Curso de movilización de datos sobre biodiversidad

Secretaría de GBIF

Versión 12, May 2021



Tabla de Contenido

Descripción del curso . . . 1
- Audiencia . . . 1
- Prerequisitos . . . 1
- Objetivos de aprendizaje . . . 1
- Certificación . . . 2

Archivos para descargar . . . 2
- Videos . . . 2
- Datos del ejercicio . . . 3
- Hojas de ejercicio . . . 3

Bases . . . 3
- Terminología . . . 3
  - Definiciones . . . 3
  - Software . . . 3
  - Estructuras . . . 4
- Calidad de los datos . . . 4
- Documentación . . . 8
- Flujos de trabajo de digitalización . . . 8
- Herramientas de software . . . 9
- Instalar OpenRefine . . . 10
  - Requisitos para la instalación . . . 11
  - Instalación en MS Windows . . . 11
  - Instalación en Mac . . . 13
  - Instalación en Linux . . . 19
- Revisión de las bases . . . 21

Caso de uso I: especímenes de herbario . . . 24
- Escenario . . . 24
  - Descripción . . . 25
  - Recopilación de datos . . . 26
  - Descripción del conjunto de datos . . . 26
- Ejercicios . . . 26
  - Planificación . . . 26
  - Captura de datos . . . 27
  - Gestión de datos . . . 27
  - Publicación de datos . . . 27
- Hoja de ejercicio . . . 27

Planificación . . . 27
- Recursos . . . 27
- Organizar . . . 27
- Ejercicio 1a-c . . . 27
  - Ejercicio 1a . . . 28
  - Ejercicio 1b . . . 29
  - Ejercicio 1c . . . 30
- Revisión . . . 30

Captura de datos . . . 31
Estándares y Darwin Core . . . 31
Orígenes y tipos de datos . . . 31
Captura, procesamiento y calidad de datos . . . 31
Ejercicio 2 . . . 32
Revisión . . . 32
Gestión de datos . . . 38
Principios de la gestión de datos . . . 38
Herramientas de gestión de datos . . . 38
Ejercicio 3a-c . . . 38
Ejercicio 3a . . . 38
Ejercicio 3b . . . 39
Ejercicio 3c . . . 39
Consejos para el ejercicio . . . 40
Comprobaciones de validación . . . 40
Herramientas útiles . . . 41
Revisión . . . 41
Publicación de datos . . . 42
Conceptos de publicación de datos . . . 42
Descripción general del IPT . . . 42
Instalaciones IPT de capacitación . . . 42
Demostración del IPT . . . 42
Ejercicio 4 . . . 43
Revisión . . . 43
Evaluación y certificación . . . 44
Rúbrica de planificación . . . 44
Rúbrica de captura de datos . . . 46
Rúbrica de gestión de datos . . . 48
Rúbrica de publicación de datos . . . 51
Caso de uso II: especies invasoras . . . 53
Escenario . . . 53
Descripción . . . 54
Recopilación de datos . . . 55
Descripción de datos digitales . . . 57
Hoja de ejercicios de Invasoras . . . 57
Ejercicio 1 . . . 57
Ejercicio 1a . . . 57
Ejercicio 1b . . . 58
Ejercicio 2 . . . 58
Ejercicio 3 . . . 58
Ejercicio 4 . . . 59
Caso de uso II: avistamiento de lepidópteros . . . 59
Escenario . . . 59
Descripción . . . 60
Recopilación de datos . . . 61

Ejemplo de captura de datos análoga . . . 61
Descripción de datos digitales . . . 62
Hoja de ejercicios Lepidoptera . . . 64
Ejercicio 1 . . . 65
Ejercicio 1a . . . 65
Ejercicio 1b . . . 66
Ejercicio 2 . . . 66
Ejercicio 3 . . . 66
Ejercicio 4 . . . 67
Caso de uso III: aves de la literatura . . . 67
Escenario . . . 67
Descripción . . . 68
Recopilación de datos originales . . . 69
Descripción de datos análogos . . . 69
Descripción de datos escaneados y traducidos . . . 70
Descripción de datos digitales . . . 71
Hoja de ejercicios de aves de la literatura . . . 71
Ejercicio 1 . . . 71
Ejercicio 2 . . . 72
Ejercicio 3 . . . 73
Ejercicio 4 . . . 73
Proyectos finales . . . 74
CASO DE USO II . . . 74
CASO DE USO III . . . 74
Envío del proyecto . . . 74
Evaluación del curso . . . 74
Documentación clave . . . 74
Darwin Core . . . 75
Publicación de datos . . . 75
Publicación de datos: IPT . . . 75
Digitalización . . . 76
GBIF . . . 76
Georreferenciación . . . 76
Especies invasoras . . . 76
Living Atlases . . . 76
Misceláneos . . . 76
OpenRefine . . . 77
Planificación/Colaboración . . . 77
Calidad . . . 77
Especies sensibles . . . 78
Taxonomía . . . 78
Glosario . . . 78
Apéndice: Artículos de datos . . . 81
Apéndice: Soluciones . . . 82
Soluciones de la revisión de las bases . . . 82

Soluciones de la revisión de planificación . . . 85
Soluciones de la revisión de captura de datos . . . 86
Soluciones de la revisión de gestión de datos . . . 87
Soluciones de la revisión de publicación de datos . . . 88
Solución sugerida del Caso de Uso I . . . 90
Agradecimientos . . . 90
Diseño del curso e instrucción . . . 90
Traductores . . . 90
Francés . . . 90
Portugués . . . 91
Español . . . 91
Recursos . . . 91
Apoyo en recursos . . . 91
Colofón . . . 91
Cita sugerida . . . 91
Colaboradores . . . 92
Licencia . . . 92
URI persistente . . . 92
Control del documento . . . 92

Descripción del curso

Este curso permitirá a los participantes planificar e implementar de manera eficaz los esfuerzos de movilización de datos de biodiversidad utilizando estándares aceptados por la comunidad. Su objetivo es aumentar el volumen, la riqueza y la calidad de los datos publicados a través de la red GBIF. Este curso se desarrolló por primera vez como parte del programa Información sobre biodiversidad para el desarrollo (BID) financiado por la Unión Europea.

Los temas incluyen:

- Gestión de proyectos
- Captura de datos
- Gestión de datos
- Publicación de datos

Este curso se compone de instrucciones en vídeo combinadas con cuestionarios y ejercicios prácticos. Cuando se ofrece como taller presencial o virtual, se fomenta el trabajo en grupo y la interacción social.

Audiencia

Este curso está diseñado para personas que trabajan como investigadores o técnicos en instituciones de investigación o políticas sobre biodiversidad. Las instrucciones proporcionadas son particularmente útiles para aquellos que tienen la necesidad o el deseo de planificar proyectos de movilización y/o movilizar información sobre biodiversidad para sus respectivas instituciones.

Prerequisitos

1. Curso de introducción a GBIF
2. Además, para aprovechar al máximo las actividades de este curso, los participantes deben poseer las siguientes habilidades y conocimientos:
    - Conocimientos básicos en el uso de computadores e Internet y, en particular, en el uso de hojas de cálculo.
    - Conocimientos básicos de geografía e informática de la biodiversidad: conceptos de geografía y cartografía, reglas básicas de taxonomía y nomenclatura.
    - Disposición de difundir los conocimientos aprendidos en el taller con los socios y colaboradores de su proyecto, adaptando los materiales de formación sobre movilización de datos sobre biodiversidad a contextos e idiomas específicos, manteniendo siempre su valor didáctico.
    - Se recomienda que los participantes tengan un buen dominio de inglés. Sin embargo se han realizado esfuerzos para proporcionar los materiales en otros idiomas, los vídeos estarán disponibles en inglés con subtítulos en otros idiomas.

Objetivos de aprendizaje

- Aprender los conceptos clave (fundamentos) en informática de la biodiversidad, en particular sobre la gestión de datos digitales sobre biodiversidad.

- Recibir una introducción al estándar Darwin Core y sus componentes.
- Entender las diferentes etapas para planificar un proyecto de movilización y cómo adaptar estas etapas a un proyecto específico.
- Evaluar una estrategia de movilización de datos para identificar posibles vacíos, deficiencias y dificultades.
- Desarrollar una estrategia de movilización de datos adaptada a un marco institucional determinado.
- Aprender a identificar los tipos de datos y la mejor manera de capturar la información relevante utilizando las buenas prácticas, los software existentes, las herramientas y las técnicas.
- Utilizar herramientas de software diseñadas para facilitar la captura de datos sobre biodiversidad para producir datos digitales a partir de fuentes analógicas.
- Aprender conceptos de calidad de datos y recibir una introducción a las herramientas utilizadas para estandarizar, validar y limpiar datos.
- Utilizar herramientas informáticas para evaluar la idoneidad de un conjunto de datos sobre biodiversidad
- Utilizar herramientas de software diseñadas para la limpieza de datos (de biodiversidad).
- Conocer el proceso de hacer que los datos de biodiversidad estén disponibles en línea de forma gratuita, también conocido como publicación de datos, utilizando la herramienta de publicación integrada de GBIF, IPT (Integrated Publishing Toolkit).
- Definir los tipos y subtipos de datos publicables (si los hay) para un conjunto de datos de biodiversidad.
- Utilice el IPT de GBIF para publicar conjuntos de datos de biodiversidad utilizando las extensiones apropiadas.
- Capacitar a otros en la planificación, captura, gestión y publicación de datos sobre biodiversidad.

