedicularis hirsuta</i> (O. Berg) Nied.	José Luis Pineda C.	454	1984-10-03	Ela O. Guzmán
BUNAM:MEXU:112938	Celastraceae	<i>Echeveria mucronata</i> Schrad.	Máximo Pérez	110	1957-07-28	H. Díaz Barriga
BUNAM:MEXU:112909	Hydrocotylaceae	<i>Mirabilis jalapa</i> L.	E. Ventura y E. López	8930	1990-10-11	J. Rzedowski
BUNAM:MEXU:112891	Caryophyllaceae	<i>Arenaria bryodes</i> Willd. ex Schrad.	John B. Paxson	535	1947-08-07	Armando Rincón G.
BUNAM:MEXU:112897	Moraceae	<i>Ficus velutina</i> Humb. & Bonpl. ex Willd.	Howard Scott Gentry	10945	1951-07-25	Guillermo Barra Manríquez
BUNAM:MEXU:112909	Garyaceae	<i>Garya ovalis</i> Benth.	Anónimo	134	1979-07-28	E. Martínez S.
BUNAM:MEXU:112938	Gramineae	<i>Digitaria castanea</i> (Benth.) Neesard	Bill Clark	51-1738	1950-08-10	G. Davidse
BUNAM:MEXU:112960	Sapotaceae	<i>Pouteria campechiana</i> (Kunth) Benth.	Jesús M. León C.		1905-06-06	Myrtle
BUNAM:MEXU:112953	Sapotaceae	<i>Chrysophyllum mexicanum</i> Brongniart	Mario Sousa	859	1961-08-13	Mario Sousa
BUNAM:MEXU:112933	Crotonaceae	<i>Epilobium ciliatum</i> Raf.	Juan H. Espinosa	2336	1962-07-16	J. Rzedowski
BUNAM:MEXU:112927	Gramineae	<i>Zea mays</i> L.	Mario Sousa	3371	1967-12-09	México (MX)

8. La información sobre la Plataforma se puede consultar en el capítulo 1, apartado 1.3.

9. LINEAMIENTOS, *op.cit.*, capítulo III, apartado 13, Anexo 1.

b) Capas geoespaciales y sus metadatos:

Cada elemento de la base de datos tiene valor en su relación con un sistema de información geográfica (SIG) por su asociación con coordenadas, polígonos o imágenes (fig. 8).

Figura 8. Geovisualizador del Portal de Datos Abiertos UNAM.



c) Metadatos de objetos digitales:

Los objetos digitales son archivos creados en formato electrónico o están convertidos a éste y se accede a ellos a través de dispositivos informáticos. Generalmente se trata de imágenes, documentos y archivos multimedia (fig. 9).

Figura 9. Tipos de objetos digitales.



d) Bases de datos y sus metadatos:

Es un conjunto de registros relacionados o pertenecientes a un mismo contexto que se encuentran dentro de un manejador de bases de datos o en una tabla (fig. 10). Estos datos deben publicarse en conjunto para que no pierdan valor ya que tienen una relación directa entre sí; por ejemplo, en una tabla con mediciones de la temperatura de un lago en un periodo determinado, cada registro corresponde a una lectura; la evolución de dicha variable se verá reflejada en el conjunto de lecturas, que es lo que le da valor a la base de datos.

Figura 10. Ejemplo de manejador de bases de datos de registros relacionados.

srid	auth_r	auth_srid	srttext	proj4text
2560	EPSG	2560	PROJCS["Pulkovo 1942 / 3-degree	+proj=tmerc +lat_0=0 +lon_0=129 +k=1 +x_0=
2561	EPSG	2561	PROJCS["Pulkovo 1942 / 3-degree	+proj=tmerc +lat_0=0 +lon_0=132 +k=1 +x_0=
2562	EPSG	2562	PROJCS["Pulkovo 1942 / 3-degree	+proj=tmerc +lat_0=0 +lon_0=135 +k=1 +x_0=
2563	EPSG	2563	PROJCS["Pulkovo 1942 / 3-degree	+proj=tmerc +lat_0=0 +lon_0=138 +k=1 +x_0=
2564	EPSG	2564	PROJCS["Pulkovo 1942 / 3-degree	+proj=tmerc +lat_0=0 +lon_0=141 +k=1 +x_0=
2565	EPSG	2565	PROJCS["Pulkovo 1942 / 3-degree	+proj=tmerc +lat_0=0 +lon_0=144 +k=1 +x_0=
2566	EPSG	2566	PROJCS["Pulkovo 1942 / 3-degree	+proj=tmerc +lat_0=0 +lon_0=147 +k=1 +x_0=
2567	EPSG	2567	PROJCS["Pulkovo 1942 / 3-degree	+proj=tmerc +lat_0=0 +lon_0=150 +k=1 +x_0=
2568	EPSG	2568	PROJCS["Pulkovo 1942 / 3-degree	+proj=tmerc +lat_0=0 +lon_0=153 +k=1 +x_0=
2569	EPSG	2569	PROJCS["Pulkovo 1942 / 3-degree	+proj=tmerc +lat_0=0 +lon_0=156 +k=1 +x_0=
2570	EPSG	2570	PROJCS["Pulkovo 1942 / 3-degree	+proj=tmerc +lat_0=0 +lon_0=159 +k=1 +x_0=
2571	EPSG	2571	PROJCS["Pulkovo 1942 / 3-degree	+proj=tmerc +lat_0=0 +lon_0=162 +k=1 +x_0=
2572	EPSG	2572	PROJCS["Pulkovo 1942 / 3-degree	+proj=tmerc +lat_0=0 +lon_0=165 +k=1 +x_0=

Los pasos anteriores permiten identificar, priorizar y clasificar los tipos de colecciones y datos. Una vez realizada esta selección, es posible continuar con el procedimiento para la integración y publicación en el *Portal*. La siguiente sección describe las directrices y pasos a seguir en colaboración con la CCUD.

3.2 Protocolo de integración y publicación de colecciones en el Portal de Datos Abiertos UNAM

La CCUD colabora con las entidades y dependencias proveedoras de datos por medio de proyectos de integración y publicación de Colecciones Universitarias Digitales. Éstos son de carácter institucional y siempre cuentan con el respaldo de su director.

El “Protocolo de integración y publicación de colecciones en el *Portal de Datos Abiertos UNAM, Colecciones Universitarias*” describe el procedimiento a seguir en los proyectos para que los datos de una colección se integren de manera estandarizada. Una parte de este proceso se realiza en estrecha colaboración con los proveedores de datos, principalmente con los curadores, responsables de los datos y del proyecto. El proceso que se describe muestra únicamente los 18 pasos de las tres etapas en las que están involucrados:¹⁰

- Definición del proyecto (pasos del 1 al 6).
- Arquitectura y gestión de datos (pasos 7 al 13).
- Publicación de colecciones (pasos 14 al 18).

Definición del proyecto

Paso 1. Reunión exploratoria

La entidad o dependencia proveedora de datos presenta las colecciones que desea integrar y publicar en el *Portal de Datos Abiertos UNAM*. La CCUD presenta, para su conocimiento y conformidad, el protocolo de integración. En esta reunión se define una fecha para entregar una muestra representativa de los datos (paso 2) para su análisis por el personal de la CCUD.

Entregable: Minuta con acuerdos de la reunión.

Responsable: Inventario de Colecciones y Comunicación (IC-CCUD).

10. El procedimiento de integración cuenta con pasos técnicos adicionales que se realizan en la CCUD. El total es de 27 pasos y éstos se muestran en el desplegable de este *Manual*. Sin embargo, el capítulo sólo incluye aquellos en los que hay una vinculación con las entidades y dependencias.

Paso 2. Entrega de la muestra de datos

La entidad o dependencia envía la muestra de la base de datos por correo electrónico u otro medio digital a la CCUD. En cuanto se recibe, ésta se envía al área de Informática y Estadística de la coordinación para su análisis.

Entregable: Muestra de datos.

Responsable: Responsable de la colección (proveedor de datos).

Paso 3. Análisis de la muestra de datos

El área de Informática y Estadística realiza un análisis de la estructura de la base de datos que permita contar con indicadores para evaluar que la integración se puede llevar a cabo. Las características principales a considerar son:

- a) Grado de avance en la digitalización de la colección.
- b) Catálogos disponibles.
- c) Grado de avance en la estandarización.
- d) Homogeneidad de la información.
- e) Objetos digitales (imágenes, documentos, audio, otros).
- f) Valoración de los requerimientos humanos y materiales, así como del tiempo necesario para su ejecución.

De igual manera, se define si la colección es un acervo universitario, un proyecto académico o publicaciones y repositorios.¹¹ Los jefes de área de la CCUD evalúan la viabilidad del proyecto para su integración. Si el resultado es positivo, se convoca a una reunión de inicio de proyecto (paso 4). En caso de identificarse aspectos que impidan cubrir alguna de las características anteriores, se debe replantear la propuesta original.

Entregable: Resultado del análisis.

Responsable: Coordinación Técnica (CT-CCUD).

11. Para más información, consultar la definición de colecciones universitarias digitales en el capítulo 1.

Paso 4. Reunión de inicio de proyecto

El inicio del proyecto implica apegarse a un plan de trabajo entre las partes (proveedor de datos y CCUD). En esta reunión se definen las fechas para llevar a cabo el plan de trabajo y la entrega de documentación:

- a) Oficio de solicitud de integración (la CCUD entrega el formato) dirigido a la Coordinación General de la CCUD, indicando:
 - Listado con los nombres de las colecciones.
 - Tipo de datos (registros de colecciones, bases de datos, objetos digitales y capas geoespaciales).
 - Responsable de la colección.
 - Entidad o dependencia que resguarda la colección y que es responsable de la base de datos.
 - Indicar el enlace para la colaboración con la CCUD (es necesario nombrar a un responsable del proyecto).
 - Firma del titular de la entidad o dependencia.
- b) Lineamientos de publicación (la CCUD entrega el formato): es un documento en el cual la entidad o dependencia establece los criterios de publicación de las colecciones en el *Portal de Datos Abiertos UNAM*, de conformidad con el capítulo IV de los Lineamientos.¹²
- c) Fecha para enviar el primer borrador de bases de colaboración al responsable del proyecto para revisión de la entidad o dependencia.
- d) Fecha para realizar el censo de colecciones (descrito en el paso 6).
- e) Fecha para realizar la transferencia de los datos y objetos digitales (la CCUD entrega requerimientos).

Adicionalmente, en la reunión se definen los requerimientos técnicos específicos necesarios para integrar la colección, el estándar que se usará o sobre el que se desarrollará uno nuevo, así como los catálogos para el control de calidad.

Entregable: Formato de solicitud de integración y de lineamientos; requerimientos técnicos.

Responsable: Coordinación Técnica (CT-CCUD) y proveedor de datos.

12. LINEAMIENTOS, *op. cit.*

Paso 5. Formalización del proyecto

Bases de colaboración firmadas por los titulares de las dependencias y entidades involucradas.

Entregable: Bases de colaboración.

Responsable: Inventario de Colecciones y Comunicación (ic-CCUD) y responsable del proyecto (proveedor de datos).

Paso 6. Registro del censo de colecciones

El objetivo del censo es contar con información que da contexto a cada uno de los registros de las colecciones que se integran al *Portal de Datos Abiertos UNAM*. De esta manera, el usuario final podrá conocer la procedencia, los responsables y las características principales de cada registro (ejemplar u objeto) consultado. El censo es un cuestionario breve que se responde en una reunión con el curador o responsable de los acervos, proyectos o repositorios. Ésta se realiza en coordinación con el responsable del proyecto.

