

Curso Manejo de Datos de Biodiversidad

La Plata – Agosto 2017

Programa

UNIDAD 1

Sección **Digitalización y Manejo de Datos de Colecciones de Historia Natural**

Día 1, 09:00 – 10:30 hs; Día 1, 11:00 – 12:30 hs

Workshop

Objetivos: a) Presentar la digitalización en el contexto de compartir los datos de biodiversidad globalmente, enfatizando la importancia de la integración de esta actividad en flujos de trabajo más amplios. b) Identificar los pasos críticos en los que la digitalización – y actividades relacionadas – pueden ser mejorados para aumentar la eficiencia de los proyectos de digitalización de datos.

Tópicos:

- ✓ Importancia de la digitalización de datos en el contexto de un flujo de trabajo para compartir los datos globalmente.
- ✓ ¿En qué consiste un flujo de trabajo? De las etiquetas a los portales de datos en línea: qué tiene y que debería suceder en el medio para que los datos de biodiversidad estén ampliamente disponibles.
- ✓ Importancia de digitalizar lo mejor posible la primera vez: desafíos relacionados con la modificación de flujos de trabajo completos más tarde en el proceso.
- ✓ Tratando de ser eficiente: qué y cómo priorizar: a) estableciendo objetivos de acuerdo a intereses de una institución particular o de investigación, b) fuentes y recursos disponibles y necesarios para lograr con éxito la digitalización de colecciones de historia natural.
- ✓ Claves para mejorar el flujo de trabajo paso a paso y en el proceso completo: conocimiento que debería ser adquirido y colaboraciones clave que deberían establecerse para conseguir los resultados esperados de una manera más eficiente y en un tiempo razonable.

Sección **Herramientas para el manejo de datos de biodiversidad**

Día 1, 14:00 – 15:30 hs

Workshop

Objetivos: Presentar a la audiencia diferentes herramientas y programas para el manejo de datos comúnmente utilizadas para capturar datos de biodiversidad en colecciones de historia natural, y proveer claves sobre cómo escoger la opción que mejor se adapta a los intereses y características de una Institución/Colección.

Tópicos:

- ✓ Cómo organizar datos de biodiversidad: planillas vs. bases de datos relacionales. Ventajas de las bases de datos relacionales y de la integración de los datos.
- ✓ Herramientas y programas para manejo de datos: cuáles son las opciones y cómo elegir de acuerdo a objetivos particulares, basado en el alcance de los datos, considerando las prácticas de las colecciones y buscando máxima eficiencia. Breve reseña sobre Specify, Arctos, Zoorbar y otros sistemas de manejo de datos de biodiversidad.
- ✓ Una vez que se ha escogido un esquema de manejo de datos, cómo lograr una digitalización más eficiente. Importancia de tener pasos de prueba (evaluación de recursos – personal, tiempo y dinero – necesarios) y de ajustar el proceso antes de embarcarse en una actividad de digitalización a gran escala.

UNIDAD 2

Sección Estándares para compartir datos: Darwin Core

Día 1, 16:00 – 17:30 hs

Workshop

Objetivos: a) Reseñar los estándares de datos de biodiversidad disponibles actualmente resaltando la importancia de su uso por parte de los proveedores de datos. b) Presentar el estándar Darwin Core en detalle, enfatizando en las mejores prácticas de uso.

Tópicos:

- ✓ ¿Qué es un estándar? Términos de un estándar vs. contenido estándar. ¿Por qué se necesitan estándares para compartir datos de biodiversidad? Variedad de formatos en que pueden presentarse los datos e importancia de la estandarización.
- ✓ Estándares de datos de biodiversidad más utilizados: Dublin Core, ABCD, Darwin Core, Audubon Core. Introducción a TDWG (Biodiversity Information Standards).
- ✓ Estándar Darwin Core: estructura general y definición de términos. Cómo ajustar datos de biodiversidad dentro del estándar Darwin Core. Recomendaciones y mejores prácticas.

Sección Estándares para compartir datos: Darwin Core Archive (DwC-A)

Día 2, 09:00 – 10:30 hs

Workshop

Objetivos: a) Presentar diferentes herramientas para estandarizar datos y ejemplificar los desafíos con tipos particulares de datos. b) Introducir Darwin Core Archive y su uso.

Tópicos:

- ✓ Cómo estandarizar sin perder información: herramientas para solucionar problemas particulares de estandarización y desafíos a la hora de mapear datos de colecciones de historia natural a formatos estándar. Se presentan ejemplos de los campos correspondientes a fechas, datos geográficos taxonomía.
- ✓ Darwin Core Archive: ¿de qué se trata? Introducción a su construcción y uso para compartir datos de biodiversidad.

Sección Estándares para compartir datos: Integrated Publishing Toolkit (IPT)

Día 2, 11:00 – 12:30 hs

Workshop

Objetivos: Introducir la instalación y uso del Integrated Publishing Toolkit (IPT), GBIF.

Tópicos:

- ✓ Integrated Publishing Toolkit (IPT): descripción y uso actual en flujos de intercambio de datos de biodiversidad.
- ✓ Integrated Publishing Toolkit (IPT): Cómo instalar IPT y demostración de cómo utilizarlo con datos de colecciones de historia natural.

Sección Estándares para compartir datos: Darwin Core Archive (DwC-A)

Día 2, 14:00 – 15:30 hs

Laboratorio

Objetivos: Proveer la oportunidad de practicar estandarización de un conjunto de datos dado y construir y validar el correspondiente Darwin Core Archive.