Certificación

Una vez completado con éxito el curso y la evaluación de las tareas (por parte de los formadores y mentores), los participantes tienen la oportunidad de recibir una certificación oficial en forma de Open Badge. Ver Evaluación y certificación para más detalles.

Archivos para descargar

Todos los archivos del curso se pueden descargar desde esta página. O si lo prefiere, todos los archivos están enlazados individualmente a lo largo del curso, tal y como aparecen en el currículum. Los archivos de vídeo también están incrustados a lo largo del curso, y se reproducen desde YouTube. Los subtítulos están disponibles cuando se reproduce desde YouTube, para la mayoría de los vídeos. Si tiene dificultades para acceder a los vídeos incrustados, descargue los archivos mp4 para reproducirlos localmente en su computador.

Videos

Los videos están narrados en inglés. Los subtítulos no están disponibles para los videos que sean descargados.

Foundations1.zip (73.7 MB)

Foundations2.zip (90.2 MB)

Planning.zip (51.3 MB)

Capture.zip (63.1 MB)

Management.zip (30.2 MB)

Publishing.zip (77.9 MB)

Appendix-Data-Papers.zip (97.5 MB)

Datos del ejercicio

Este archivo comprimido (ZIP 37.7 MB) contiene los datos del ejercicio para todos los casos de uso.

Hojas de ejercicio

Este archivo comprimido (ZIP 1.4 MB) contiene las hojas de ejercicio para todos los casos de uso. Las hojas de ejercicios están escritas en inglés y deben completarse en inglés.

Bases



Este módulo incluye instrucciones para garantizar que todos los participantes estén al mismo nivel antes de profundizar en los temas de movilización de datos. Recibirá una introducción al lenguaje, la terminología y las definiciones de algunos de los conceptos, funciones y procesos básicos que pondrá en práctica durante el resto del curso. También recibirá una introducción a la calidad de los datos y aprenderá la importancia de la documentación. Por último, como parte de este módulo, se le pedirá que instale el software OpenRefine.

Terminología

Definiciones



En este video (12:02), revisará la terminología utilizada en este curso. Si no puede ver el vídeo incrustado, puede descarga localmente. (MP4 - 38.5 MB)

▶ https://www.youtube.com/watch?v=FZAF5Sy8Nsc *(YouTube video)*

Software



En este video (05:58), usted revisará ejemplos de los diferentes tipos de aplicaciones y software disponibles en el mundo de la informática para la movilización de la biodiversidad. Si no puede ver el vídeo incrustado, puede descargar localmente. (MP4 - 18.9 MB)

▶ https://www.youtube.com/watch?v=vYfDIgBBKXY *(YouTube video)*

Estructuras

En este video (13:10), usted revisará los tipos de datos y campos que contienen datos, las estructuras que ayudan a organizar y proteger los datos y lo que estos significan para la integridad y seguridad de sus datos. Si no puede ver el vídeo incrustado, puede descargarlo localmente. (MP4 - 38.8 MB)

▶ https://www.youtube.com/watch?v=msnVbZvIy2E *(YouTube video)*

Calidad de los datos

En este vídeo (12:26), se repasa la terminología utilizada en este curso. Si no puede ver el vídeo incrustado, puede descargalo localmente. (MP4 - 38,5 MB)

▶ https://www.youtube.com/watch?v=5o7TcS2K7Cw *(YouTube video)*

A continuación encontrará una lectura seleccionada de la guía de Arthur Chapman “Principles of data quality”. Puede encontrar el documento completo, las referencias y traducciones en GBIF.org.

Antes de que pueda tener lugar una discusión detallada sobre calidad de los datos y su aplicación a datos de presencia de especies, hay una serie de conceptos que necesitan ser definidos y descritos. Estos incluyen el término calidad de los datos en sí, los términos exactitud y precisión que a menudo son mal aplicados, y lo que entendemos por datos primarios de especies y datos de presencia de especies.

Datos de presencia de especies

El término datos de presencia de especies se utiliza aquí para incluir los datos de las etiquetas de los especímenes o los lotes que se encuentran en los museos y herbarios, los datos de observaciones y de estudios ambientales. En general, los datos son lo que denominamos "basados en puntos", aunque también se incluyen los datos lineales (datos de transectos de estudios ambientales, colectas a lo largo de un río), poligonales (observaciones dentro de un área definida, como un parque nacional) y cuadriculados (observaciones o registros de encuestas de una cuadrícula regular). En general, se trata de datos georreferenciados, es decir, registros con referencias geográficas que los vinculan a un lugar concreto en el espacio, ya sea con una coordenada georreferenciada (por ejemplo, latitud y longitud, UTM) o no (descripción textual de una localidad, altitud, profundidad), y tiempo (fecha, hora del día).

En general, los datos también están vinculados a un nombre taxonómico, pero las colectas no identificadas también pueden ser incluidas. De vez

en cuando, el término se ha utilizado de forma intercambiable con el término “datos primarios de especies”.

Datos primarios de especies

El término "datos primarios de especies" se utiliza para describir los datos crudos de colecta y los datos sin atributos espaciales. Incluye datos taxonómicos y de nomenclatura sin atributos espaciales, como nombres, taxones y conceptos taxonómicos sin referencias geográficas asociadas.

Exactitud y precisión

La exactitud y la precisión se confunden a menudo y generalmente las diferencias no se entienden.

Exactitud se refiere a la cercanía de los valores medidos, de las observaciones o de las estimaciones al valor real o verdadero (o con un valor que se acepta como verdadero, por ejemplo, las coordenadas de un punto de control).

La precisión (o resolución) se puede dividir en dos tipos principales. La precisión estadística es la cercanía con la que las observaciones repetidas se ajustan a sí mismas. No tienen nada que ver con su relación con el valor real y pueden tener alta precisión, pero poca exactitud. La precisión numérica es la cantidad de dígitos significativos en los que se registra una observación y se ha vuelto mucho más obvia con la llegada de las computadoras. Por ejemplo, una base de datos puede mostrar un registro de latitud/longitud con 10 decimales, es decir, unos 0.01 mm, cuando en realidad el registro tiene una resolución no superior a 10-100 m (3-4 decimales). Esto a menudo conduce a una falsa impresión tanto de la resolución como de la precisión.

Estos términos (exactitud y precisión) pueden aplicarse tanto a los datos no espaciales como a los espaciales. Por ejemplo, una colecta puede tener una identificación a nivel de subespecie (es decir, tener alta precisión), pero ser del taxón incorrecto (es decir, tener baja exactitud) o estar identificada solo a nivel de Familia (alta exactitud, pero baja precisión).

Calidad de los datos

La calidad de los datos es multidimensional e implica la gestión, el modelado y el análisis de los datos, así como el control de calidad y de seguridad, el almacenamiento y la presentación. Como se indica de forma independiente por Chrisman (1991) y Strong et al. (1997), la calidad

de los datos está relacionada con el uso y no puede evaluarse independientemente del usuario. En una base de datos, los datos no tienen calidad ni valor real (Dalcin 2004); sólo tienen un valor potencial que se realiza sólo cuando alguien utiliza los datos para hacer algo útil. La calidad de la información está relacionada con su capacidad de satisfacer a sus clientes y de responder a sus necesidades (English 1999).

Redman (2001), sugiere que para que los datos sean aptos para el uso deben ser accesibles, precisos, oportunos, completos, coherentes con otras fuentes, pertinentes, exhaustivos, proporcionar un nivel de detalle adecuado, ser fáciles de leer y fáciles de interpretar.

Un tema que un custodio de datos necesita considerar es lo que puede ser necesario hacer con la base de datos para aumentar su usabilidad a un público más amplio (es decir, aumentar su potencial uso o relevancia) y, por lo tanto, hacerla apta para una gama más amplia de propósitos. En este caso, habrá un equilibrio entre el aumento de la usabilidad y la cantidad de esfuerzo necesario para añadir funcionalidad y usabilidad adicionales. Esto puede requerir cosas como segmentar los campos de datos, añadir información de georreferenciación, etc.

Garantía de calidad/Control de calidad

La diferencia entre control de calidad y garantía de calidad no siempre es clara. Taulbee (1996) hace la distinción entre "control de calidad" y "garantía de calidad" y enfatiza que uno no puede existir sin el otro si se quieren alcanzar los objetivos de calidad. Ella define el "control de calidad" como un juicio sobre la calidad basado en estándares, procesos y procedimientos internos establecidos para controlar y monitorear la calidad; y "garantía de calidad" como un juicio sobre la calidad basado en estándares externos al proceso y es la revisión de las actividades y los procesos de control de calidad para asegurar que los productos finales cumplan con los estándares de calidad predeterminados.

En un enfoque más orientado a los negocios, Redman (2001) define la garantía de calidad como “aquellas actividades que están diseñadas para producir productos de información libres de defectos para satisfacer las necesidades más importantes de los clientes más importantes, al menor costo posible”.

No está claro cómo deben aplicarse estos términos en la práctica y, en la mayoría de los casos, parecen utilizarse como sinónimos para describir la práctica general de la gestión de la calidad de los datos.

Incertidumbre

La incertidumbre puede considerarse como una "medida del grado de incompletitud del conocimiento o de la información sobre una cantidad desconocida cuyo valor real podría establecerse si se dispusiera de un dispositivo de medición perfecto" (Cullen y Frey 1999). La incertidumbre es una propiedad de la comprensión de los datos por parte del observador, y se trata más del observador que de los datos en sí. Siempre hay incertidumbre en los datos; la dificultad está en registrar, comprender y visualizar esa incertidumbre para que otros también puedan comprenderla. La incertidumbre es un término clave para comprender el riesgo y la evaluación del riesgo.