La elaboración del censo tiene como referencia el proyecto del estándar de descripción de colecciones biológicas *Natural Collection Description* (NCD) de la organización Biodiversity Information Standards (Taxonomic Databases Working Group, TDWG) y tiene como principales objetivos:

- a) Concentrar en un documento sencillo la descripción general de cada colección.
- b) Contar con datos para generar un inventario que documenta su ubicación, estado y características principales de cada acervo.
- c) Entregar un documento a cada curador con la información obtenida, misma que puede contribuir a los procesos de gestión y manejo de las colecciones.

Las actividades para elaborar el censo de colecciones se describen en la siguiente página:

1. Reunión inicial	Es la presentación del procedimiento y de los objetivos del censo a los curadores y personal que participará. En esta etapa se definen los nombres de las colecciones y se entrega un listado a la CCUD a través de un oficio firmado por los curadores o responsables.
2. Registro	Se definen las entrevistas con los curadores o responsables de las colecciones seleccionadas. El cuestionario incluye:
a. Datos institucionales:	Dirección, acrónimo de la institución, descripción de la entidad o dependencia universitaria, nombre del titular, página web y palabras clave, así como la información de la o las sedes alternas.
b. Datos generales de la colección:	Nombre oficial, acrónimos (códigos de la colección), año de creación, departamento al que pertenece, entre otros.
c. Datos de los responsables:	Información sobre el personal asociado (curador, responsable de la colección, técnicos, investigadores, fotógrafos, asesores).
d. Datos legales y de contacto:	Leyendas y términos de uso adicionales a los “Términos de Libre Uso de Datos Abiertos de la UNAM” y datos de contacto de la colección.
e. Descripción de la colección:	Textos breves que describen el contenido mayoritario de la colección, con palabras clave, para público general y especializado.
f. Colecciones especializadas:	Información de subcolecciones, colecciones accesorias y otras relacionadas a un acervo central o colección madre. Se elabora un cuestionario para cada una.
g. Recursos:	Información de páginas web, recursos digitales, catálogos y publicaciones digitales.
3. Aprobación	Una vez concluido el registro, la CCUD revisa, edita o agrega información adicional que se encuentre en otra fuente autorizada por el curador o responsable. Los resultados se envían para su visto bueno y, una vez aprobado, se capturan en el inventario de colecciones.
4. Actualización en el catálogo de colecciones	La descripción obtenida a través del censo se mantiene actualizada en el catálogo público del <i>Portal de Datos Abiertos UNAM</i> .

Esta descripción permite construir un inventario para que la CCUD genere un catálogo de colecciones que debe contar con metadatos mínimos, algunos de los cuales se documentan desde el censo.¹³ Estos son:

- a) Nombre de la colección.
- b) Entidad académica o dependencia a la que pertenece.
- c) Código de la colección.
- d) Código de la entidad o dependencia.
- e) Nombre y cargo del responsable.
- f) Contacto de la colección.

Al concluir, los resultados del censo se entregan a la entidad o dependencia, con información que puede ser de utilidad para la planeación de estrategias de trabajo y difusión de las colecciones universitarias digitales. La duración del registro varía de acuerdo con la cantidad de acervos o proyectos que se ingresan y es una actividad paralela al avance de las siguientes fases del procedimiento.

Entregable: Censo de las colecciones.

Responsable: Inventario de Colecciones y Comunicación (IC-CCUD) y responsable del proyecto (proveedor de datos).

13. LINEAMIENTOS, *op.cit.*, capítulo III, apartado 17, Anexo 1.

Paso 7. Transferencia de datos y objetos digitales

Es la entrega de la base de datos de la colección en el formato original a la CCUD. Debe ir acompañada de:

- a) Oficio de entrega por parte de la entidad o dependencia.
- b) Lineamientos de publicación (mencionado en la reunión de trabajo, paso 4).
- c) Catálogos para el control de calidad estructural (si aplica; descritos en el paso 10).

Para la entrega de datos y objetos digitales, es necesario considerar las siguientes especificaciones:

- a) Integridad: los datos o archivos deben ser entregados en una versión completa y actualizada.
- b) Identidad: todos los registros de las bases de datos de una colección o metadatos de objetos digitales, así como mapas o bases de datos, deben contar con un identificador (numérico de preferencia) que haga único el registro dentro del conjunto de datos.
- c) Fecha de actualización: cada registro deberá contener la fecha de su última modificación.
- d) Documentación: es el diccionario de la base de datos y las relaciones entre las tablas.
- e) Formato: la recepción de las bases de datos, de los metadatos de objetos digitales y sus archivos puede realizarse de la siguiente manera:
 - Formato abierto (txt, csv).
 - Formato propietario (como Excel o Access). Se evaluará el manejo por parte de la CCUD.

- f) Servicios web estandarizados (como el protocolo de transmisión de contenidos OAI-PMH).¹⁴
- Versión de la aplicación: indicar la aplicación y la versión utilizada para generar los datos (por ejemplo, DSpace 1.0).
 - Codificación de caracteres (por ejemplo, UTF8).
 - URL del servicio web.
 - Credenciales de acceso: son los datos necesarios para acceder a los recursos como nombre de usuario y contraseña (si aplica).
 - Instrucciones mínimas de uso: protocolos que utiliza el servicio, si existe alguna limitante en la descarga, horario para realizarla, filtros necesarios para generar el conjunto de datos y vigencia de la conexión.
- g) Relaciones con objetos digitales: indicar los campos que tienen la función de definir la relación con los registros de la base de datos, o bien, entregar una tabla que muestre la relación entre ambos. Un ejemplo se muestra en la figura 11 de la siguiente página.
- h) Lineamientos de publicación: se refiere al documento solicitado en la reunión de inicio de proyecto (paso 4). Su importancia radica en que durante el proceso de control de calidad se aplicarán los criterios señalados por los proveedores de datos a las colecciones (acervos, proyectos de investigación, repositorios y publicaciones).

Entregable: Oficio de entrega, base de datos, lineamientos de publicación de la entidad o dependencia y catálogos para el control de calidad.

Responsable: Responsable de la colección (proveedor de datos).

Paso 8. Definición del estándar

La integración de la diversidad de datos, colecciones y proyectos universitarios requiere de estándares adecuados para lograr un acceso unificado de la información. En la UNAM, la CCUD desempeña la función de definir estándares para estructurar los datos abiertos y coordinar sus políticas de integración y publicación; esto incluye documentar

14. Open Archives Initiative (OAI) desarrolló un protocolo de transferencia de metadatos: el *Protocol for Metadata Harvesting* (PMH) que facilita recolectar metadatos de repositorios.

Figura 11. Ejemplo de posibles relaciones entre registros.

Tabla 2. Relaciones entre los registros del acervo y metadatos de imágenes utilizando los identificadores (occurrenceID y guid, respectivamente) y precisa la secuencia en caso necesario.

Tabla 3. Metadatos de imagen con identificador del metadato (guid). Precisa la relación con los datos del acervo utilizando el campo (IsDerivedFrom).

Tabla 1. Registros del acervo con identificador de registro en el campo occurrenceID.

occurrenceID	...
IGL:IGM:14585	...
IGL:IGM:14586	...
IGL:IGM:14587	...
IGL:IGM:14588	...
IGL:IGM:14589	...
IGL:IGM:14590	...
IGL:IGM:14591	...
IGL:IGM:14592	...
IGL:IGM:14593	...
IGL:IGM:14594	...
IGL:IGM:14595	...
IGL:IGM:14596	...

occurrenceID	secuencia	guid	...
IGL:IGM:14585	1	5689	...
IGL:IGM:14586	1	5690	...
IGL:IGM:14587	1	5692	...
IGL:IGM:14587	2	5693	...
IGL:IGM:14587	3	5694	...
IGL:IGM:14587	4	5691	...
IGL:IGM:14588	1	5695	...
IGL:IGM:14589	1	5696	...
IGL:IGM:14590	1	5697	...
IGL:IGM:14591	1	5697	...
IGL:IGM:14591	2	5698	...
IGL:IGM:14592	1	5697	...
IGL:IGM:14592	2	5699	...
IGL:IGM:14593	1	5697	...
IGL:IGM:14593	2	5700	...
IGL:IGM:14594	1	5701	...
IGL:IGM:14594	2	5702	...
IGL:IGM:14595	1	5703	...
IGL:IGM:14596	1	5704	...

guid	IsDerivedFrom	...
5689	IGL:IGM:14585	...
5690	IGL:IGM:14586	...
5691	IGL:IGM:14587	...
5692	IGL:IGM:14587	...
5693	IGL:IGM:14587	...
5694	IGL:IGM:14587	...
5695	IGL:IGM:14588	...
5696	IGL:IGM:14589	...
5697	IGL:IGM:14590	...
5698	IGL:IGM:14591	...
5699	IGL:IGM:14592	...
5700	IGL:IGM:14593	...
5701	IGL:IGM:14594	...
5702	IGL:IGM:14594	...
5703	IGL:IGM:14595	...
5704	IGL:IGM:14596	...

y aplicar las normas internacionales y adecuarlas a las características específicas de la información de las colecciones universitarias digitales.¹⁵

El estándar es una norma que regula la estructura de los datos, las palabras que se utilizan (vocabulario controlado), y los campos básicos (definidos en diccionarios). Sugiere o hace obligatorio el uso de convenciones como la escritura de las fechas, códigos de los países, el idioma, las unidades métricas, los acrónimos, etcétera. Trabajar con bases de datos estandarizadas tiene varias ventajas:

15. Acuerdo que reorganiza las funciones y estructura de la Secretaría de Desarrollo Institucional de la Universidad Nacional Autónoma de México (México: *Gaceta UNAM*, 30 de noviembre de 2015).

- a) Permiten la descarga de datos que se pueden reutilizar sin necesidad de volverse a procesar (es decir, darles un nuevo formato).
- b) Facilitan el uso de los datos al contenerlos de manera organizada y clara.
- c) Ahorran tiempo y agilizan el análisis.
- d) Regulan la captura de una base de datos y la información básica que debe contener un registro.
- e) Ofrecen datos interoperables.
- f) Facilitan la validación académica al formar parte de un consenso de especialistas.

La elección de un estándar debe cumplir con las siguientes características:

- a) Simple: evita complicaciones en el registro de los datos.
- b) Flexible: permite la captura de registros sin que sea obligatorio conocer la totalidad de los atributos del estándar.
- c) Con independencia sintáctica: cuenta con datos que aportan el contexto necesario para tener significado por sí mismo.
- d) Con interoperabilidad semántica: permite que los datos sean legibles por sistemas informáticos sin necesidad de intervención humana.
- e) Con un alto nivel de normalización formal: registra datos que no presentan variantes, lo más atomizado posible, y sin perder su significado.
- f) Que su crecimiento y actualización esté regulado por instituciones estables y formales.
- g) Que cuente con un consenso internacional: es recomendable que participe el mayor número posible de especialistas para definir y actualizar el estándar.
- h) Que cuente con una arquitectura de metadatos para la web: define etiquetas para su uso por sistemas informáticos.
- i) Que presente modularidad de metadatos en la web: define extensiones para bloques de atributos no considerados.

En el proceso de definición, es conveniente adoptar estándares existentes, de preferencia internacionales, ya que son de amplia aceptación y aumentan las posibilidades de combinar y relacionar bases de datos de orígenes diversos (fig. 12). Algunos

ejemplos de los utilizados en la CCUD son *Darwin Core* (colecciones de biodiversidad), *Plinian Core* (especies de biodiversidad) y *Dublin Core* (repositorios). En caso de que no exista uno para algún tipo de colección, éste se definirá con el apoyo del personal de la CCUD que, a través del área de Informática y Estadística, elaborará la documentación correspondiente. Este proceso se realiza considerando las características anteriores.

