



CCUD

COORDINACIÓN DE COLECCIONES
UNIVERSITARIAS DIGITALES

PROTOCOLO DE INTEGRACIÓN
Y PUBLICACIÓN EN EL
PORTAL DE DATOS ABIERTOS UNAM,
COLECCIONES UNIVERSITARIAS

COORDINACIÓN DE COLECCIONES UNIVERSITARIAS DIGITALES

Presentación

La Coordinación de Colecciones Universitarias Digitales (CCUD) integra, administra y difunde, como datos abiertos, los acervos universitarios —científicos, históricos y culturales— y los productos concluidos de la investigación. Éstos se publican a través de la Plataforma Informática de la CCUD  **aurora** para su consulta en el *Portal de Datos Abiertos UNAM, Colecciones Universitarias* (datosabiertos.unam.mx).

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 del director (a) de la entidad o dependencia proveedora de datos.

El “Protocolo de integración y publicación de colecciones en el *Portal de Datos Abiertos UNAM, Colecciones Universitarias*”¹ 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 los curadores, responsables de los datos y del proyecto. Los pasos que se describen en este folleto muestran únicamente las diferentes etapas en las que están involucrados.

Las colecciones universitarias digitales

Las colecciones universitarias digitales son conjuntos de datos de bienes tangibles e intangibles, reunidos y clasificados por su valor académico, cultural, histórico y de identidad universitaria, propiedad de la UNAM. El *Portal de Datos Abiertos UNAM* publica colecciones que son producto de las actividades universitarias de docencia, investigación, difusión y extensión de la cultura. Sus datos, codificados en formatos electrónicos, provienen de tres fuentes principales:

a) Acervos: fuentes de consulta, referencia o preservación, compuestas principalmente de datos primarios. Por ejemplo, las 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: concebidas para la obtención de nuevo conocimiento y para ser aplicadas en la solución de problemas o preguntas académicas. Se componen principalmente de datos secundarios como datos de experimentos o resultados de análisis estadístico;

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

Protocolo de integración y publicación

El “Protocolo de integración y publicación de colecciones en el *Portal de Datos Abiertos UNAM, Colecciones Universitarias* consta de 18 pasos divididos en tres secciones:

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

Áreas de la CCUD

CG: Coordinación General

CT: Coordinación Técnica

DT: Desarrollo Tecnológico

IE: Informática y estadística

IC: Inventario de Colecciones y Comunicación

I. Definición del proyecto

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 presente 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: IC

2. Entrega de una muestra de datos

La entidad o dependencia envía la muestra 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 para su análisis.

Entregable: Muestra de datos

Responsable: Responsable de la colección (proveedor de datos)

3. Análisis de la muestra

El área de Informática y Estadística realiza un análisis de la estructura de la base de datos para contar con indicadores que permitan 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. 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: CT

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 contenidos (objetos digitales, mapas o datos).
- 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 dependencia o entidad.

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 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.

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: Formatos de solicitud de integración y lineamientos; requerimientos técnicos

Responsable: CT y proveedor de datos

5. Formalización del proyecto

Bases de colaboración o convenio firmado por los titulares de las dependencias y entidades involucradas.

Entregable: Bases de colaboración

Responsable: IC y responsable del proyecto (proveedor de datos)

6. Registro del censo de colecciones

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

Entregable: Censo de las colecciones

Responsable: IC y responsable del proyecto (proveedor de datos)

II. Arquitectura y gestión de datos

7. Transferencia de datos y objetos digitales

Es la entrega de la base de datos de la colección en su 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 inicio de proyecto, paso 4*).
- c) Catálogos para el control de calidad estructural (*si aplica*).

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)

	No	Tarea	Responsable	Documentos	Notas
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 participación	

8. Definición del estándar

La integración de datos de diversas fuentes requiere de estándares para lograr un acceso unificado de la información. 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. Las bases de datos que cumplen con éstas nos permiten acceder a la información unificada de diversas fuentes, eliminando barreras como idioma, lugar de origen o formatos, y que sea comprendida y utilizada por la mayor cantidad posible de usuarios. En algunos casos, el estándar sugiere o hace obligatorio el uso de convenciones como la escritura de las fechas, las claves de país, el idioma, las unidades métricas, los acrónimos, etcétera.

La CCUD, en colaboración con el curado o responsable de la colección, define las normas para cada colección que se integra al *Portal de Datos Abiertos UNAM* priorizando el uso de estándares internacionales. En caso de que no exista, se define uno nuevo. Ambos casos se documentan considerando las definiciones de los campos y las adecuaciones.