Error

El error abarca tanto la imprecisión de los datos como su inexactitud. Hay muchos factores que contribuyen al error. En general, se considera que el error es aleatorio o sistemático. El error aleatorio suele referirse a la desviación del estado real de manera aleatoria. El error o sesgo sistemático surge de un cambio uniforme en los valores y a veces se describe como de "exactitud relativa" en el mundo cartográfico (Chrisman 1991). Al determinar la "aptitud para el uso", el error sistemático puede ser aceptable para algunas aplicaciones y no apto para otras.

Un ejemplo puede ser el uso de un datum geodésico diferente, que, si se utiliza en todo el análisis, puede no causar ningún problema importante. Sin embargo, los problemas surgirán cuando un análisis utilice datos de diferentes fuentes y con diferentes sesgos, por ejemplo, fuentes de datos que utilicen diferentes datums geodésicos, o cuando las identificaciones se hayan realizado utilizando una versión anterior de un código de nomenclatura.

"Dado que el error es ineludible, debe reconocerse como una dimensión fundamental de los datos" (Chrisman 1991). Sólo cuando el error se incluye en una representación de los datos es posible responder a preguntas sobre las limitaciones de los datos e incluso sobre las limitaciones del conocimiento actual. Los errores conocidos en las tres dimensiones de espacio, atributo y tiempo deben medirse, calcularse, registrarse y documentarse.

Validación y limpieza

La validación es un proceso que se utiliza para determinar si los datos son inexactos, incompletos o irrazonables. El proceso puede incluir verificaciones de formato, verificaciones de integridad, verificaciones de razonabilidad, verificaciones de límites, revisión de los datos para

identificar valores atípicos (geográficos, estadísticos, temporales o ambientales) u otros errores, y evaluación de datos por expertos en el área temática (por ejemplo, especialistas en taxonomía). Estos procesos suelen dar lugar a la señalización, documentación y posterior comprobación de los registros sospechosos. Los controles de validación también pueden implicar la comprobación del cumplimiento de los estándares, reglas y convenciones aplicables. Una etapa clave en la validación y limpieza de datos es identificar las causas fundamentales de los errores detectados y centrarse en evitar que esos errores vuelvan a producirse (Redman 2001).

La limpieza de datos se refiere al proceso de "corregir" errores en los datos que se han identificado durante el proceso de validación. El término es sinónimo de "limpieza de datos", aunque algunos utilizan la limpieza de datos para abarcar tanto la validación como la limpieza de datos. Es importante que en el proceso de limpieza de datos no se pierdan datos por descuido y que los cambios en la información existente se realicen con mucho cuidado. A menudo es mejor conservar tanto los datos antiguos (datos originales) como los nuevos (datos corregidos) uno al lado del otro en la base de datos para que, si se cometen errores en el proceso de limpieza, se pueda recuperar la información original.

Documentación

En este video (09:47), proporcionaremos una visión general de la importancia de la documentación en relación con la gestión de datos y la publicación de datos. Aprenderá sobre mapeo de datos, relaciones de datos y metadatos. Si no puede ver el vídeo incrustado, puede descargalo localmente. (MP4 - 29.2 MB)

▶ https://www.youtube.com/watch?v=Z5-SYImGRGc *(YouTube video)*

Flujos de trabajo de digitalización

Este video (07:20) sobre flujos de trabajo de digitalización identifica cinco grupos (o etapas) en el proceso de digitalización de objetos de colección de historia natural mediante imágenes digitales, y estas etapas se pueden adaptar fácilmente a otras fuentes de datos de biodiversidad. Si no puede ver el vídeo incrustado, puede descargalo localmente. (MP4 - 26.8 MB)

▶ https://vimeo.com/120369455 *(Vimeo video)*

Como destaca el vídeo, los protocolos de digitalización varían de una institución a otra, pero es esencial que el protocolo elegido sea acordado, documentado y respetado.

No profundizaremos en la digitalización, per se, durante el taller, ya que puede ser un curso de una semana de duración por sí solo, en su lugar, nos centramos en la introducción básica a la captura de datos sobre biodiversidad. Sin embargo, queremos proporcionarle recursos en materia de digitalización, ya que sabemos que muchos están interesados en ello.

Hay muchas maneras de organizar los esfuerzos de digitalización y, por lo tanto, la digitalización puede parecer desalentadora al principio. Es importante recordar que en la mayoría de los casos alguien ya ha intentado digitalizar los mismos tipos de especímenes y objetos que usted está planeando. En este ejercicio le presentamos algunos recursos prácticos del flujo de trabajo de la digitalización para ayudarle a empezar. Estas también constituirán la base para el trabajo que realizaremos en el taller sobre la selección, modificación y evaluación de los flujos de trabajo.

Algunos pasos en el proceso pueden incluir:

- **Curación y etapa de pre-digitalización**: comprende la preparación de la fuente de datos para el proceso de digitalización, incluyendo la asignación de identificadores únicos que ayuden a referirse a la fuente sin errores y a mantener junta toda la información derivada.
- **Captura de imagen**: incluye una gran cantidad de planificación, no sólo sobre la captura de imágenes en sí (por ejemplo, la definición de la secuencia de trabajo, la selección del hardware adecuado), sino también sobre cómo y dónde se almacenarán y manejarán las imágenes.
- **Procesamiento de imagen**: incluye el control de calidad, conversión de archivos, etc.
- **Captura electrónica de datos**: es el núcleo del proceso de digitalización, incluye la captura de información clave en una base de datos. El vídeo resalta que el método más común de ingresar la información es a través de un teclado, pero cada vez son más las instituciones que recurren a tecnologías avanzadas de entrada de datos.
- **Georreferenciación**: La información geográfica es muy importante para el análisis de la biodiversidad, por lo que los proyectos de digitalización deben intentar extraer la información geográfica más exacta posible.

Integrated Digitized Biocollections (iDigBio) es el centro de coordinación del United States National Resource for Advancing Digitization of Biodiversity Collections (ADBC). Lideran un esfuerzo nacional para poner a disposición datos e imágenes de millones de especímenes biológicos en un formato electrónico estándar para la comunidad investigadora, organismos gubernamentales, estudiantes, educadores y el público en general. Han producido varios videos que discuten el proceso de digitalización.

Hay otros vídeos en la serie iDigBio en los que puede estar interesado si desea aprender más sobre flujos de trabajo específicos para diferentes tipos de especímenes:

- “Digitalización de colecciones húmedas” (4:34 mins) https://vimeo.com/120369690
- “Flujos de trabajo en proceso de digitalización de especímenes vertebrados conservados en seco” (7:25 mins) https://vimeo.com/160615629
- “Digitalización de especímenes de herbarios” (7:34 mins) https://vimeo.com/120369768

Herramientas de software



Revisión de herramientas de software utilizadas en informática de la biodiversidad

Durante las actividades del curso, demostraremos y trabajaremos con muchas herramientas de software relacionadas con la digitalización, la calidad y la transformación de los datos. Probablemente ya utilice varias de ellas en su trabajo diario.

Los instructores, mentores y antiguos participantes de los cursos han compilado una lista con información sobre herramientas informáticas de biodiversidad. En ella se ofrecen enlaces a sus principales sitios web, datos clave y un resumen de las ventajas y desventajas.

Enlace de descarga Software-database-EN.xlsx. (23 KB)

Cuando evalúe un software de biodiversidad que no haya utilizado antes, debe considerar cómo lo adaptaría a sus fines. A continuación encontrará una lista con la que puede comenzar su evaluación. Están inspiradas en el capítulo "características de una buena solución de base de datos" del manual de GBIF "Iniciar un proyecto de digitalización":

- **Precio**: uno de los factores más determinantes. Hay que tener cuidado con otros costos más allá del precio de la licencia del software, como el hardware necesario para ejecutarlo, el mantenimiento, las actualizaciones y la experiencia para ejecutarlo.
- **Funcionalidad**: hay que tener claro qué se espera que consiga el software y asegurarse de que lo hace con eficacia. No se distraiga con funcionalidades adicionales que puedan hacer el software innecesariamente más complejo.
- **Estabilidad**: algunas soluciones llevan mucho tiempo en el mercado y cuentan con el apoyo de instituciones o empresas sólidas, por lo que es más probable que no tengan errores o que dispongan de buenos sistemas para resolver cualquier problema que surja. También será más probable que se actualicen y porten a sistemas operativos más modernos.
- **Escalabilidad**: Algunos software funcionan muy bien cuando se les hace una prueba, pero su rendimiento se degrada al cabo de un tiempo o cuando se utilizan con mayores cantidades de datos, también cuando varios usuarios acceden a ellos simultáneamente. Consulte las opiniones de otros usuarios en Internet.
- **Integración**: asegúrese de que el software acepta y produce los formatos de datos que utiliza y necesita. La transformación de datos es una tarea que consume mucho tiempo.
- **Soporte de idioma**: es esencial que todos los usuarios del software puedan entender su interfaz y la documentación que hará posible su uso.
- **Soporte técnico y documentación**: asegúrese de explorar la documentación y los mecanismos de apoyo existentes. Puede estar seguro de que en algún momento la necesitará.
- **Curva de aprendizaje**: algunos software pueden requerir capacitación específica para aprender a usarlo, mientras que otros son más intuitivos y pueden aprenderse mientras se usan con el apoyo de sistemas de ayuda en línea.