Entregable: Estándar documentado.

Responsable: Informática y Estadística (IE-CCUD).

Figura 12. Datos de tres fuentes en diferentes formatos integrados al estándar *Darwin Core*.

Tabla 1. Datos no estandarizados de lugar, fecha y nombre científico de colectas de plantas.

Campo	Valor
Nombre científico	Erythrina cristata-galli L.
Localización	Como ornamental en la Ciudad de Morelia
Fecha de colecta	18.VII.'80

Los datos se atomizan y validan.

Tabla 2. Datos no estandarizados de lugar, fecha y nombre científico de colectas de arácnidos.

Campo	Valor
species	Pardosa valens
locality	Delegación Iztacalco, Colonia Viaducto. UTM N: 2145000 UTM E: 485900
eventdate	13-ago-93

Se atomizan y se realizan las conversiones necesarias conservando los valores originales.

Tabla 3. Datos no estandarizados de lugar, fecha y nombre científico de colectas de mariposas.

Campo	Valor
Especie	Xylophanes eumedon
Localización	México. Morelos, Villa de Ayala. Rcho. El Polverin
Fecha	1-2.x.1972 7:30-10 pm

Los datos se atomizan y validan, asignando los valores ausentes.

Se da el formato de fecha requerido por el estándar ISO 8601 y se conserva el valor original.

Tabla 4. Datos de tres fuentes con diferentes formatos integrados utilizando los lineamientos del estándar Darwin Core.

Campo (estándar Darwin Core)	Valor estandarizado (Darwin Core)		
	Plantas	Arañas	Mariposas
genus	Erythrina	Pardosa	Xylophanes
specificEpithet	cristata-galli	valens	eumedon
scientificNameAuthorShip	L.	Barnes, 1959	(Boisduval, 1875)
country	México	México	México
stateProvince	Michoacán	Ciudad de México	Morelos
county	Morelia	Iztacalco	Villa de Ayala
verbatimLocality	Como ornamental en la Ciudad de Morelia	Delegación Iztacalco, Colonia Viaducto. UTM N: 2145000 UTM E: 485900	México. Morelos, Villa de Ayala. Rancho El Polverin
locality		Colonia Viaducto	Rancho El Polverin
verbatimLatitude		UTM N: 2145000	
verbatimLongitude		UTM E: 485900	
decimalLatitude		19.399159	
decimalLongitude		-99.13429	
verbatimEventdate	18.VII.'80	13-ago-93	1-2.x.1972 7:30-10 pm
eventDate	1980-07-18	1993-10-13	1972-10-01T19:30/22:00

Paso 9. Migración de datos al estándar

A partir del estándar definido y documentado, la CCUD diseña una *base de datos curatorial* a la cual transfieren los datos originales de la colección. Esta nueva base de datos es de uso del área de Informática y Estadística y servirá para llevar a cabo el control de calidad. Todos los pasos de este proceso se documentan, así como las modificaciones realizadas. La migración es un paso que se automatiza en la medida de lo posible.

Entregable: Base de datos estandarizada con datos migrados y documentación.

Responsable: Informática y Estadística (IE-CCUD).

Paso 10. Control de calidad estructural

El control de calidad estructural es una evaluación técnica que se aplica a la base de datos curatorial para reducir la presencia de errores y contar con datos de calidad, coherentes y actualizados. Los métodos de revisión deben considerar los formatos, idiomas y estándares, así como los diversos temas de los acervos, proyectos de investigación, repositorios y publicaciones.

Para realizar el control de calidad se utilizan catálogos, los cuales son listados de valores que siguen convenciones válidas de uso y nomenclatura para cada colección. A cada clase de dato le corresponderán uno o varios catálogos específicos para establecer criterios homogéneos de formato tales como fechas, nombres de personas, nombres geográficos, o bien, determinar criterios particulares a los temas y tipos de colección como nombres científicos (biodiversidad), género fotográfico (obra artística), tipo de publicación (objetos digitales) y disciplina (proyectos universitarios), entre otros.

Para definir los catálogos que se requieren, la CCUD solicita las referencias más utilizadas en la colección y, adicionalmente, realiza una investigación para definir los que servirán para validar la información. Se dará prioridad a las fuentes de las colecciones aunque se incluyen otros criterios de selección que tienen como fin agilizar y automatizar el proceso lógico de limpieza de datos. La selección de catálogos contempla:

- a) Que estén digitalizados.
- b) Que se puedan descargar.
- c) Que se consulten a través de servicios web.

- d) Que sean de uso extendido (cantidad de gente que lo ocupa).
- e) Que se puedan actualizar.

Una vez que han sido seleccionados, se inicia el proceso de validación de los datos, que se realiza a un nivel atomizado y en su relación con otros para dar homogeneidad y estandarizarlos. A partir de un análisis de comparación lógica, se busca la equivalencia con los catálogos, así como la consistencia entre los datos de un mismo registro. Las modificaciones se registran en bitácoras, es decir, tablas de control de calidad y monitoreo. Durante el proceso se califica:

- a) Si el dato existe.
- b) Si es consistente de origen: existe en catálogos y es coherente en su relación con otros campos.
- c) Si está desactualizado o no tiene lógica por sí mismo.
- d) Si está desactualizado o no tiene lógica en el conjunto.
- e) Si se aplica una modificación que no afecta su significado (acentos, mayúsculas, traducción de un nombre, otros).

El área de Informática y Estadística de la CCUD es la responsable de definir y aplicar los métodos que permitan estimar la calidad de un dato, así como proponer modificaciones para homogeneizarlos sin perder información. De esta manera, el control de calidad permite:

- a) Detectar datos anómalos.
- b) Reducir la presencia de errores.
- c) Proveer datos homogéneos.
- d) Ubicar datos duplicados, datos faltantes e incompletos, errores ortográficos, datos ambiguos, datos desactualizados y datos obsoletos.
- e) Contar con datos de calidad.

Derivado de este análisis, los datos se clasifican en tres tipos:

- a) **Consistentes** con los catálogos: continúan el proceso de integración y publicación.
- b) **Modificados** en el proceso: son consistentes con catálogos, pero requieren de una validación por parte del curador o responsable de los datos, previo a su publicación. Se corrigen y documentan errores como asignación, omisión, ortográficos y sistemáticos. Una vez validados por el responsable de los datos o curador, los registros continúan el proceso de publicación.
- c) **Inconsistentes** con los catálogos: se reportan al curador o responsable de la colección para su revisión. Estos registros se publicarán hasta que hayan sido revisados y cumplan con los pasos del protocolo a partir de la transferencia de datos (paso 7) como una nueva entrega.

Finalizado el control de calidad, el siguiente paso consiste en filtrar los datos de acuerdo con los lineamientos de publicación establecidos por las entidades y dependencias.

Entregable: Excel con diagnóstico de control de calidad.

Responsable: Informática y Estadística (IE-CCUD).

Paso 11. Filtrado de datos

Se detectan y etiquetan los datos que no cumplieron con los criterios definidos en los lineamientos de la entidad o dependencia (descritos en el paso 4 “Reunión de inicio de proyecto” y entregados en el paso 7 “Transferencia de datos”). Los resultados filtrados y clasificados se presentan a los responsables de la colección con un diagnóstico que describe el análisis realizado.

Entregable: No aplica.

Responsable: Informática y Estadística (IE-CCUD).

Paso 12. Presentación del diagnóstico de control de calidad estructural

Es una reunión donde la CCUD presenta al curador y al responsable del proyecto la documentación del proceso de gestión de datos que incluye:

- a) Análisis de los estándares.
- b) Catálogos utilizados en el control de calidad estructural.
- c) Mapeo documentado: describe las equivalencias entre la base de datos original y el estándar, así como las adecuaciones realizadas.
- d) Análisis del control de calidad estructural con la presentación de:
 - mapas
 - estadísticas
 - listados
 - catálogos
 - clasificación de los datos
- e) Tablas de validación: son tablas que registran únicamente los datos clasificados como modificados e inconsistentes (fig. 13), para facilitar la validación por parte del curador o responsable.¹⁶

Figura 13. Ejemplo de una tabla de validación de registros modificados.

Nombre del campo	Valor original	Valor estandarizado	Respuesta del curador	Valor aceptado	Observación
eventDate	18.VII.'80	1980-07-18	aceptada	1980-07-18	
eventDate	1-2.x.1972 7:30-10 pm	1972-10-01T19:30/22:00	aceptada	1972-10-01T19:30/22:00	
eventDate	18.XII.'13	2013-12-18	revisar		Es necesario verificar si no se trata de 1913
verbatimLatitud	UTM N: 2145000	19.399159	aceptada	19.399159	
verbatimLongitud	UTM E: 485900	-99.13429	aceptada	-99.13429	
genus	Erythrina	Erythrina	aceptada	Erythrina	
specificEpithet	cristata-galli	cristata galli	rechazada	cristata-galli	
locality	Rancho El Polverin	Rancho El Polvorin	otro	Rancho El Polvorin	

16. Esta clasificación se describe en el paso 10 "Control de calidad estructural".

Como conclusión de la presentación del diagnóstico:

- a) El curador o responsable de la colección aprueba los registros consistentes para que continúen su proceso de publicación.
- b) La CCUD entrega las tablas de validación que incluyen únicamente los datos clasificados como “modificados” e “inconsistentes”.
- c) Se define una fecha para la entrega de la validación de registros modificados.
- d) Los registros inconsistentes se devuelven al curador o responsable para su revisión. Una vez revisados se reintegran al paso 7 (Transferencia de datos).

Entregable: Excel con el diagnóstico y las tablas de validación.

Responsable: Informática y Estadística (IE-CCUD).

Paso 13. Entrega de la validación de datos modificados

Durante esta etapa, el curador o responsable de la colección revisa las modificaciones realizadas a los registros. Para ello, en la presentación del diagnóstico de control de calidad estructural se entregaron las tablas de validación sobre las cuales se debe aceptar, rechazar o comentar la propuesta señalada. Una vez concluida, se entregan a la CCUD.

Entregable: Excel con registros modificados aprobados.

Responsable: Curador y responsable del proyecto (proveedor de datos).

Publicación de colecciones

Paso 14. Diseño de la ficha

Cada registro de la colección se publica en una ficha digital independiente y con una dirección única, para que pueda ser localizada por motores de búsqueda (como *Google*) y citada por los usuarios. Esta dirección le atribuye identidad digital a cada objeto o ejemplar, un elemento esencial de las normas bajo las cuales opera el *Portal de Datos Abiertos UNAM*. La CCUD diseña la ficha que se publicará en línea y, en caso necesario, se lleva a cabo una reunión con el curador o responsable de la colección.

La CCUD asigna un nombre de recurso uniforme (Uniform Resource Name, URN por sus siglas en inglés) a cada registro dentro de una colección. Este número, denominado *triple*, es necesario para integrar bases de datos independientes y se compone de la siguiente información:

- a) Siglas o acrónimo de la entidad o dependencia: se obtiene del listado oficial de las iniciales de las entidades o dependencias universitarias, publicado dentro de la normatividad de la Secretaría Administrativa de la Dirección General de Personal. En algunos casos, los códigos tienen una variación en su uso y, en ese caso, la CCUD evaluará si conserva las iniciales oficiales o aquéllas que por su uso ayudan a una mejor identificación por parte de la comunidad, con el visto bueno de la entidad o dependencia universitaria.
- b) Siglas o acrónimo de la colección: puede estar definido por una institución oficial como Semarnat (Secretaría de Medio Ambiente y Recursos Naturales), el Archivo General de la Nación, o bien, por la entidad o dependencia. En caso de no contar con uno, el código se define con la CCUD.
- c) Número de catálogo del registro: es el código con el que se identifica a cada objeto o ejemplar dentro de la colección; este número es único.

