

Curso Manejo y Calidad de Datos de Biodiversidad

Mar del Plata – Noviembre 2018

Programa

APERTURA E INTRODUCCIÓN

Tópicos:

- ✓ Presentación del curso. Objetivos.
- ✓ Presentación de los docentes y de los participantes.

UNIDAD 1

Sección **Portales de datos sobre biodiversidad**

Tópicos:

- ✓ Presentación del portal de datos del Global Biodiversity Information Facility (GBIF).
- ✓ Participación en GBIF. Estructura y mecanismos de adhesión.
- ✓ Uso del portal de GBIF. Consultas de registros, conjuntos de datos y metadatos.
- ✓ Presentación del portal de datos del Sistema Nacional de Datos Biológicos (SNDB).
- ✓ Participación en el SNDB. Estructura y mecanismos de adhesión. Relación entre SNDB y GBIF.
- ✓ Uso del portal del SNDB. Herramientas de control de calidad de los datos integradas en el portal del SNDB.

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

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**

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.
- ✓ 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.

Sección **Tipos de datos**

Tópicos:

- ✓ Tipos de datos sobre biodiversidad: características.
- ✓ Reconocimiento de tipos de datos propios de los participantes.
- ✓ Usos de los distintos tipos de datos en diferentes contextos.
- ✓ Dónde y cómo compartir distintos tipos de datos. Portales disponibles, condiciones de adhesión y usos.

UNIDAD 2

Sección **Estándares para compartir datos: Darwin Core**

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 - práctica**

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.

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

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: Extensiones a Darwin Core**

Tópicos:

- ✓ Qué es un Core y qué son las extensiones a Darwin Core. Esquema de estrella.
- ✓ Ejemplos prácticos del uso de extensiones con Occurrence Core y Event Core.
- ✓ Qué extensiones hay disponibles y cómo utilizarlas.

Sección **Estándares para compartir datos: Extensiones a Darwin Core - práctica**

Actividades:

- ✓ Identificación de Cores y extensiones apropiadas para utilizar con los datos propios de los participantes.

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

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.
- ✓ ¿Qué es un recurso? Características y requerimientos.
- ✓ Cores y extensiones en IPT.

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

Actividades:

- ✓ Uso de IPT con datos propios. Creación de recursos.
- ✓ Mapeo a Darwin Core dentro de IPT.
- ✓ Uso de extensiones a Darwin Core dentro de IPT.
- ✓ Cómo completar los metadatos.
- ✓ Publicación de datos usando IPT y visualización del DwC-A generado a partir del IPT.

UNIDAD 3

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

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.
- ✓ Concepto de "fitness for use". Calidad de datos en función del objetivo.

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

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 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: Herramientas: Open Refine**

Tópicos:

- ✓ Uso básico de Open Refine para la evaluación y mejoramiento de la calidad de los datos.
- ✓ Carga de proyectos y demostración de funciones básicas.
- ✓ Llamado a servicios externos para la limpieza de fechas, taxonomía y geografía.

Sección **Evaluación y mejoramiento de la calidad de los datos: Herramientas: Open Refine - práctica**

Actividades:

- ✓ Uso de Open Refine para la limpieza de datos propios.
- ✓ Evaluación de calidad de datos y mejoramiento específico de campos seleccionados.

CONSIDERACIONES FINALES Y CIERRE

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

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.