Instalar OpenRefine

Instalar el software necesario para las actividades posteriores del curso

OpenRefine es una herramienta con un conjunto de características para trabajar con datos tabulares que mejora la calidad general de un conjunto de datos. Es una aplicación que se ejecuta en su propio computador como un pequeño servidor web y para poder usarlo su navegador debe dirigirse a ese servidor web. Por lo tanto, considere OpenRefine como una aplicación web personal y privada.

Utilizaremos OpenRefine durante la parte de movilización de datos del curso, especialmente durante los ejercicios prácticos. Será necesario instalar OpenRefine en su computador. Si tiene conocimientos de informática, puede seguir estos pasos para instalar el software. Si no se siente seguro, pida ayuda. Consulte la página página de descarga de OpenRefine para más detalles.



Es posible que se necesiten contraseñas administrativas para instalar el software.

Requisitos para la instalación

1. Sólo para usuarios de Linux: Java JRE instalado.
2. Google Chrome, Microsoft Edge o Mozilla Firefox instalados. Internet Explorer no es compatible.

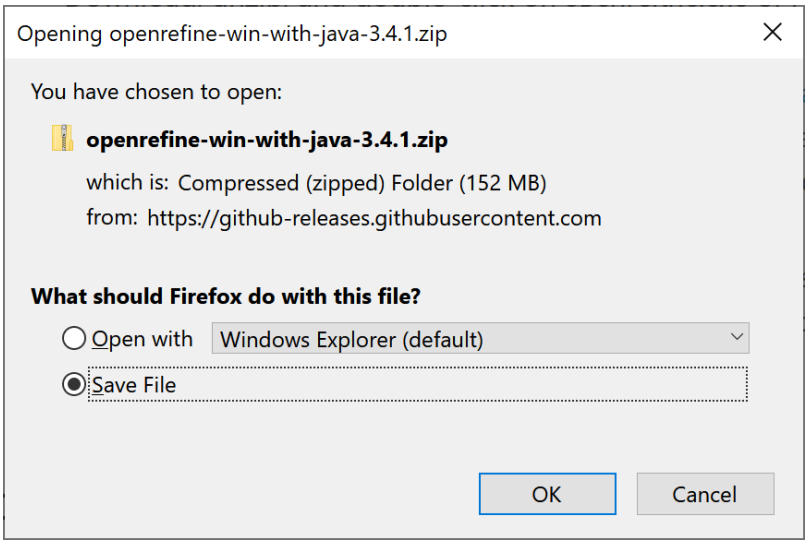


La última versión estable es OpenRefine 3.5.1, publicada el 19 de diciembre de 2021. Las instrucciones detalladas de instalación están disponibles en https://docs.openrefine.org/manual/installing/.

Instalación en MS Windows

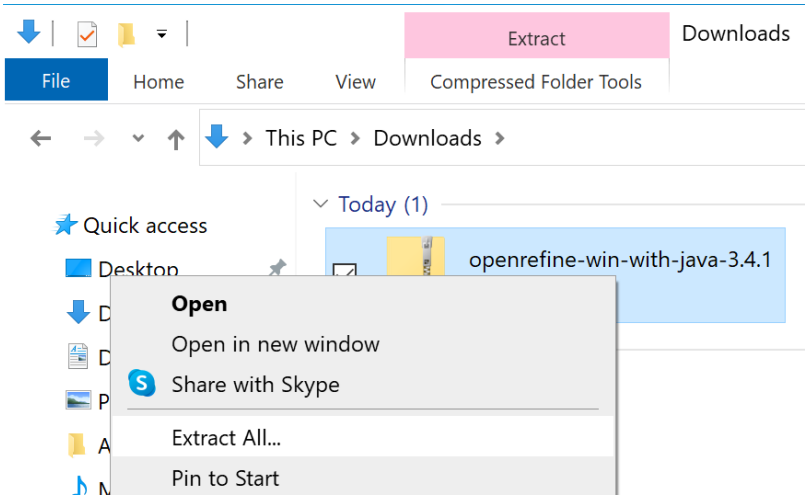
1. Descargue Windows kit **with embedded Java**. Elija guardar el archivo en lugar de abrirlo.
2. Encuentre el archivo descargado. Haga clic derecho en él y elija "Extraer todo…". Descomprima y haga doble clic en openrefine.exe o refine.bat si la primera no funciona.
3. Aparecerá una ventana de comandos (no la cierre) y poco después, una nueva ventana del navegador web mostrará la aplicación.

▼ *Instrucciones detalladas para MS Windows (haga clic para expandir)*

1. 

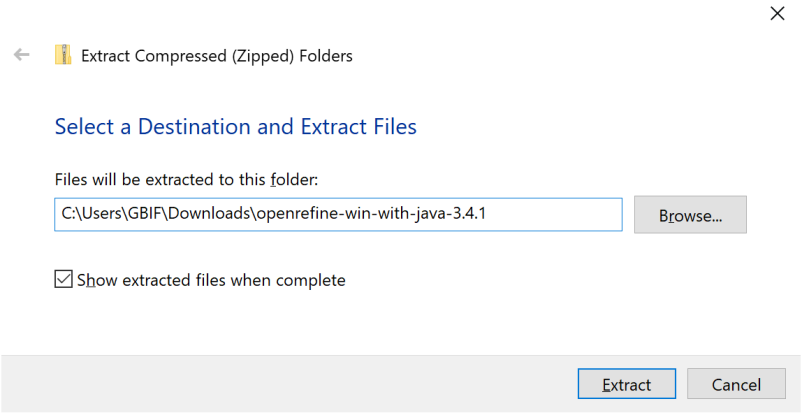
   Descargue Windows kit **with embedded Java**. Elija guardar el archivo en lugar de abrirlo.

2.



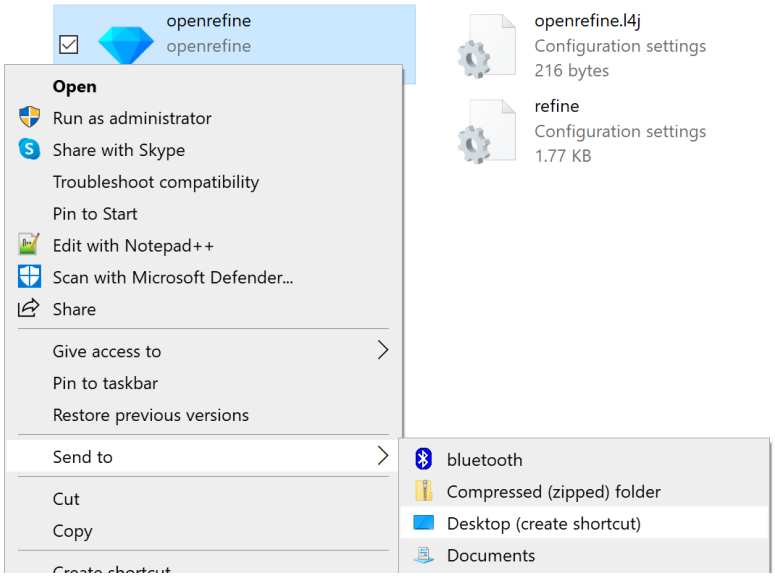
Encuentre el archivo descargado. Haga clic derecho en él y elija "Extraer todos..."

3.



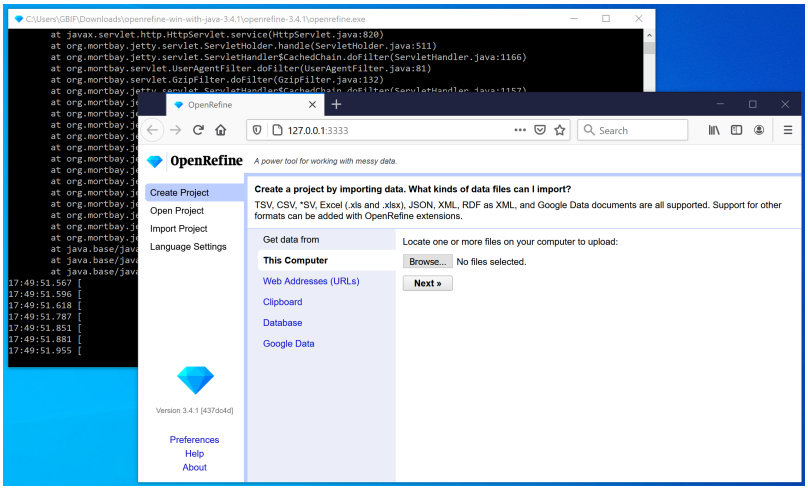
Haga clic en "Extraer"

4.



Busque los archivos extraídos. Opcionalmente, haga clic derecho en "openrefine" y elija "Enviar a → Escritorio (crear acceso directo)" para crear un acceso directo en su computador. Luego haga doble clic en "openrefine"

5.



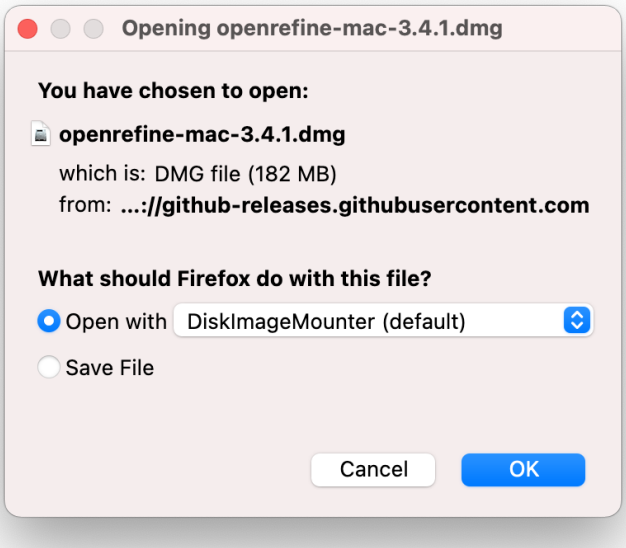
Se abre una ventana de consola negra y poco después se abre el navegador. OpenRefine ya está listo para usarse.