Por ejemplo, un ejemplar biológico (LE32770) dentro de la Colección Nacional de Insectos (CNIN) del Instituto de Biología (IBUNAM) se mostraría de la siguiente manera:

IBUNAM:CNIN:LE32770

El *triple* también se incluye como parte de la dirección de la ficha de cada registro que, de acuerdo con el ejemplo anterior, sería:

www.datosabiertos.unam.mx/IBUNAM:CNIN:LE32770

De esta manera, se etiquetan los registros para que sean únicos y cada uno equivale a una publicación en línea, recuperada por diferentes buscadores.

Entregable: Campos y descripción de campos en la ficha.

Responsable: Coordinación Técnica (CT-CCUD).

Paso 15. Versión final de la base de datos a publicar

El área de Informática y Estadística genera la base de datos definitiva para su publicación, en el estándar definido, con los registros filtrados y las modificaciones validadas. Ésta se entregará en un disco compacto (CD) y un oficio en el que se registra el nombre del archivo con una cadena alfanumérica o *checksum* (MD5),¹⁷ con la que se resguarda el original. Junto con la base de datos, se anexa un “formato de permisos de publicación” (paso 16).

Entregable: Disco con la base de datos a publicarse, formato para el oficio de permisos de publicación (la CCUD entrega).

Responsable: Informática y Estadística (IE-CCUD).

Paso 16. Permisos de publicación

El director otorga el permiso para que la CCUD publique las colecciones de la entidad o dependencia a través de un oficio, basado en el “formato de permisos” entregado junto con la base de datos final (paso 15). Deberá entregarse firmado, en hoja membretada y dirigido a la Coordinación General de la CCUD.

Entregable: Oficio de permisos de publicación con la firma del director.

Responsable: Inventario de Colecciones y Comunicación (IC-CCUD).

17. El *checksum* (MD5) o “suma de verificación” es un respaldo para garantizar a las entidades, dependencias y a la CCUD, la identidad de la base de datos entregada. Es un algoritmo que calcula un valor único a cada archivo para detectar si el documento sufrió algún cambio. El manual del algoritmo md5sum se encuentra disponible en: <http://www.gnu.org/software/coreutils/manual/html_node/md5sum-invocation.html#index-md5sum>.

Paso 17. Publicación de la colección

En esta etapa se llevan a cabo diversos procesos técnicos del área de Desarrollo Tecnológico.¹⁸ Una vez concluidos, la CCUD hace pública la colección en el *Portal de Datos Abiertos UNAM, Colecciones Universitarias* (datosabiertos.unam.mx).

Entregable: Colección disponible en línea.

Responsable: Desarrollo Tecnológico (DT-CCUD).

Paso 18. Entrega de proyecto

Se notificará al responsable de la colección que ésta se encuentra disponible para su consulta en línea. Al término del proceso, la CCUD hará entrega de constancias de publicación a los curadores o responsables de los datos.

Entregable: Constancias de publicación.

Responsable: Coordinación General (CG-CCUD).

18. Los pasos técnicos se pueden consultar en el desplegable del *Manual*.

3.3 Actualizaciones

Como se menciona en el capítulo 2, una de las características principales de los datos abiertos es que sean oportunos, es decir, actualizados periódicamente. Las colecciones son dinámicas, sus contenidos se actualizan constantemente ya que se completan, modifican o corrigen durante el trabajo de curación. Un buen ejemplo de este dinamismo es el caso de las colecciones biológicas, donde los ejemplares cuentan con una determinación taxonómica que los especialistas modifican conforme se actualizan los criterios de clasificación científica de los organismos.

Es muy importante que los cambios de la colección se vean reflejados en el *Portal de Datos Abiertos UNAM*. Hay cuatro maneras en que las actualizaciones se pueden integrar:

- a) Nueva transferencia (paso 7 del Protocolo): cuando la colección cuenta con una acumulación de cambios significativa (a criterio del curador o responsable) se realiza una nueva entrega.
- b) Catálogos y estándares: ambos están en constante actualización por las organizaciones que los emiten. Cada vez que se modifican se revisa la arquitectura y gestión de los datos.
- c) Lineamientos: si una entidad o dependencia modifica sus criterios de publicación, éstos se deben actualizar y filtrar en la base de datos de la CCUD para una nueva publicación en el *Portal* en el menor tiempo posible.
- d) Usuarios: detectan y reportan errores en las colecciones a través del *Portal*. En este caso, la CCUD turna las observaciones al responsable de la colección.¹⁹

19. Para cualquier información adicional, escribir a la CCUD: contacto@ccud.unam.mx



REFERENCIAS

Protocolos y procedimientos de la ccud

El *Manual de Datos Abiertos de Colecciones Universitarias Digitales* toma como principal referencia los siguientes protocolos y procedimientos internos de la Coordinación de Colecciones Universitarias Digitales:

PR.CCUD/CT/002/2014/f Ver_Procedimiento de integración y publicación de colecciones en el *Portal de Datos Abiertos UNAM, Colecciones Universitarias* (México: Universidad Nacional Autónoma de México, 2016).

PO.CCUD/IC/002/2015/B Ver_Procedimiento de operación para elaborar el censo de colecciones universitarias digitales (México: Universidad Nacional Autónoma de México, 2016).

PO.CCUD/IE/001/2016/A Ver_Procedimiento de operación para la definición de estándares de datos de colecciones que ingresan a la coordinación de colecciones universitarias digitales (México: Universidad Nacional Autónoma de México, 2016).

PO.CCUD/IE/002/2016/A Ver_Procedimiento de operación para el control de calidad de los datos que ingresan al área de Informática y Estadística (México: Universidad Nacional Autónoma de México, 2016).

PO.CCUD/IE/003/2016/A Ver_Procedimiento de operación para la presentación del diagnóstico de control de calidad de las bases de datos que ingresan al área de Informática y Estadística (México: Universidad Nacional Autónoma de México, 2016).

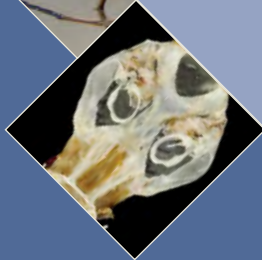
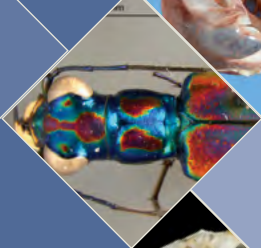
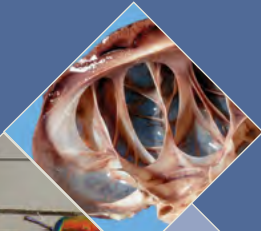
Bibliografía

- Acuerdo por el que se aprueba la Norma Técnica para el acceso y publicación de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional. *Diario Oficial de la Federación*. México: Secretaría de Gobernación, 4 de diciembre de 2014. Documento en línea: <http://www.dof.gob.mx/nota_detalle.php?codigo=5374183&fecha=04/12/2014>; fecha de última consulta: 2 de febrero de 2016.
- Acuerdo por el que se establecen los Lineamientos Generales para la Política de Acceso Abierto de la Universidad Nacional Autónoma de México. México: *Gaceta UNAM*, 10 de septiembre de 2015. Disponible en línea: <http://www.unamenlinea.unam.mx/files/TUL_AcuerdoLineamientosGeneralesPoliticaAccesoAbierto.pdf>; fecha de última consulta: 16 octubre de 2016.
- Acuerdo que Reorganiza las Funciones y Estructura de la Secretaría de Desarrollo Institucional de la Universidad Nacional Autónoma de México. México: *Gaceta UNAM*, 30 de noviembre de 2015.
- Banco Interamericano de Desarrollo, *Numbers for development*. <<https://data.iadb.org/>>; fecha de última consulta: 3 de octubre de 2016.
- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003). Disponible en: <<https://openaccess.mpg.de/Berlin-Declaration>>; fecha de última consulta: 7 de octubre de 2016.
- Coordinación de la Investigación Científica. “Sistema de informática para la biodiversidad y el ambiente (SIBA-UNAM)”, Universidad Nacional Autónoma de México. Disponible en línea: <http://www.cic-ctic.unam.mx/cic/mas_cic/megaproyectos/impulsa_5.cfm>; fecha de última consulta: 7 de octubre de 2016.
- Declaración de Bethesda sobre publicación de acceso abierto (2003). Disponible en: <http://www.senado.gob.mx/comisiones/ciencia_tecnologia/docs/accesoinfo_b3-3.pdf>; fecha de última consulta: 7 de octubre de 2016.
- Decreto por el que se establece la regulación en Materia de Datos Abiertos. *Diario Oficial de la Federación*. México: Secretaría de Gobernación, 20 de febrero de 2015. Documento en línea: <http://dof.gob.mx/nota_detalle.php?codigo=5382838&fecha=20/02/2015>; fecha de última consulta: 2 de febrero de 2016.
- Decreto por el que se expide la Ley General de Transparencia y Acceso a la Información Pública. *Diario Oficial de la Federación*. México: Secretaría de Gobernación, 4 de mayo de 2015. Documento en línea: <http://www.dof.gob.mx/nota_detalle.php?codigo=5391143&fecha=04/05/2015>; fecha de última consulta: 2 de febrero de 2016.

- Decreto por el que se reforman y adicionan diversas disposiciones de los artículos 6º, 7º, 27, 28, 73, 78, 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos, en materia de Telecomunicaciones. *Diario Oficial de la Federación*. México: Secretaría de Gobernación, 11 de junio de 2013. Documento en línea: <http://www.dof.gob.mx/nota_detalle.php?codigo=5301941&fecha=11/06/2013>; fecha de última consulta: 2 de febrero de 2016.
- Decreto por el que se reforman y adicionan diversas disposiciones de la Ley de Ciencia y Tecnología, de la Ley General de Educación y de la Ley Orgánica del Consejo Nacional de Ciencia y Tecnología. *Diario Oficial de la Federación*. México: Secretaría de Gobernación, 20 de mayo de 2014. Documento en línea: <https://goo.gl/nLeCebhttp://www.dof.gob.mx/nota_detalle.php?codigo=5345503&fecha=20/05/2014>; fecha de última consulta: 2 de febrero de 2016.
- GenBank, “What Is GenBank?”, National Center for Biotechnology Information. Documento en línea: <<https://www.ncbi.nlm.nih.gov/genbank/>>; fecha de última consulta: 7 de octubre de 2016.
- Global Biodiversity Information Facility. <<http://www.gbif.org/>>; fecha de última consulta: 7 de octubre de 2016.
- GNU, “md5sum: Print or check MD5 digests”, GNU Coreutils. Documento en línea: <http://www.gnu.org/software/coreutils/manual/html_node/md5sum-invocation.html#index-md5sum>; fecha de última consulta: 7 de octubre de 2016.
- Guía de Implementación de la Política de Datos Abiertos. *Diario Oficial de la Federación*. México: Secretaría de Gobernación, 18 de junio de 2015. Documento en línea: <http://www.dof.gob.mx/nota_detalle.php?codigo=5397117&fecha=18/06/2015>; fecha de última consulta: 2 de febrero de 2016.
- Iniciativa de Budapest para el acceso abierto (2002). Documento en línea: <<http://www.budapestopenaccessinitiative.org/translations/spanish-translation>>; fecha de última consulta: 7 de octubre de 2016.
- International Aid Transparency Initiative. <<http://www.aidtransparency.net/>>; fecha de última consulta: 7 de octubre de 2016.
- International Council for Science. *Open Data in a Big Data World*, Science International. Disponible en: <<http://www.icsu.org/science-international/accord>>; fecha de última consulta: 2 de febrero de 2016.
- Iniciativa Lationamericana por los Datos Abiertos (ILDA). <<https://idatosabiertos.org/>>; fecha de última consulta: 21 de octubre de 2016.