Entregable: Estándar documentado

Responsable: IE

9. Migración de datos al estándar

La CCUD diseña una “base de datos curatorial” donde los datos originales se migran al estándar definido. Contiene todos los datos de la colección y permite estructurarlos de acuerdo con los estándares definidos. La base de datos curatorial es de uso del área de Informática y Estadística para el desarrollo de los procesos de integración. Todos los pasos de este proceso se documentan, así como las modificaciones realizadas, y se automatiza en la medida de lo posible.

Entregable: Base de datos curatorial (estandarizada) con datos migrados y documentación

Responsable: IE

10. Control de calidad estructural

La CCUD aplica una evaluación técnica donde los datos se califican con diferentes métodos de validación y catálogos. El control de ca-

lidad estructural se realiza a la base de datos curatorial para que cumpla con los lineamientos y estándares que reducen la presencia de errores y permita contar con información de calidad, coherente y actualizada. Durante el proceso, 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 datos continúan el proceso de publicación;
- c) **Inconsistentes** con los catálogos: se reportan al curador o responsable de la colección para su revisión. Estos datos 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.

Entregable: Excel con diagnóstico de control de calidad

Responsable: IE

11. Filtrado de datos

Es el paso para detectar y etiquetar los datos que no cumplan con los criterios definidos en los lineamientos de publicación de la entidad o dependencia (*descritos en el paso 4 “Reunión de inicio de proyecto” y entregados en el paso 7 “Transferencia de datos”*).

Entregable: No aplica

Responsable: IE

12. Presentación del diagnóstico del control de calidad estructural

Es una reunión donde la CCUD presenta al curador o responsable del proyecto la documentación del proceso de gestión de datos, que incluye: documentación de estándares, análisis de la transferencia, catálogos utilizados y “tablas de validación” en las que se encuentran las propuestas de modificación. Como resultado de esta reunión:

- a) La CCUD entrega las tablas de validación que incluyen únicamente los datos clasificados como “modificados” e “inconsistentes” para su validación.
- b) Se solicita fecha de entrega de la validación de registros modificados.
- c) Los registros consistentes se aprueban para continuar su proceso de publicación.
- d) Los registros inconsistentes se devuelven al responsable de los datos o curador 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: IE

13. Entrega de la validación de datos modificados

Durante esta etapa, el curador o responsable revisa y valida las modificaciones propuestas en el control de calidad estructural. Una vez concluida la revisión, entregan las tablas de validación.

Entregable: Tablas de validación con datos modificados aprobados

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

III. Publicación de colecciones

14. Diseño de la ficha

La CCUD publica cada registro de la colección en una ficha digital independiente y con una dirección única (URL) para que pueda ser localizada por motores de búsqueda (como *Google*) y recuperada por los usuarios. Su diseño contempla las particularidades de cada colección publicada en el *Portal de Datos Abiertos UNAM* sin perder una estructura homogénea que las integra.

Entregable: Definición de campos en la ficha

Responsable: CT

15. Versión final de la base de datos a publicar

La CCUD genera la base de datos final en el estándar definido, con los registros filtrados y las modificaciones validadas. Ésta se entrega en un CD con la copia exacta de la base de datos que se publicará en el *Portal de Datos Abiertos UNAM* y se anexa un formato de permisos de publicación (*paso 16*).

Entregable: Disco con la base de datos a publicarse, formato de permisos de publicación (*la CCUD entrega*)

Responsable: IE

16. Permisos de publicación

El director(a) 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 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(a)

Responsable: IC y proveedor de datos

17. Publicación de la colección

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

Entregable: Colección disponible en datosabiertos.unam.mx

Responsable: DT

18. Entrega de proyecto

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

Entregable: Constancias de publicación

Responsable: CG

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

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.

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;

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

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;

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;

Plataforma Informática de la CCUD (Aurora). Sistema interoperativo desarrollado por la CCUD, conformado por bases de datos estandarizadas y servicios web que integran datos de colecciones, capas geoespaciales y objetos digitales.



Universidad Nacional Autónoma de México

Coordinación de Colecciones Universitarias Digitales

Directorio

Dra. Tila María Pérez Ortiz
Coordinación General

Biól. Joaquín Giménez Héau
Coordinación Técnica

M. en C. Ariana Chávez Méndez
Líder de Proyecto

Lic. en C. C. Rubén Sáenz González
Desarrollo Tecnológico

Biól. Daniel Pérez Castillo
Informática y Estadística

Lic. Edurne Uriarte Santillán
Inventario de Colecciones y Comunicación

Para mayor información comunicarse con:

Inventario de Colecciones y Comunicación: 5623 7723

Conmutador: 5622 8971

Correo: contacto@ccud.unam.mx