Instalación en Mac

1. Descargue el Mac kit.
2. Descargue, abra y arrastre el icono a la carpeta de aplicaciones. No es necesario instalar Java por separado.
3. Haga doble clic en él y una nueva ventana del navegador mostrará la aplicación.

▼ *Instrucciones detalladas para Mac (haga clic para expandir)*

1.



Descargue el Mac kit y elija abrirlo.

2.

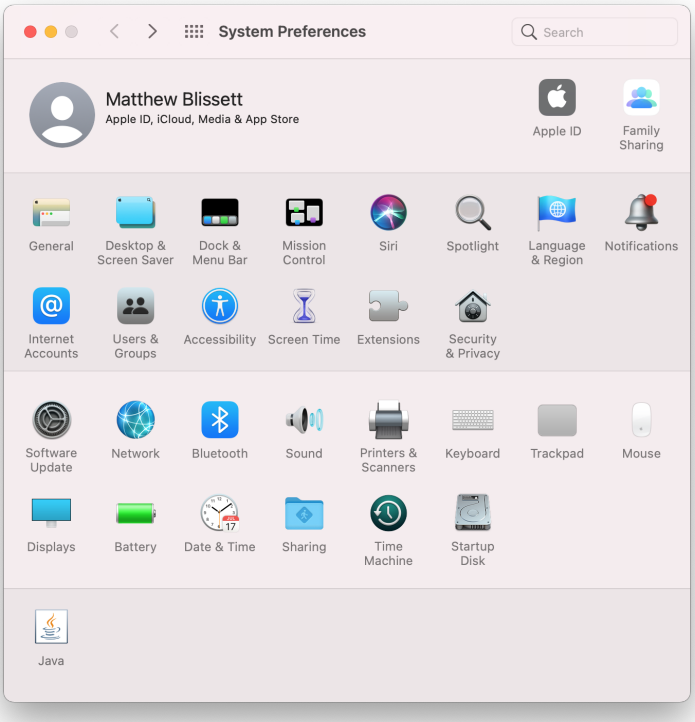


Se muestra una advertencia. Haga clic en "OK".

3. 

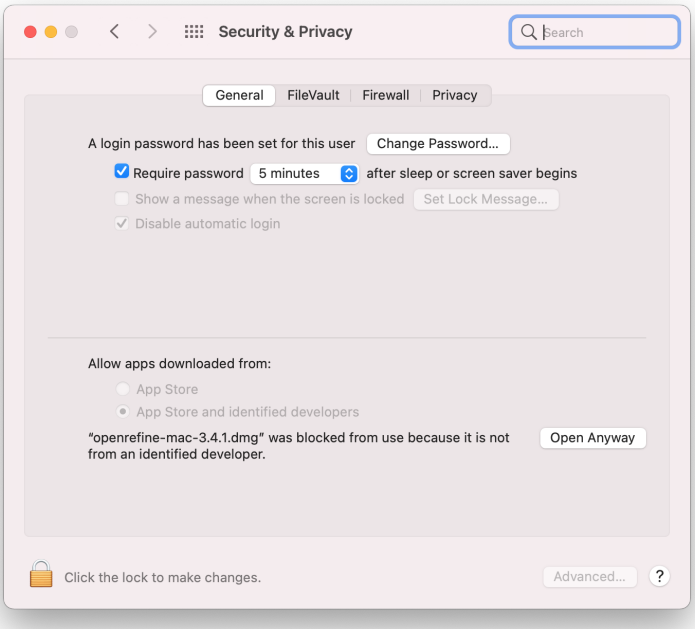
Abra las Preferencias del Sistema.

4.



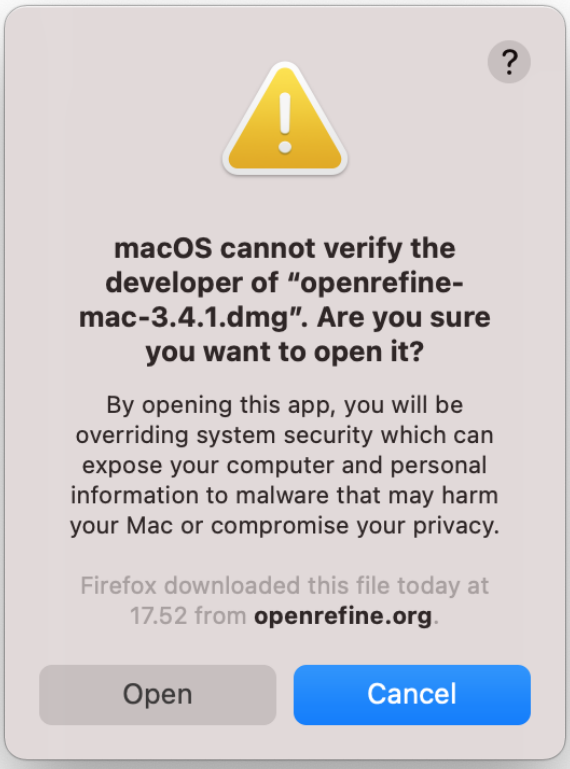
Abrir Seguridad y Privacidad

5.

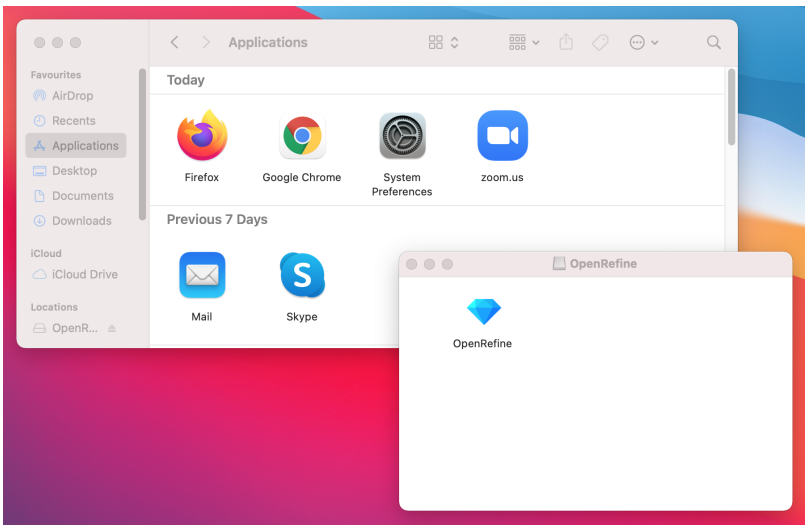


Elija "Abrir de todas formas" en la parte inferior.

6.

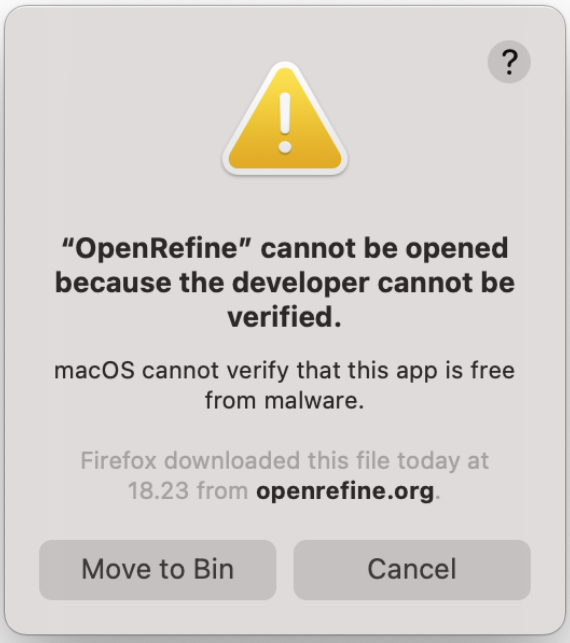


Elija "Abrir"

7. 

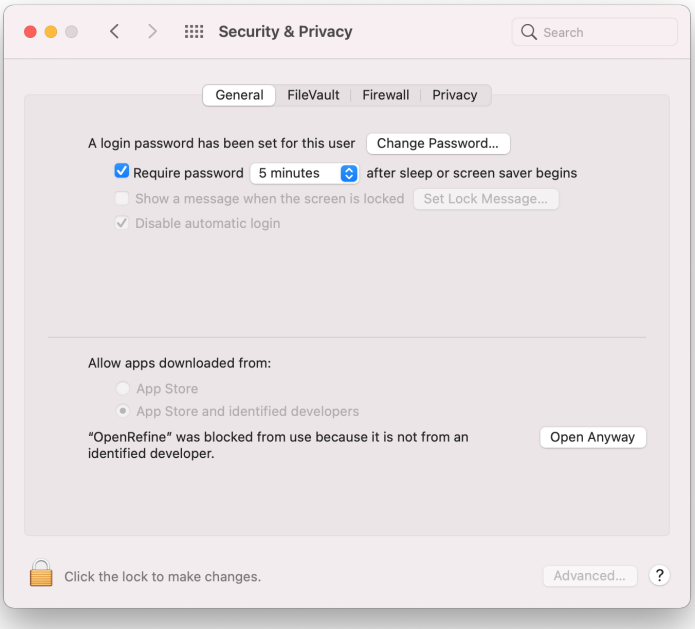
¡Finalmente se abre el archivo de la aplicación! Arrástrelo a su carpeta de Aplicaciones.