- International Open Data Charter. <<http://opendatacharter.net>>; fecha de última consulta: 9 de mayo de 2016.
- LINEAMIENTOS. Ver_ "Lineamientos para la integración y publicación de las colecciones universitarias digitales en el *Portal de Datos Abiertos UNAM, Colecciones Universitarias*". México: *Gaceta UNAM*, 24 de septiembre de 2015.
- Ministerio de Tecnologías de la Información y las Comunicaciones. *Guía para la apertura de datos en Colombia, Gobierno en línea*. Colombia: Ministerio de Tecnologías de la Información y las Comunicaciones. Documento en línea: <http://programa.gobiernoenlinea.gov.co/apc-aa-files/da4567033d075590cd3050598756222c/Datos_Abiertos_Guia_v2_0.pdf>; fecha de última consulta: 9 de octubre de 2016.
- Natural Collections Descriptions Interest Group. *Natural Collections Descriptions (NCD): A data standard for exchanging data describing natural history collections*, v0.9 2008, Biodiversity Information Standards (TDWG).
Documento en línea:
<http://www.biodiversidad.gob.mx/especies/colecciones/pdf/NCD_090.pdf>, fecha de última consulta: 9 de octubre de 2016.
- Open Data Institute. <<http://theodi.org>>; fecha de última consulta: 7 de octubre de 2016.
- Open Knowledge International. <<http://okfn.org>>; fecha de última consulta: 7 de octubre de 2016.
- _____. *Global Open Data Index*. Documento en línea: <<http://global.census.okfn.org/>>; fecha de última consulta: 1 de febrero de 2016.
- _____. "¿Qué son los datos abiertos?", en *Open Data Handbook*.
Documento en línea: <<http://opendatahandbook.org/guide/es/what-is-open-data/>>; fecha de última consulta: 7 de octubre de 2016.
- Organización de las Naciones Unidas. *UNdata, a world of information*. Disponible en línea: <<http://data.un.org/>>; fecha de última consulta: 7 de octubre de 2016.
- _____. United Nations Development Program. Disponible en: <<http://open.undp.org/#2016>>; fecha de última consulta: 7 de octubre de 2016.
- _____. Independent Expert Advisory Group Secretariat. *A world that counts*, Documento en línea: <<http://undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>>; fecha de última consulta: 7 de octubre de 2016.
- Organización para la Cooperación y el Desarrollo Económico. *Principles and Guidelines for Access to Research Data from Public Funding*. Documento en línea: <<http://www.oecd.org/sti/sci-tech/38500813.pdf>>; última consulta: 2 de febrero de 2016.

- Política Nacional de Datos Abiertos. México. Disponible en línea:
<<http://mxabierto.github.io/iniciativa-datos-abiertos/documentos/politica/pnda/>>;
fecha de última consulta: 2 de febrero de 2016.
- Programa de Información Científica. *Datos científicos abiertos: la ciencia la hacemos entre todos*. Chile: Comisión Nacional de Investigación Científica y Tecnológica, 2014. Documento en línea:<http://datos.bcn.cl/es/obtienearchivo?id=documentos/10221.1/44377/1/datos_cientificos_abiertos_conicyt.pdf>; fecha de última consulta, 9 de octubre de 2016.
- Reglamento de transparencia, acceso a la información pública y protección de datos personales para la Universidad Nacional Autónoma de México. *Gaceta UNAM*. México: Universidad Nacional Autónoma de México, 25 de agosto de 2016. Documento en línea:
<http://www.transparencia.unam.mx/files/documentos/reglamento_transparencia2016.pdf>; fecha de última consulta: 7 de octubre de 2016.
- Términos de Libre Uso de Datos Abiertos en la UNAM. 2016. México: Universidad Nacional Autónoma de México. Disponible en:
<<https://datosabiertos.unam.mx/informacion/terminosdeuso.html>>; fecha de última consulta: 7 de octubre de 2016.
- Toda la UNAM en línea. Universidad Nacional Autónoma de México. Disponible en:
<<http://www.unamenlinea.unam.mx/>>; fecha de última consulta: 7 de octubre de 2016.
- Uhlir, Paul F. y Peter Schröder, “Open data for global science” en *Data Science Journal*, Volumen 6, Open Data Issue, 17 de junio de 2007. Documento en línea: <<http://www.spatial.maine.edu/icfs/Uhlir-SchroederPaper.pdf>>; fecha de última consulta: 2 de febrero de 2016.
- UNESCO. *UNESCO Transparency Portal*. Disponible en: <<http://opendata.unesco.org/>>;
fecha de última consulta: 7 de octubre de 2016.



GLOSARIO

Acceso Abierto. Vía para acceder a investigaciones, materiales educativos, académicos, científicos, tecnológicos y de innovación, por medio de una plataforma digital que no requiere suscripción, registro o pago, sin perjuicio de las disposiciones en materia de patentes, protección de la propiedad intelectual o industrial, seguridad nacional y derechos de autor, entre otras, así como de aquella información que, por razón de su naturaleza y decisión del autor, sea confidencial o reservada. Para más información véase el capítulo 2, sección “Conceptos generales”.

Acervo digital. Conjunto de bienes culturales o patrimoniales puestos a disposición del público como colecciones digitalizadas.

Base de datos. Conjunto de datos relacionados o pertenecientes a un mismo contexto, almacenados sistemáticamente para su posterior uso. Las bases de datos permiten almacenar grandes cantidades de datos para su organización y administración, así como para construir relaciones entre ellos, ya que por sí solos y de manera aislada, no reflejan ningún tipo de información.

Capas geoespaciales. Conjunto de datos con referencia geoespacial que tienen valor en un contexto geográfico y que deben ser descritas por metadatos.

Catálogo de colecciones. Listado o inventario de las colecciones universitarias digitales proveedoras de Datos Abiertos así como sus metadatos, que se encuentra registrado en la CCUD.

Censo de colecciones. Cuestionario para registrar información general de las colecciones (acervos, proyectos de investigación, repositorios o publicaciones) de las entidades y dependencias universitarias. Para más información véase el capítulo 3, sección “Protocolo de integración y publicación de colecciones universitarias digitales”.

Ciencia Abierta. Movimiento que tiene como objetivo que las investigaciones científicas (sus resultados, datos y procedimientos) se abran a todos los ciudadanos. Conocido por su término en inglés *Open Science*. Para más información véase el capítulo 2, sección “Conceptos generales”.

Colección. Conjunto de datos primarios de la misma clase, reunidos y clasificados por su valor académico, cultural, histórico y de identidad universitaria.

Colección Digital. Conjunto de datos primarios de los recursos que conforman las colecciones codificadas en formatos de medios electrónicos que se pueden leer y compartir por medio de computadoras y otros dispositivos electrónicos.

Colección Universitaria. Conjunto ordenado de bienes tangibles e intangibles, reunidos y clasificados por su valor académico, cultural, histórico y de identidad universitaria, propiedad de la UNAM.

Colección Universitaria Digital. Conjunto de datos primarios de bienes tangibles e intangibles, reunidos y clasificados por su valor académico, cultural, histórico y de identidad universitaria, propiedad de la UNAM, codificados en formatos de medios electrónicos que se pueden leer y compartir por medio de computadoras y otros dispositivos electrónicos. Para más información, véase el capítulo 1, sección “Las Colecciones Universitarias Digitales”.

Conjunto de datos. Serie de datos estructurados, vinculados entre sí y agrupados dentro de una misma unidad temática y física, de forma que pueden ser procesados apropiadamente para obtener información.

Constancia de publicación. Documento emitido por la CCUD a nombre del responsable de la colección, donde se hace constar que datos de la misma han sido integrados a la Plataforma Informática de la CCUD.

Datos Abiertos. Datos digitales de carácter público que administra la Coordinación de Colecciones Universitarias Digitales, y que en términos de las disposiciones aplicables

no tienen naturaleza de reservada o confidencial, siendo accesibles en línea, utilizados, reutilizados y redistribuidos por cualquier interesado. Para más información véase el capítulo 2, sección “Conceptos generales”.

Datos primarios. Datos recabados directamente de los hechos que se investigan. Su interpretación o integración deriva en datos secundarios.

Dependencias administrativas. Todas aquéllas que realizan actividades administrativas en apoyo a las funciones sustantivas de la universidad.

Entidades académicas. Todas aquéllas que realizan actividades de docencia, investigación, difusión y extensión como son los institutos, centros, facultades y escuelas y los centros de extensión universitaria.

Estándar. Modelo que sirve como punto de referencia para los campos que deben integrarse en una tabla o base de datos, así como el tipo de datos que deben capturarse en colecciones de la misma naturaleza. De esta manera pueden integrarse y ser consultadas como si se tratara de una sola.

Formato abierto. Conjunto de características técnicas y de presentación que corresponden a la estructura lógica usada para almacenar datos en un acervo digital, cuyas especificaciones técnicas están disponibles públicamente, que no suponen una dificultad de acceso y que su aplicación y reproducción no están condicionadas a contraprestación alguna.

Interoperabilidad. Del inglés *interoperability*, es un término que se utiliza para referirse a la capacidad que tienen dos o más sistemas para intercambiar, transferir y utilizar información, datos o documentos por medios electrónicos, de manera uniforme y eficiente.

Investigación transdisciplinaria. Estudio realizado por facultades, escuelas, institutos, centros o programas de diversas disciplinas que, de forma articulada e integrada,

desarrollan y fortalecen esfuerzos institucionales para el estudio y solución de algún tema, construyendo propuestas a partir de diferentes puntos de interés, y que trasciende las fronteras entre diversas disciplinas.

Metadatos. Datos estructurados y actualizados que describen el contexto y las características de contenido, captura, procesamiento, calidad, condición, acceso y distribución de un conjunto de datos, que sirven para facilitar su búsqueda, identificación y uso.

Migración de datos. Transferencia de los datos de una base de origen a otra de destino que tiene un diseño y estructura distinta. Para más información, véase el capítulo 3, sección “Protocolo de integración y publicación de colecciones universitarias digitales”.

Objetos digitales. Archivos creados en formato electrónico o la digitalización de un archivo analógico, que se accede a ellos mediante dispositivos digitales.

Plataforma informática de la CCUD. Sistema interoperativo desarrollado por la CCUD, conformada por bases de datos estandarizadas y servicios web que integran datos de colecciones, capas geoespaciales y objetos digitales. Para más información, véase el capítulo 1, sección “Plataforma Informática Aurora”.

Portal de Datos Abiertos UNAM, Colecciones Universitarias. Conjunto de páginas y puntos de acceso que permiten consultar los datos públicos de las colecciones digitales de la Universidad Nacional Autónoma de México puestas a disposición de la sociedad a través de la página electrónica (datosabiertos.unam.mx). Para más información, véase el capítulo 1, sección “El Portal de Datos Abiertos UNAM, Colecciones Universitarias.”

Propiedad de la UNAM. Objetos que conforman las colecciones producidas o custodiadas por la universidad, así como los documentos que guardan datos e información que los describe, tales como bases de datos y tablas, entre otros, que pertenecen a la UNAM.