Actividades:

- ✓ Estandarización de un conjunto de datos dado de acuerdo con las mejores prácticas de uso del estándar Darwin Core. Análisis de los problemas encontrados y cómo solucionarlos.
- ✓ Construcción del Darwin Core Archive correspondiente utilizando la herramienta GBIF Darwin Core Archive Assistant. Puesta a prueba del archivo construido utilizando la herramienta GBIF Darwin Core Archive Validator.

Sección Estándares para compartir datos: Integrated Publishing Toolkit (IPT)

Día 2, 16:00 – 17:30 hs

Laboratorio

Objetivos: a) Proporcionar la oportunidad de practicar la instalación y manejo de IPT. b) Proporcionar asistencia con IPT ya instalado en el MLP.

Actividades:

- ✓ Instalación y configuración local del Integrated Publishing Toolkit (IPT).

- ✓ Asistencia y soporte a IPT en el MLP: soluciones a problemas en IPT ya instalado.

UNIDAD 3

Sección **Evaluación y mejoramiento de la calidad de los datos: Introducción**

Día 3, 09:00 – 10:30 hs

Workshop

Objetivos: Introducir conceptos de calidad de datos de biodiversidad y los desafíos actuales, subrayando la importancia de abordajes impulsados por la comunidad en su conjunto.

Tópicos:

- ✓ Terminología sobre calidad de datos de biodiversidad: ¿qué significan términos tales como "evaluación", "mejoramiento", "procedencia", "autoridad" y "vocabulario controlado"?
- ✓ Evaluación de la calidad de los datos: ¿qué tan mal están? Reseña sobre el estado actual de los datos de biodiversidad y sobre cómo puede afectar ello a la investigación y a la toma de decisiones.
- ✓ Identificando problemas con la calidad de los datos: marcado o "flagging". Introducción al concepto de marca o "flag" como un punto de verificación de la calidad de los datos, cómo son utilizadas en GBIF y iDigBio.
- ✓ Ejemplo de evaluación de la calidad de los datos por parte de la comunidad: datos de biodiversidad del MLP a los ojos de GBIF y iDigBio ("flags") y de VertNet ("issue reports").

Sección **Evaluación y mejoramiento de la calidad de los datos: Ejemplo práctico**

Día 3, 11:00 – 12:30 hs

Workshop / Laboratorio

Objetivos: Proveer a la audiencia con la oportunidad de evaluar y mejorar la calidad de los datos de un conjunto de datos de real del MLP, utilizando información de marcas o "flags" y reportes de GBIF, iDigBio y VertNet, y de re-proveer los datos mejorados a los correspondientes repositorios.

Actividades:

- ✓ Evaluación de la calidad de los datos de un conjunto de datos real del MLP a través del análisis de las marcas o "flags" provistas por GBIF y iDigBio y de los reportes de VertNet. Identificación de diferentes tipos de problemas y las formas en que podrían solucionarse.
- ✓ Resolución de uno de los problemas identificados anteriormente. Corrección del defecto identificado en la calidad de los datos y re-provisión del conjunto de datos mejorado a los correspondientes repositorios.

Sección **Evaluación y mejoramiento de la calidad de los datos: Herramientas**

Día 3, 14:00 – 15:30 hs

Workshop

Objetivos: Introducir herramientas de calidad de datos de biodiversidad, distinguiendo entre aquellas dirigidas a solucionar problemas puntuales y aquellas integradas en flujos de trabajo, proveyendo un panorama de las herramientas que se encuentran disponibles hoy en día para el uso por la comunidad.

Tópicos:

- ✓ Herramientas para evaluar y mejorar la calidad de los datos de biodiversidad. Función única vs. integración en flujos de trabajo, ventajas y desventajas del uso de cada una sobre un conjunto de datos.
- ✓ Herramientas de evaluación: uso básico de Open Refine.
- ✓ Herramientas con función única: reseña de las herramientas disponibles, introducción a Darwin Test, Taxonomic Name Resolution Service (TNRS), Name Parser, Species Link sp Outlier, GeoLocate, entre otras.
- ✓ Herramientas integradas en flujos de trabajo: Herramientas basadas en flujos: reseña de las funciones provistas a través de los procesos de VertNet “Migrators”, y de los actores y flujos de Kurator.

Sección **Evaluación y mejoramiento de la calidad de los datos: Cerrando el círculo**

Día 3, 16:00 – 17:30 hs

Workshop

Objetivos: a) Relacionar las prácticas de mejoramiento de la calidad de los datos de biodiversidad con procesos impulsados por la comunidad. Introducir los conceptos de retroalimentación o “feedback”, y cómo los proveedores de datos pueden beneficiarse de esfuerzos conjuntos llevados a cabo por diferentes fuentes. b) Presentar elementos para comprender cabalmente las potenciales ventajas de mejorar la calidad de los datos de biodiversidad y los desafíos asociados.

Tópicos:

- ✓ Mecanismos de retroalimentación o “feedback”: de la fuente, al mundo entero, a la fuente. ¿Cómo funcionan los mecanismos de “feedback”? ¿Cómo implementar estos mecanismos eficientemente?
- ✓ Simulando pasos de verificación de la calidad de los datos de biodiversidad: ¿cuántos más datos pueden obtenerse luego de los pasos de verificación de calidad? Mejoramiento en completitud y estandarización, implicancias.
- ✓ Simulando pasos de verificación de la calidad de los datos de biodiversidad: ¿cómo lidiar con inconsistencias entre diferentes fuentes de datos? Implicancias para la

investigación: ¿en qué y en quién confiar? Efectos de la mala calidad de los datos en la investigación y en la toma de decisiones.

Consideraciones finales y Discusión

Día 3, 17:30 – 18:00 hs

Workshop