8.



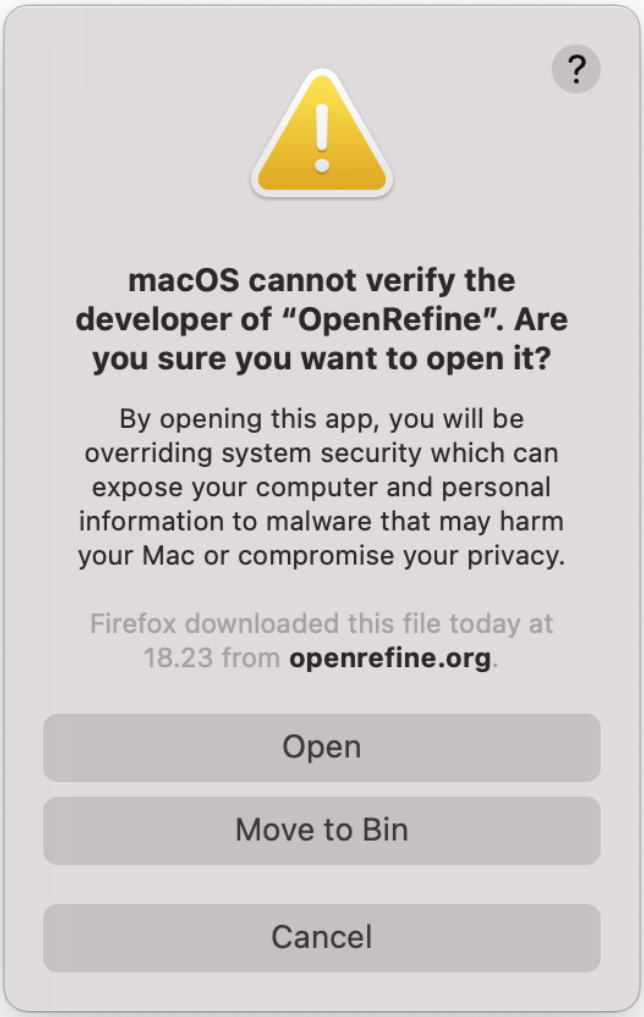
Haga doble clic en el icono de OpenRefine. ¡Aparece otra advertencia de seguridad!

9.



Vuelva a "Seguridad y Privacidad" y haga clic en "Abrir de todas formas".

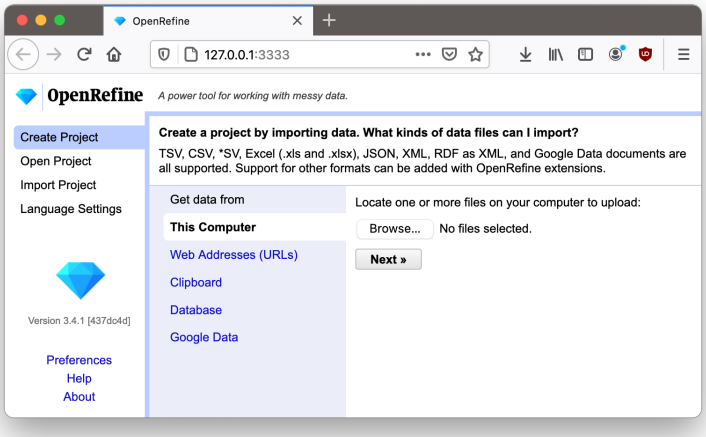
10.



(Para evitar estas advertencias, los desarrolladores de OpenRefine tendrían que pagar a Apple.)

Haga clic en "Abrir".

11.



¡Finalmente! La aplicación se está ejecutando.

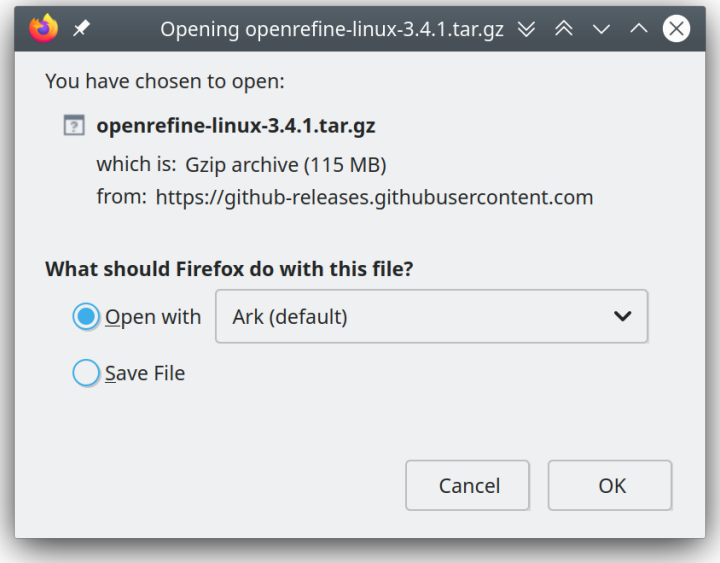
Instalación en Linux

1. Descargue el Linux kit.
2. Descargue, extraiga y escriba ./refine para empezar. Esto requiere que Java esté instalado en su ordenador.

▼ *Instrucciones detalladas para Linux (haga clic para expandir)*

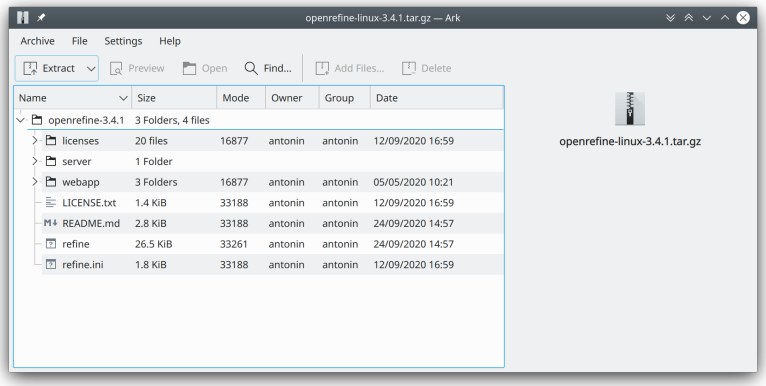
Estas instrucciones son para KDE (por ejemplo, Kubuntu, SuSE), pero el proceso es similar para Gnome (por ejemplo, Ubuntu, Red Hat, CentOS).

1.



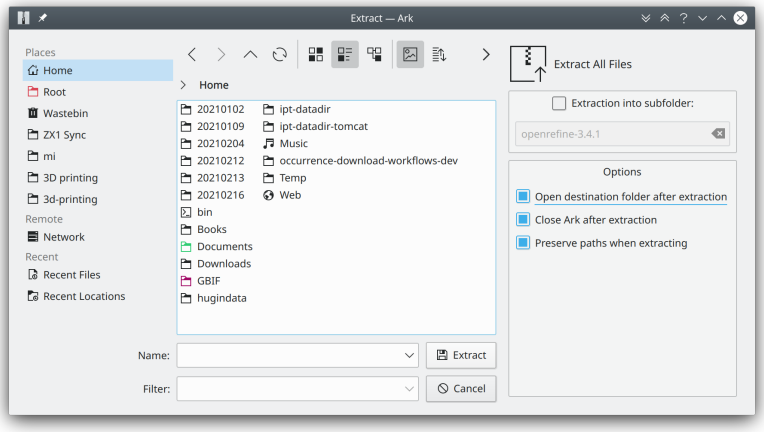
Descargue el Linux kit. Ábralo.

2.

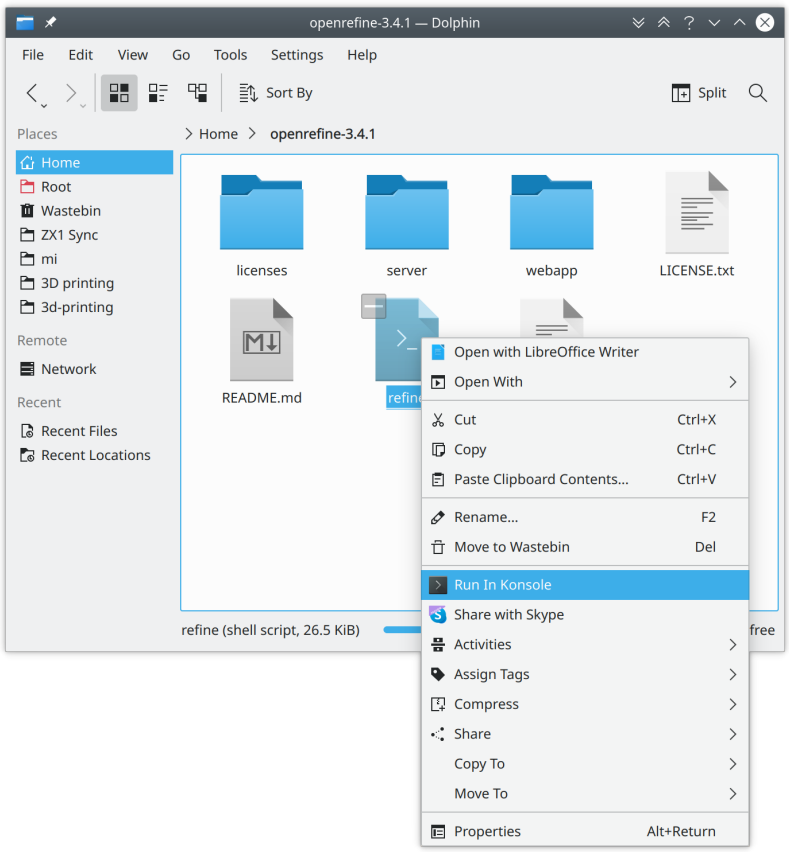


Haga clic en "Extraer" para descomprimir la aplicación descargada.