Proveedor de datos. Es la entidad o dependencia que custodia la colección y es propietaria de sus versiones digitales (bases de datos, objetos digitales). En el proceso de integración, la CCUD debe tener interacción cercana con:

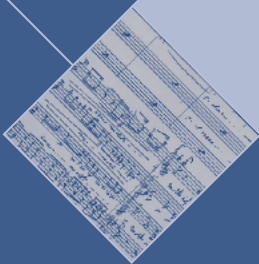
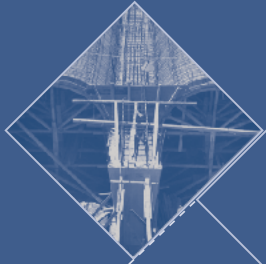
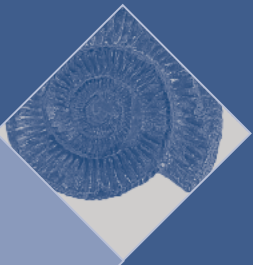
- a) Curador: persona designada como responsable de una colección. Este es el académico adscrito a la entidad o dependencia que resguarda el acervo.
- b) Responsable de proyecto: es la persona autorizada por el titular de la entidad o dependencia como intermediario entre ésta y la CCUD durante el proceso de integración y publicación de colecciones. Es el mediador para la entrega de documentos y bases de datos; también colabora en la agenda de reuniones de trabajo.
- c) Responsable de los datos: es la persona encargada de los datos y objetos digitales de una colección. Puede ser una persona diferente al curador o responsable del proyecto y, en ese caso, también debe contar con el respaldo institucional.

Registros de colecciones. Conjunto de datos correspondientes al registro de una base de datos que tienen valor por sí mismos.

Repositorio. Plataforma digital que, siguiendo los estándares internacionales, almacena, mantiene y preserva la información educativa, académica, científica, tecnológica y de innovación, la cual se deriva de las investigaciones, productos educativos y académicos.

Servicios web. Tecnología de transmisión e intercambio de información en internet por diferentes dispositivos electrónicos a partir de protocolos y estándares. También se consideran *software* de código abierto con la capacidad de acceder a múltiples sistemas operativos, permitiendo la interoperabilidad máquina a máquina en una red.

Términos de libre uso de datos abiertos en la UNAM. Autorización que otorga la universidad, como titular de los derechos patrimoniales de una(s) base(s) de datos, a un usuario, indicando lo que puede o no hacer con la misma al ser publicado como datos abiertos.



ANEXOS

**LINEAMIENTOS PARA LA INTEGRACIÓN Y PUBLICACIÓN DE LAS COLECCIONES UNIVERSITARIAS
DIGITALES EN EL *PORTAL DE DATOS ABIERTOS UNAM*
*COLECCIONES UNIVERSITARIAS***

**CAPÍTULO I
DISPOSICIONES GENERALES**

1. Los presentes Lineamientos tienen por objeto fijar las normas a las que se deben sujetar la operación, funcionamiento y control de las colecciones digitales del *Portal de Datos Abiertos UNAM Colecciones Universitarias*, así como establecer las disposiciones para que el conjunto de datos generados por las entidades académicas y dependencias universitarias, administrados por la Coordinación de Colecciones Universitarias Digitales, se pongan a disposición como Datos Abiertos, con el propósito de facilitar su acceso, uso, consulta, reutilización y redistribución.

2. Las disposiciones contenidas en los presentes Lineamientos son de observancia general y obligatoria para todas las entidades académicas y dependencias universitarias que generen, produzcan o administren información contenida en los acervos digitales susceptible de ser publicada en el *Portal de Datos Abiertos UNAM Colecciones Universitarias*.

3. Para efectos de los presentes Lineamientos, se entenderá por:

I. Acceso abierto. Vía para acceder a investigaciones, materiales educativos, académicos, científicos, tecnológicos y de innovación, por medio de una plataforma digital que no requiere suscripción, registro o pago, sin perjuicio de las

disposiciones en materia de patentes, protección de la propiedad intelectual o industrial, seguridad nacional y derechos de autor, entre otras, así como de aquella información que, por razón de su naturaleza y decisión del autor, sea confidencial o reservada;

II. Acervo digital. Conjunto de bienes culturales o patrimoniales puestos a disposición del público como colecciones digitalizadas;

III. Base de datos. Conjunto de datos relacionados o pertenecientes a un mismo contexto, almacenados sistemáticamente para su posterior uso. Las bases de datos permiten almacenar grandes cantidades de datos para su organización y administración, así como para construir relaciones entre ellos ya que éstos por sí solos y de manera aislada no reflejan ningún tipo de información;

IV. Capas geoespaciales. Conjunto de datos con referencia geoespacial que tienen valor en un contexto geográfico y que deben ser descritas por metadatos;

V. Catálogo de colecciones. Listado o inventario de las colecciones universitarias digitales proveedoras de Datos Abiertos así como sus metadatos, que se encuentra registrado en la CCUD;

VI. CCUD. Coordinación de Colecciones Universitarias Digitales;

VII. Colección. Conjunto de datos primarios de la misma clase, reunidos y clasificados por su valor académico, cultural, histórico y de identidad universitaria;

VIII. Colección digital. Conjunto de datos primarios de los recursos que conforman las colecciones codificadas en formatos de medios electrónicos que se pueden leer y compartir por medio de computadoras y otros dispositivos electrónicos;

IX. Colección universitaria digital. Conjunto de datos primarios tangibles e intangibles, reunidos y clasificados por su valor académico, cultural, histórico y de identidad universitaria, propiedad de la Universidad Nacional Autónoma de México, codificados en formatos de medios electrónicos que se pueden leer y compartir por medio de computadoras y otros dispositivos electrónicos;

X. Colección universitaria. Conjunto ordenado de bienes tangibles e intangibles, reunidos y clasificados por su valor académico, cultural, histórico y de identidad universitaria, propiedad de la Universidad Nacional Autónoma de México;

XI. Conjunto de datos. Serie de datos estructurados, vinculados entre sí y agrupados dentro de una misma unidad temática y física, de forma que pueden ser procesados apropiadamente para obtener información;

XII. Constancia de publicación. Documento emitido por la CCUD a nombre del responsable de la colección, donde se hace constar que datos de la misma han sido

integrados a la Plataforma Informática de la CCUD;

XIII. Datos abiertos. Datos digitales de carácter público que administra la Coordinación de Colecciones Universitarias Digitales, y que en términos de las disposiciones aplicables no tienen naturaleza de reservada o confidencial, siendo accesibles en línea, utilizados, reutilizados y redistribuidos por cualquier interesado;

XIV. Datos primarios. Los datos recabados directamente de los hechos que se investigan. Su interpretación o integración deriva en datos secundarios;

XV. Dependencias administrativas. Todas aquéllas que realizan actividades administrativas en apoyo a las funciones sustantivas de la Universidad;

XVI. Entidades académicas. Todas aquéllas que realizan actividades de docencia, investigación, difusión y extensión como son los institutos, centros, facultades y escuelas y los centros de extensión universitaria;

XVII. Estándar. Modelo que sirve como punto de referencia para los campos que deben integrarse en una tabla o base de datos, así como el tipo de datos que deben capturarse en colecciones de la misma naturaleza. De esta manera pueden integrarse y ser consultadas como si se tratara de una sola;

XVIII. Formato abierto. Conjunto de características técnicas y de presentación que corresponden a la estructura lógica usada para almacenar datos en un acervo

digital, cuyas especificaciones técnicas están disponibles públicamente, que no suponen una dificultad de acceso y que su aplicación y reproducción no están condicionadas a contraprestación alguna;

XIX. Investigación transdisciplinaria. Estudio realizado por facultades, escuelas, institutos, centros o programas de diversas disciplinas, que de forma articulada e integrada desarrollan y fortalecen esfuerzos institucionales para el estudio y solución de algún tema, construyendo propuestas a partir de diferentes puntos de interés, y que trasciende las fronteras entre diversas disciplinas;

XX. Metadatos. Datos estructurados y actualizados que describen el contexto y las características de contenido, captura, procesamiento, calidad, condición, acceso y distribución de un conjunto de datos, que sirven para facilitar su búsqueda, identificación y uso;

XXI. Migración de datos. Transferencia de los datos de una base de datos de origen a otra de destino que tiene un diseño y estructura distinta;

XXII. Objetos digitales. Archivos creados en formato electrónico o la digitalización de un archivo analógico, que se accede a ellos mediante dispositivos digitales;

XXIII. Plataforma informática de la CCUD. Sistema interoperativo desarrollado por la CCUD, conformada por bases de datos estandarizadas y servicios web que integran datos de colecciones, capas geoespaciales y objetos digitales;

XXIV. *Portal de Datos Abiertos UNAM Colecciones Universitarias*. Conjunto de páginas y puntos de acceso que permiten consultar los datos públicos de las colecciones digitales de la Universidad Nacional Autónoma de México puestas a disposición de la sociedad a través de la página electrónica www.datosabiertos.unam.mx.

XXV. Propiedad de la UNAM. Objetos que conforman las colecciones universitarias, así como los documentos que guardan datos e información que los describe, tales como bases de datos y tablas, entre otros, que pertenecen a la UNAM.

XXVI. Registros de colecciones. Conjunto de datos correspondientes a un registro de una base de datos que tienen valor por sí mismos;

XXVII. Repositorio. Plataforma digital que siguiendo los estándares internacionales, almacena, mantiene y preserva la información educativa, académica, científica, tecnológica y de innovación, la cual se deriva de las investigaciones, productos educativos y académicos;

XXVIII. Términos de libre uso de datos abiertos en la UNAM. Autorización que otorga la UNAM, como titular de los derechos patrimoniales de una(s) base(s) de datos, a un usuario, indicando lo que puede o no hacer con la misma al ser publicado como datos abiertos;

XXIX. UNAM. Universidad Nacional Autónoma de México.

CAPÍTULO II**DE LA INTEGRACIÓN E INTEROPERABILIDAD DE DATOS
ABIERTOS DE COLECCIONES UNIVERSITARIAS**

4. La CCUD pondrá a disposición los conjuntos de datos como Datos Abiertos de conformidad con la normativa universitaria a través del *Portal de Datos Abiertos UNAM Colecciones Universitarias*.

5. La integración de los datos y conjuntos de datos del *Portal de Datos Abiertos UNAM Colecciones Universitarias* se realizará a través de la Plataforma que para el efecto se desarrolla, mantiene y actualiza, con el objeto de asegurar el buen funcionamiento y la disponibilidad de la información.

6. Los datos vinculados por el *Portal de Datos Abiertos UNAM Colecciones Universitarias*, también pueden serlo por otras páginas o portales de esta Casa de Estudios o de otras entidades académicas, gubernamentales y civiles, en términos de los presentes Lineamientos o de acuerdo a los instrumentos que en su caso se suscriban.

7. El *Portal de Datos Abiertos UNAM Colecciones Universitarias* debe cumplir con los principios generales siguientes:

I. Accesibilidad: Tener las características de acceso reconocidas a nivel internacional, incluyendo las que se refieren a respetar y considerar las necesidades específicas de las personas con capacidades especiales;

II. Adecuación Tecnológica: Buscar la neutralidad tecnológica y el aprovechamiento de estándares abiertos;

III. Asociación: Compartir información y conocimiento con otros sistemas de entidades académicas y dependencias;

IV. Confidencialidad: Garantizar, en términos de las disposiciones jurídicas aplicables, la no divulgación de datos o información a terceros o a sistemas no autorizados de datos personales, información no autorizada por terceros y la derivada de cláusulas de confidencialidad;

V. Conservación: Asegurar la integridad, confiabilidad y disponibilidad de los datos e información contenidos en los mismos a través del tiempo;

VI. Equilibrio: Establecer un balance entre los aspectos de seguridad de los datos e información y las aplicaciones de acceso a los mismos, de forma que no sea un obstáculo para la interoperabilidad con otros servicios digitales de esta Casa de Estudios;

VII. Igualdad: Evitar restricciones o limitaciones para los usuarios que decidan relacionarse con el *Portal de Datos Abiertos UNAM Colecciones Universitarias*, con independencia de la utilización de medios de contacto tradicionales, en términos de las disposiciones jurídicas aplicables;

VIII. Integridad: Garantizar que los datos o información contenidos en el *Portal de Datos Abiertos UNAM Colecciones Universitarias* permanezcan completos e inalterados y, en su caso, que sólo hayan sido modificados por la fuente de origen de la información;

IX. Legalidad: Cumplir la normativa aplicable requerida para el intercambio de datos o información de contenidos en el *Portal de Datos Abiertos UNAM Colecciones Universitarias*, de la misma manera que sucede en el intercambio de información por medios físicos;

X. Reutilización: Generar los medios para poner a disposición y compartir la información, funcionalidades y soluciones tecnológicas, entre aquellas entidades académicas y dependencias universitarias que lo requieran, y

XI. Seguridad: Mantener como mínimo, el mismo nivel de garantías y seguridad que se tiene para el intercambio por medios físicos, así como adoptar procesos de ciberseguridad, en términos de lo que establezcan las disposiciones aplicables a las autoridades competentes.

8. Los estándares abiertos deben tener como mínimo, las características siguientes:

I. Disponibilidad;

II. Que los derechos de autor estén disponibles, libres de regalías y condiciones;

III. Maduros;

IV. Internacionalmente aceptados;

V. De fácil distribución, y

VI. Con amplio soporte en el mercado.

9. Corresponderá a la CCUD expedir, modificar y actualizar, previa opinión favorable

del Secretario de Desarrollo Institucional y del Comité Técnico del Programa Toda la UNAM en Línea, su Reglamento Interno y el Manual de Implementación de Datos Abiertos donde se establecerá lo siguiente:

I. Las normas de operación del *Portal de Datos Abiertos UNAM Colecciones Universitarias*;

II. Las especificaciones técnicas básicas que las entidades académicas y dependencias universitarias deben observar para garantizar su interoperabilidad con el *Portal de Datos Abiertos de la UNAM Colecciones Universitarias*;

III. Las características y los mecanismos de identidad que permitan la identificación digital de las entidades y dependencias universitarias, y

IV. Las demás que resulten necesarias para el cumplimiento del objeto del *Portal de Datos Abiertos UNAM Colecciones Universitarias*.

CAPÍTULO III DEL PORTAL DE DATOS ABIERTOS UNAM COLECCIONES UNIVERSITARIAS

10. Dentro del Programa Toda la UNAM en Línea, la CCUD dará acceso a la Información como Datos Abiertos mediante el *Portal de Datos Abiertos UNAM Colecciones Universitaria*, los cuales, para ser considerados como datos abiertos, los conjuntos de datos deben contar con las características mínimas siguientes:

I. Accesibles: Los datos están disponibles para la gama más amplia de usuarios, para cualquier propósito;

II. De libre uso: Citan la fuente de origen como único requerimiento para ser utilizados libremente;

III. Gratuitos: Se obtienen sin entregar a cambio contraprestación alguna;

IV. Integrales: Contienen el tema que describe a detalle y con los metadatos necesarios;

V. Legibles por máquinas: Deberán estar estructurados, total o parcialmente, para ser procesados e interpretados por equipos electrónicos de manera automática;

VI. No discriminatorios: Los datos están disponibles para cualquier persona, sin necesidad de registro;

VII. Oportunos: Son actualizados, periódicamente, conforme se generen;

VIII. Permanentes: Se conservan en el tiempo, para lo cual, las versiones históricas relevantes para uso público se mantendrán disponibles con identificadores adecuados al efecto, y

IX. Primarios: Proviene de la fuente de origen con el máximo nivel de desagregación posible.

11. La integración de los datos y conjuntos de datos en el *Portal de Datos Abiertos UNAM Colecciones Universitarias* se realizará a través de la Plataforma Informática de la CCUD,

de acuerdo con las directrices que para tal efecto establezca el Manual de Implementación de Datos Abiertos de la CCUD.

12. La priorización para la publicación de Datos Abiertos y conjuntos de datos de colecciones universitarias en el *Portal de Datos Abiertos UNAM Colecciones Universitarias*, será de acuerdo a los criterios siguientes:

I. Datos de las entidades académicas y dependencias de la UNAM que ya son de carácter público o son similares a los publicados por otras iniciativas nacionales e internacionales;

II. Datos con alta demanda de consulta por parte de entidades académicas, de gobierno o de la sociedad civil, y

III. Datos que sean relevantes para la comprensión y solución de los grandes retos nacionales, así como para el diseño e implementación de políticas públicas.

13. En el *Portal de Datos Abiertos UNAM Colecciones Universitarias* podrán publicarse datos provenientes de distintas fuentes, como:

I. Registros de colecciones;

II. Bases de datos y sus metadatos;

III. Capas geoespaciales y sus metadatos, y

IV. Metadatos de objetos digitales.

14. Las colecciones susceptibles a ser publicadas en el *Portal de Datos Abiertos UNAM Colecciones Universitarias* son las que resul-

ten de las actividades de docencia, investigación y difusión y extensión de la cultura, realizadas por esta Casa de Estudios.

15. Las colecciones que se publiquen como Datos Abiertos pertenecen a la UNAM y deben contar con al menos un responsable; para tal efecto la CCUD le otorgará constancia de publicación y, en su caso, el órgano colegiado que corresponda determinará el valor curricular de la misma.

16. La entidad académica o dependencia que resguarda la colección susceptible de ser publicada en el *Portal de Datos Abiertos UNAM Colecciones Universitarias*, debe contar con documentación que acredite que cuenta con los derechos de publicación y divulgación de los registros, bases de datos, capas geoespaciales y objetos digitales.

17. El *Portal de Datos Abiertos UNAM Colecciones Universitarias* contará con un Catálogo de Colecciones Universitarias elaborado por la CCUD, mismo que debe incluir los metadatos mínimos siguientes:

- I. Nombre de la colección;
- II. Entidad académica o dependencia a la que pertenece;
- III. Código de la colección;
- IV. Código de la entidad académica o dependencia;
- V. Número de catálogo;
- VI. Nombre y cargo del responsable;

VII. Contacto de la colección, y

VIII. Diccionario de datos documentado.

18. Los datos deben estructurarse siguiendo el estándar definido por la CCUD y cumplir con los criterios de control de calidad estructural descritos en el Manual de Implementación de Datos Abiertos emitido por la CCUD, mismos que deben ser adecuados para cada colección.

19. Para la identificación única de los datos en la UNAM, debe emplearse el triplete conformado por lo siguiente:

- I. El código de la institución;
- II. El código de la colección, y
- III. El número del catálogo.

20. La publicación de los datos en línea debe estar acompañada por su forma de citar, así como la fecha y hora de su última actualización.

21. Los componentes mínimos que deben integrar los metadatos de un conjunto de datos para la información publicada como Datos Abiertos son:

- I. Título: Nombre descriptivo de la base de datos que facilita la búsqueda, identificación y entendimiento del conjunto de datos;
- II. Descripción: Una explicación de los datos, con suficiente detalle para que los usuarios puedan entender su contexto y determinar si es de su interés;

III. Palabras clave: Etiquetas o palabras clave que facilitan a los usuarios la búsqueda del conjunto de datos. Es importante considerar el uso de términos tanto técnicos como no técnicos;

IV. Última modificación: La fecha más reciente en que se cambió, modificó o actualizó el conjunto de datos;

V. Publicador: La entidad responsable de la publicación del conjunto de datos;

VI. Punto de contacto: Información de contacto relevante al conjunto de datos;

VII. Identificador: Debe ser único dentro del catálogo de datos de la entidad;

VIII. Distribución: Conecta un conjunto de datos con sus distribuciones disponibles, y

IX. Términos de Libre Uso: Liga directa al documento que establece los derechos de libre uso del conjunto de datos.

CAPÍTULO IV

DE LA INFORMACIÓN CONFIDENCIAL Y RESERVADA

22. La información clasificada como confidencial o reservada en términos de los principios y disposiciones aplicables en materia de transparencia, acceso a la información pública y protección de datos personales, no será susceptible de publicación en el *Portal de Datos Abiertos UNAM Colecciones Universitarias*.

23. La información que contenga datos personales es susceptible de ser publicada

en el *Portal de Datos Abiertos UNAM Colecciones Universitarias* cuando persiga fines estadísticos, científicos, históricos o de interés general, previo procedimiento por el cual no puedan asociarse los datos personales con el individuo al que se refieran.

24. La información que por su naturaleza pudiera provocar daños a la integridad de la especie u objeto descrito en Datos Abiertos, no es susceptible de publicación en el *Portal de Datos Abiertos UNAM Colecciones Universitarias*.

25. La información que con su publicación pudiera tener una mayor afectación que beneficio al ser publicada, no será susceptible de ser incorporada al *Portal de Datos Abiertos UNAM Colecciones Universitarias*.

26. Para valorar el beneficio de la publicación de la información, las entidades académicas y dependencias universitarias, deben tomar en consideración los siguientes aspectos:

I. La no afectación a la integridad de personas, especie u objeto descritos;

II. El viable desarrollo de proyectos en curso;

III. La no afectación a los acuerdos de confidencialidad;

IV. La observancia a los convenios de propiedad intelectual, y

V. El apego de los procedimientos legales en curso.

La definición de estos criterios para cada colección en particular será responsabilidad de la entidad académica o dependencia que resguardan las colecciones universitarias.

CAPÍTULO V
DE LA INTERPRETACIÓN

27. La interpretación normativa de los presentes Lineamientos quedará a cargo del Abogado General.

TRANSITORIOS

ÚNICO.- LOS PRESENTES LINEAMIENTOS ENTRARÁN EN VIGOR EL DÍA DE SU PUBLICACIÓN EN *GACETA UNAM*.

APROBADOS POR LA COORDINACIÓN DE COLECCIONES UNIVERSITARIAS DIGITALES, EL 1⁰ DE SEPTIEMBRE DE 2015.

TÉRMINOS DE LIBRE USO DE DATOS ABIERTOS DE LA UNAM

Al ingresar, consultar y utilizar los **Datos Abiertos** contenidos en este portal de la Universidad Nacional Autónoma de México ("UNAM"), acepta expresamente los siguientes:

Términos de Libre Uso de Datos Abiertos de la UNAM

Los presentes "Términos de libre uso" promueven la consulta de acervos digitales universitarios de carácter público y accesible en línea, en pleno respeto a la legislación aplicable incluida aquella en materia de propiedad intelectual.

1. Usted como "USUARIO" reconoce que los Datos Abiertos contenidos en el presente son:

- a) **Accesibles:** los datos están disponibles para la gama más amplia de usuarios, para cualquier propósito;
- b) **De libre uso:** Citan la fuente de origen como único requerimiento para ser utilizados libremente;
- c) **Gratuitos:** Se obtienen sin entregar a cambio contraprestación alguna;
- d) **Integrales:** Contienen el tema que describe a detalle y con los metadatos necesarios;
- e) **Legibles por máquinas:** Deberán estar estructurados, total o parcialmente, para ser procesados e interpretados por equipos electrónicos de manera automática;
- f) **No discriminatorios:** Los datos están disponibles para cualquier persona sin necesidad de registro;
- g) **Oportunos:** Son actualizados periódicamente, conforme se generen;
- h) **Permanentes:** Se conservan en el tiempo, para lo cual, las versiones históricas relevantes para uso público, se mantendrán disponibles a través de los identificadores institucionales adecuados para tal efecto;
- i) **Primarios:** Proviene de la fuente de origen con el máximo nivel de desagregación posible.