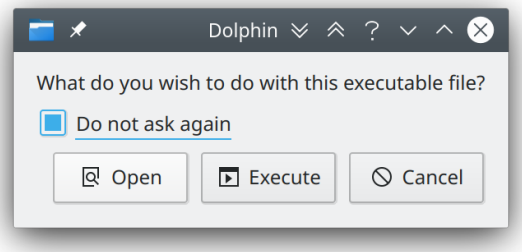
3.



Elija un lugar adecuado. También seleccioné "Abrir carpeta de destino después de la extracción" y "Cerrar el Ark después de la extracción"

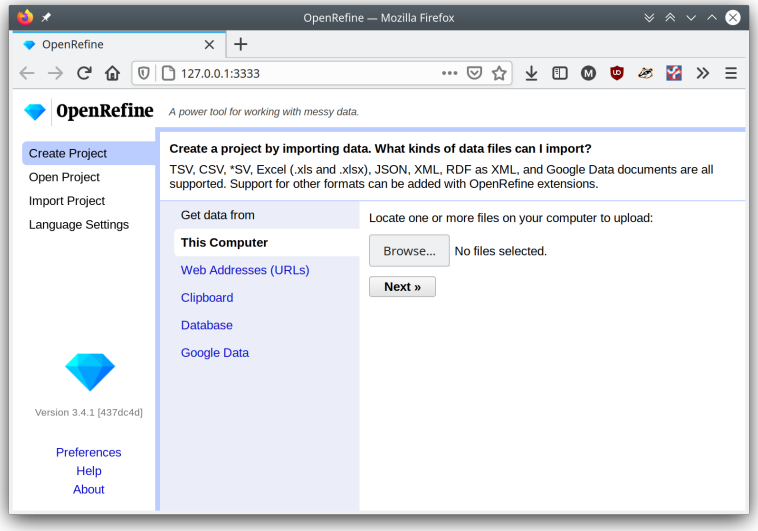
4. 

Haga clic derecho en "refine" y elija "Ejecutar en Konsole". Esto es necesario para que pueda salir de OpenRefine de forma segura más tarde, cerrando la ventana de Konsole.

5. 

Confirme que desea ejecutar la aplicación descargada.

6.



OpenRefine se está ejecutando.

Revisión de las bases

Comprueba los conceptos aprendidos en esta sección.

1. Para la declaración dada, ingrese el término correcto (base de datos, lenguaje de programación de la base de datos, programa de la base de datos)
   - combina y presenta funciones y características para manipular datos, juntas en una interfaz unificada
     _____
   - recopilación estructurada y organizada de datos y/o información almacenada en una computadora
     _____
   - la forma en que un humano se comunica con una computadora
     _____
2. Si abre un archivo de datos y ve lo siguiente, ¿cuál sospecharía que es el problema?+

   `□tre, ou ne pas □tre, c□est l□ la question.`

   - ☐ Nada
   - ☐ Está corrupto
   - ☐ Se ha utilizado una codificación incorrecta para abrir el archivo
   - ☐ El remitente usó una fuente extraña
3. Para el software dado, ingrese el tipo de software (captura de datos, gestión de datos, limpieza de datos, publicación de datos).
   - Herramienta de publicación integrada (IPT)
     _____
   - Specify
     _____ Y _____

- iNaturalist
  _____
- OpenRefine
  _____

4. Para el ejemplo dado, introduzca el tipo de datos correcto (binario, booleano, flotante, entero, entero largo, texto, texto no estructurado)

   - 1236975
     _____
   - 01101111
     _____
   - Caminamos 5 millas a lo largo de la carretera oeste desde la oficina de correos en el centro de la ciudad. Luego fuimos 2 millas al norte por un camino de tierra hacia el río, luego continuamos hacia el oeste a lo largo del río por otras 5 millas.
     _____
   - 1024
     _____
   - 29.
     _____
   - Sí/No
     _____
   - 6 conejos observados
     _____

5. ¿Cuál de estos términos describe un "nombre de campo/columna"?

   - ☐ Asignado
   - ☐ Descriptivo
   - ☐ Identificador
   - ☐ Legible
   - ☐ Único
   - ☐ interfaz de usuario

6. ¿Cuál de estos términos describe una "etiqueta de campo"?

   - ☐ Asignado
   - ☐ Descriptivo
   - ☐ Identificador
   - ☐ Legible
   - ☐ Único
   - ☐ interfaz de usuario

7. Para cada afirmación, introduzca la estructura correcta (fila, columna, tabla)

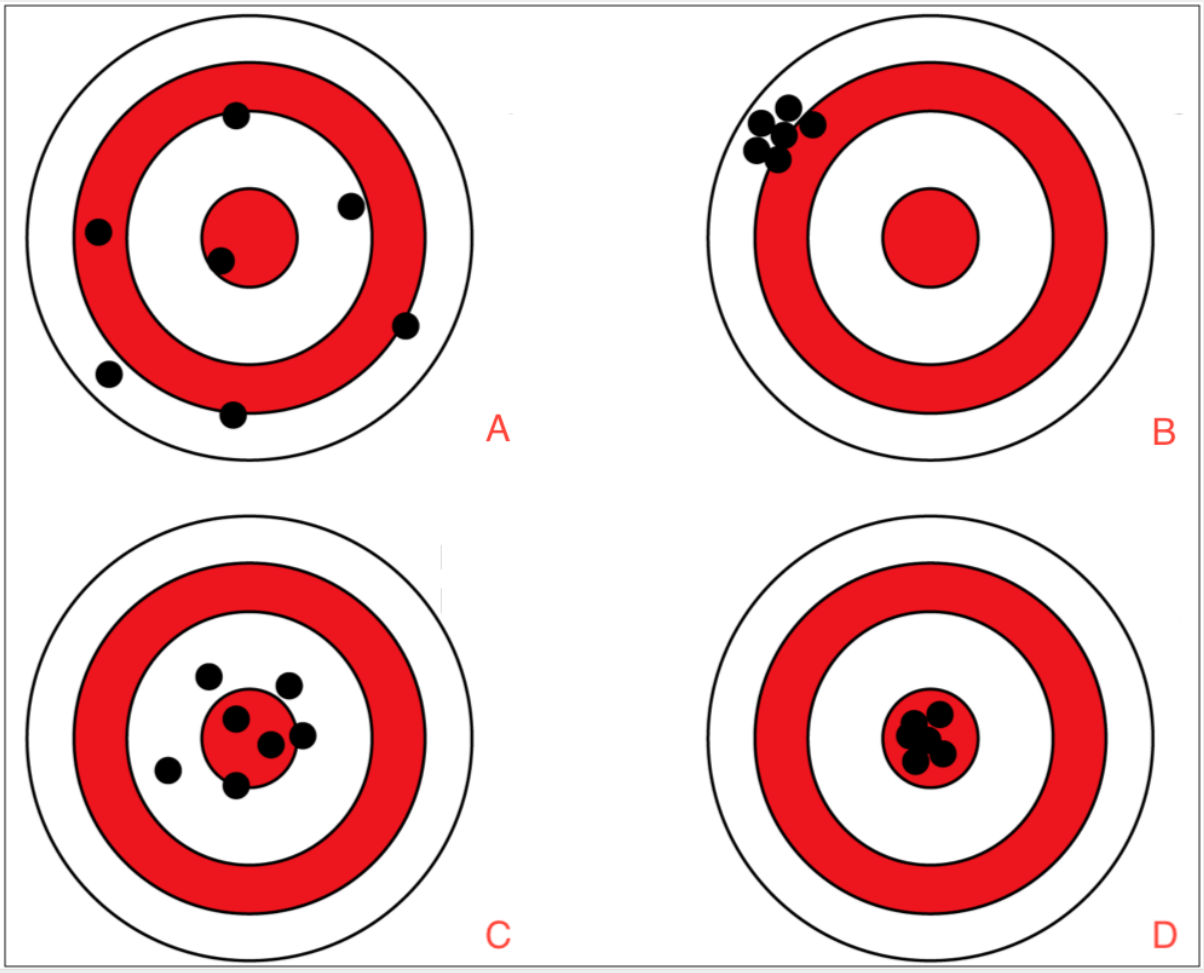
   - Todos los datos se refieren a un concepto UNICO.
     _____

- Un atributo tiene el MISMO campo / tipo de datos para cada registro.
  _____
- Los atributos de un registro SIEMPRE permanecen juntos.
  _____

8. ¿Quién determina la aptitud para el uso ("fitness for use") de sus datos?

   - ☐ El director del museo o departamento
   - ☐ Los usuarios de los datos para investigación o educación
   - ☐ El colector de los datos en el campo
   - ☐ La persona que introduce los datos en la base de datos

9. Para las sentencias dadas, introduzca la imagen coincidente. (A, B, C, D)



   - Alta exactitud, baja precisión
     _____
   - Baja exactitud, alta precisión
     _____
   - Alta exactitud, alta precisión
     _____
   - Baja exactitud, baja precisión
     _____

10. Identificar las relaciones de datos donde el conjunto de datos B necesita ser fusionado en el conjunto de datos A (0:1, 1:0, 1:1, 1:∞, ∞:1, ∞:∞). No se utilizan todas las relaciones.