2. Cada "USUARIO" se obliga a no utilizar la información con objeto de engañar o confundir al público variando el sentido original de la misma y su veracidad.

3. El "USUARIO" reconoce que la "UNAM" se inspira en los principios de libertad de cátedra e investigación, acogiendo en su seno, con propósitos exclusivos de docencia e investigación, todas las corrientes del pensamiento y las tendencias de

carácter científico y social. Por lo que por ningún motivo deberá aparentar que el uso que haga de la información, representa una postura oficial de la "UNAM" o que el mismo está avalado, integrado, patrocinado o apoyado por la fuente de origen.

4. El "USUARIO" se obliga a otorgar los créditos correspondientes a la "UNAM" en su carácter de Institución generadora y cuando sea posible, mencionar la entidad o dependencia Universitaria que haya coadyuvado en la generación de la información, precisando la fecha de actualización más reciente según los siguientes casos:

a) En caso que la información generada provenga de la consulta a un solo registro, debe citarse de la siguiente manera:

Modelo:

[Departamento o Área], [Entidad o Dependencia con acrónimo], [Nombre del registro], de: [Nombre de la Colección]. En *Portal de Datos Abiertos UNAM Colecciones Universitarias* (en línea), México, Universidad Nacional Autónoma de México.

Disponible en: [Liga de internet al registro consultado]

Fecha de actualización: dd/mm/aaaa, hora de actualización

Fecha de consulta: dd/mm/aaaa, hora de consulta

Ejemplo:

Departamento de Botánica, Instituto de Biología (IBUNAM), Plantago patagonica Jacq., ejemplar de: Herbario Nacional de México (MEXU), Plantas

Vasculares. En *Portal de Datos Abiertos UNAM Colecciones Universitarias* (en línea), México, Universidad Nacional Autónoma de México.

Disponible en:

<http://datosabiertos.unam.mx/IBUNAM;-MEXU:77459>

Fecha de actualización: 23/09/2014, 10:58:16 a.m.

Fecha de consulta: 15/06/2015, 11:14:20 p.m.

b) En caso que la información generada por la consulta provenga de más de un registro dentro de una colección, debe citarse de la siguiente manera:

Modelo: [Departamento o Área], [Entidad o Dependencia con acrónimo], [Nombre de la Colección]. En *Portal de Datos Abiertos UNAM Colecciones Universitarias* (en línea), México, Universidad Nacional Autónoma de México.

Disponible en: [Liga de internet de la consulta descargada]

Fecha de actualización: dd/mm/aaaa, hora de actualización

Fecha de consulta: dd/mm/aaaa, hora de consulta

Ejemplo:

Departamento de Botánica, Instituto de Biología (IBUNAM), Herbario Nacional de México, Plantas Vasculares. En *Portal de Datos Abiertos UNAM Colecciones Universitarias* (en línea), México, Universidad Nacional Autónoma de México.

Disponible en: <http://datosabiertos.unam.mx/>

Fecha de actualización: 23/09/2014, 10:58:16 a.m.

Fecha de consulta: 15/06/2015, 11:14:20 p.m.

c) En caso que la información generada por la consulta provenga de varios registros cuyo origen corresponde a distintas colecciones, debe citarse cada colección de manera independiente, siguiendo el modelo anterior (inciso 4.b).

5. Cada "USUARIO" reconoce la eventual inclusión de información y/o contenidos propiedad de terceros, dentro de los conjuntos de datos que se encuentran en acceso abierto por este medio. Por lo que en caso de que requiera utilizar dicha información y/o contenidos, deberá buscar forzosamente la autorización directamente con el titular de los derechos correspondientes de conformidad con la Ley Federal de Derechos de Autor y cualquier otro ordenamiento en materia de propiedad intelectual aplicable.

SALVEDADES

1. La "UNAM" no es, ni será responsable por la interpretación y aplicación que el "USUARIO" haga de los resultados obtenidos a través del uso de la información; por lo que cualquier decisión basada en su interpretación excluye a la "UNAM" de responsabilidad alguna. Asimismo, la "UNAM" no es, ni será responsable de las diferencias obtenidas por precisiones, redondeos o truncamientos numéricos, así como por cambios técnicos o tecnológicos que puedan incidir en tales resultados.

2. El PORTAL DE DATOS ABIERTO UNAM COLECCIONES UNIVERSITARIAS, así como todas las reservas, nombres de dominio, logos y marcas Universitarias son titularidad de la "UNAM" y su utilización le corresponde en exclusiva a la misma. Por consiguiente, queda prohibida su utilización por parte de cualquier tercero que carezca de autorización expresa.

3. La "UNAM" es titular de las bases de datos que se encuentran en acceso abierto por este medio, las cuales gozan de la protección que conceden las leyes nacionales, así como de los convenios y acuerdos internacionales en materia de Propiedad Intelectual, en aquellas bases de datos que por la selección o disposición de sus contenidos, constituyen creaciones intelectuales.

4. El uso no autorizado en contravención con estos "Términos de Libre Uso", podrá ser sancionado de acuerdo con la Legislación Universitaria y demás aplicable.

5. La "UNAM" se reserva el derecho a modificar en cualquier momento los presentes "Términos de Libre Uso".

6. Los presentes "Términos de Libre Uso" se rigen por las leyes mexicanas, por lo que cualquier disputa en relación a la interpretación y cumplimiento de las presentes disposiciones, así como para lo no previsto en las mismas serán sustanciadas ante la jurisdicción mexicana, sometiéndose a la jurisdicción de los Tribunales Federales de la Ciudad de México, y a las disposiciones contenidas en la Ley Federal del Derecho de Autor, en el Código Civil Federal vigente y demás aplicables al caso, renunciando el "USUARIO" expresamente al fuero que por razón de su domicilio presente o futuro pudiera corresponderle.

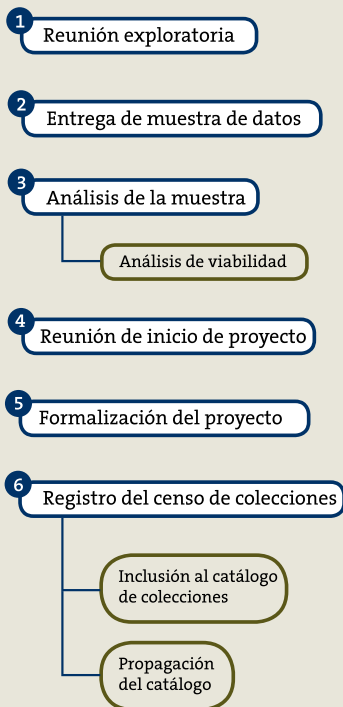
INFORMACIÓN GENERAL

El PORTAL DE DATOS ABIERTOS UNAM COLECCIONES UNIVERSITARIAS es titularidad de la Universidad Nacional Autónoma de México y es administrado por la Coordinación de Colecciones Universitarias Digitales (CCUD), el nombre de dominio <http://www.datosabiertos.unam.mx/> y todos aquellos medios que sirvan para acceder de forma directa al sitio oficial son de utilización exclusiva de la "UNAM".

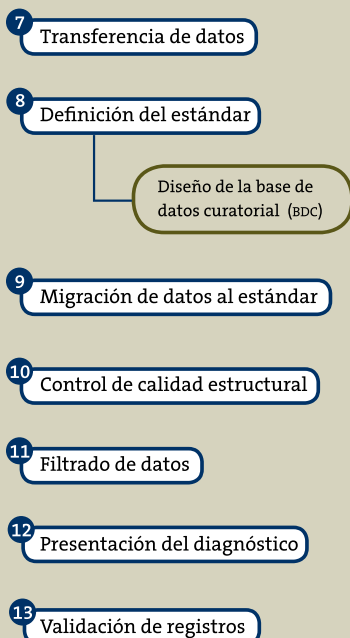
	Núm.	Tarea	Responsable	Documentos
I. Definición del proyecto	1	Reunión exploratoria	IC-CCUD	Minuta
	2	Muestra de datos	Proveedor de datos	Entrega de la muestra
	3	Análisis de la muestra	CT-CCUD	Resultado del análisis
	4	Reunión de inicio de proyecto	CT-CCUD Proveedor de datos	a) formato de solicitud de integración b) formato de lineamientos de publicación c) requerimientos técnicos de la CCUD para la transferencia de datos y objetos digitales
	5	Formalización del proyecto	IC-CCUD Proveedor de datos	Bases de colaboración
	6	Censo de colecciones	IC-CCUD Proveedor de datos	Censo
II. Arquitectura y gestión de datos	7	Transferencia de datos	Proveedor de datos	a) oficio de entrega b) base de datos c) lineamientos de publicación d) catálogos de publicación
	8	Definición el estándar	IE-CCUD	Estándar documentado
	9	Migración de datos al estándar	IE-CCUD	Base de datos con estándar, datos migrados y documentación
	10	Control de calidad estructural	IE-CCUD	Excel con diagnóstico
	11	Filtrado de datos	IE-CCUD	No aplica
	12	Presentación del diagnóstico	IE-CCUD	Excel con diagnóstico y tablas de validación
	13	Entrega de la validación de registros modificados	Proveedor de datos	Excel con registros modificados aprobados
	14	Diseño de la ficha	CT-CCUD	Descripción de campos de la ficha
III. Publicación de colecciones	15	Versión final de la base de datos	IE-CCUD	a) disco con la base de datos b) formato de permisos de publicación
	16	Permisos de publicación	Proveedor de datos	Oficio de permisos con la firma del director(a)
	17	Publicación de la colección	DT-CCUD	Colección disponible en línea
	18	Entrega de proyecto	CG-CCUD	Constancias de publicación

Protocolo de integración y publicación de colecciones en el Portal de Datos Abiertos UNAM, Colecciones Universitarias

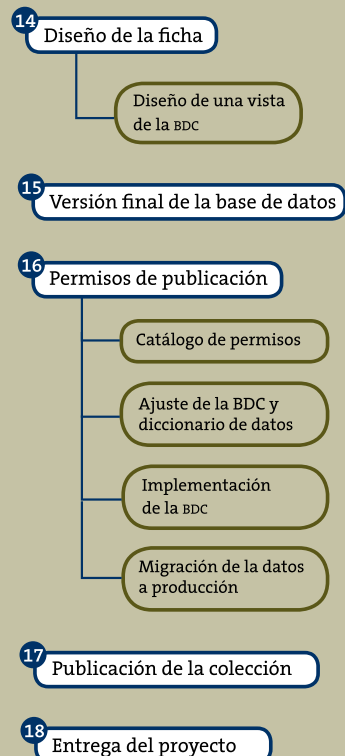
I. Definición del proyecto



II. Arquitectura y gestión de datos



III. Publicación de colecciones



Pasos en azul: realizados entre la CCUD y los proveedores de datos.
Pasos en verde: etapas técnicas de la CCUD.

Manual de Datos Abiertos
de Colecciones Universitarias Digitales
ISBN: 978-607-02-9086-2
© Derechos reservados



CCUD

COORDINACIÓN DE COLECCIONES
UNIVERSITARIAS DIGITALES



Protocolo de integración y publicación de colecciones en el Portal de Datos Abiertos UNAM, Colecciones Universitarias

Puede descargar el archivo PDF con la siguiente liga:

El *Manual de Datos Abiertos de Colecciones Universitarias Digitales*, se terminó de imprimir en marzo de 2017, en Historias en Papel, ubicados en Cascada 402, Col. Banjidal, delegación Iztapalapa, Ciudad de México, C.P. 09450. El tiraje consta de mil ejemplares. Para su formación se usó la tipografía TheSerif.