- El campo del colector existe en ambos conjuntos de datos A y B

  _____

- El campo del país sólo existe en el conjunto de datos B

  _____

- El campo del nombre existe en el conjunto de datos A, pero el conjunto de datos B contiene los campos de Nombre y Apellidos

  _____

- El campo ID existe en ambos conjuntos de datos A y B

  _____

- La elevación existe en el conjunto de datos A, pero no en el conjunto de datos B

  _____

- La fecha existe en el conjunto de datos A, pero el Día, Mes y Año son campos separados en el conjunto de datos B

  _____

11. Los metadatos son importantes porque (seleccione los enunciados VERDADEROS):

- ☐ permiten a los usuarios determinar si un conjunto de datos es adecuado para su uso.
- ☐ le permiten compartir coordenadas exactas para cada registro.
- ☐ le permiten saber bajo qué términos legales se permite la reutilización de datos.
- ☐ también se aplica a todos los materiales complementarios y asociados, incluidas imágenes, videos y otros medios.
- ☐ le permiten conocer el próximo horario de exhibición/apertura de la institución.

Caso de uso I: especímenes de herbario

Familiarícese con el escenario del caso de uso.

El caso de uso I es un caso de uso práctico para los módulos de planificación, captura, gestión y publicación de datos. Se recomienda que descargue hoja de ejercicios (MS Word 345 KB) para que pueda tomar notas mientras trabaja en los ejercicios. Se proporcionará una solución sugerida en el apéndice de soluciones. El caso de uso I no está calificado.

Escenario

Un proyecto de movilización de datos en un herbario regional



Eriocaulon bilobatum Morong *colectado en Guatemala por Rapid Reference Collection (RRRC) | Field Museum of Natural History - Keller Action Science Center (licenciado bajo* CC-BY-NC 4.0*)*

Esta narrativa fue desarrollada como base para ejercicios prácticos en el curso de movilización de datos de biodiversidad y el concepto y contenido del ejercicio fue desarrollado por Alberto González-Talaván, Néstor Beltrán, Nicolas Noé, Sharon Grant. Los datos provienen de un conjunto de datos real, pero se han modificado para los fines de estos ejercicios. Es un escenario ficticio y solo tiene fines educativos.

Descripción

La Universidad de White Plains es una institución de educación terciaria reconocida en Guatemala y una referencia nacional para la investigación de la biodiversidad. El Departamento de Biología Vegetal mantiene un herbario de tamaño mediano que contiene aproximadamente 80,000 especímenes recolectados en y alrededor de la región local y que data de mediados del siglo XX hasta el presente. La colección incluye especímenes importantes, incluidos tipos y endémicos.

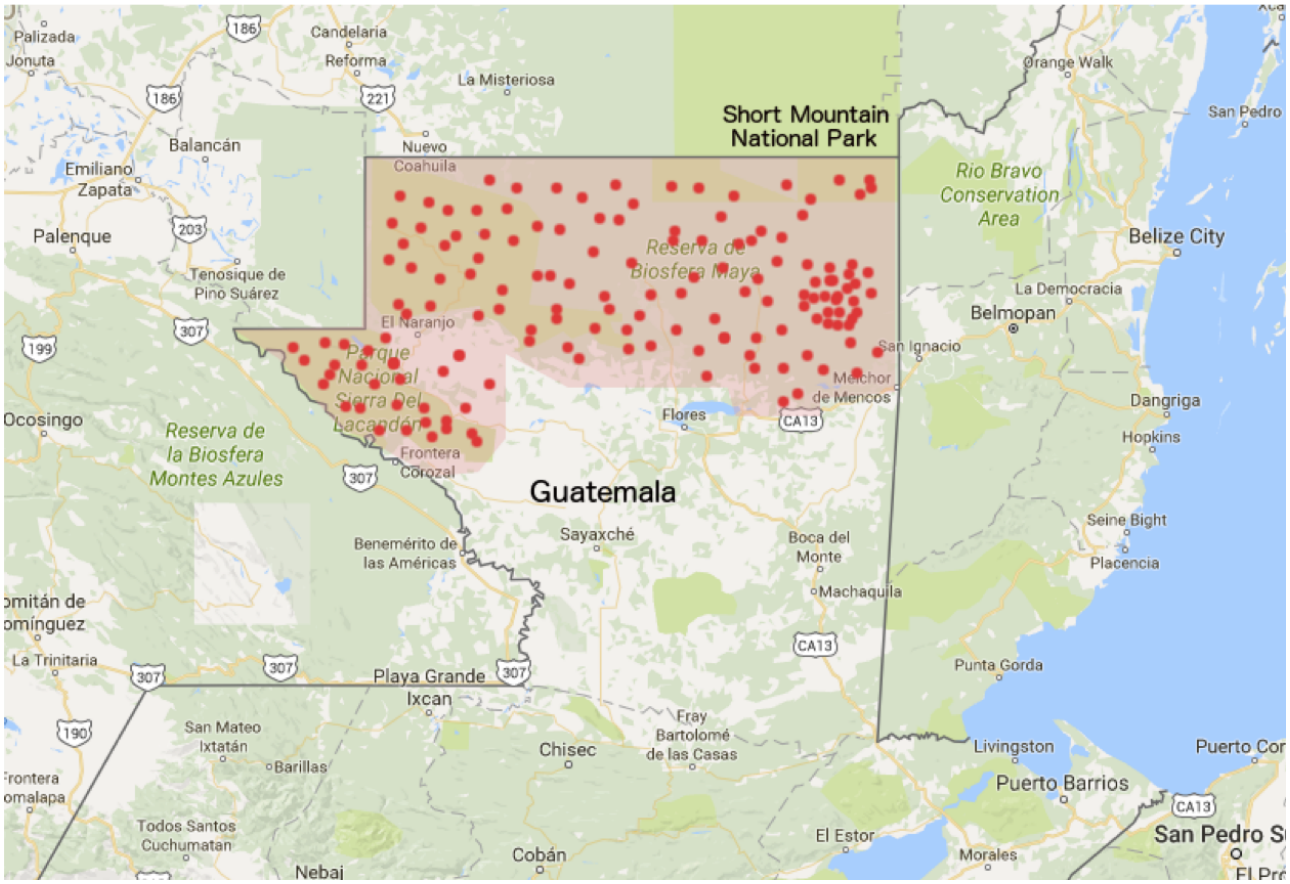
Actualmente, el cuidado de la colección está a cargo del profesor de Sistemática Vegetal, quien realiza tareas curatoriales así como su labor habitual de investigación y docencia. El Administrador Departamental es responsable de las tareas administrativas diarias del herbario, como la compra de insumos, incluidos papel y etiquetas. El personal docente y los estudiantes que colaboran dentro del departamento ocasionalmente trabajan y actualizan las identificaciones de los especímenes, y hay dos botánicos jubilados que regularmente se ofrecen como voluntarios en la colección, ayudando al profesor a preparar los préstamos.

La Universidad ya tiene una búsqueda en línea para las colecciones de su biblioteca que es mantenida por el equipo web central de la Universidad en servidores alojados externamente. Actualmente, no están disponibles especímenes de historia natural a través de esta plataforma.

El Jefe del Departamento de Biología Vegetal ha obtenido recientemente una subvención de por años

de 50.000 USD para generar imágenes y publicar la información de las colecciones botánicas en línea. El equipo quiere aprovechar esta oportunidad para establecer un protocolo de digitalización y publicación permanente que le dará mayor visibilidad al herbario y seguirá atrayendo financiación.

Recopilación de datos



El herbario comprende aproximadamente 80.000 especímenes colectados entre 1960 y 2015 concentrados principalmente en una zona de alta biodiversidad: el Parque Nacional Short Mountain. La colección sigue creciendo como resultado de intercambios, donaciones y varios proyectos de investigación activos. Para cada espécimen, se documenta información sobre el colector, la hora, la fecha, la ubicación y la taxonomía. Hasta ahora no se han llevado a cabo medidas de control de calidad sobre los datos y no se ha realizado una toma de imágenes sistemática de los especímenes.

Descripción del conjunto de datos

El profesor dispone de una sencilla base de datos no relacional en su computador que sirve como un índice de muchos de los especímenes, pero no de todos. Cualquier imagen existente de los especímenes es mantenida localmente por los investigadores que la tomaron.

Ejercicios

Los ejercicios individuales para este caso de uso se encuentran en los módulos correspondientes.

Planificación

Ejercicio 1a-c

Captura de datos

Ejercicio 2

Gestión de datos

Ejercicio 3a-c

Publicación de datos

Ejercicio 4

Hoja de ejercicio

UC1-Herbarium-Exercise_EN.docx (MS Word 345 KB)

Planificación

En el módulo de planificación, revisará las etapas clave de planificación del proyecto y aprenderá a crear un flujo de trabajo viable. Además, creará un plan idealizado de proyecto/flujo de trabajo basado en el CASO DE USO I. Usted identificará metas, tareas, actores clave y roles, también asignará tareas específicas a las etapas.

Recursos

En este video (15:09), aprenderá cómo definir los elementos que tendrán un impacto en un proyecto y analizará las interacciones entre los elementos. Si no puedes ver el vídeo incrustado, puedes descargar localmente. (MP4 - 25.1 MB)

▶ https://www.youtube.com/watch?v=VRvUdMjd93c *(YouTube video)*

Organizar

Este video (15:52) se centra en las tareas y las formas en que pueden organizarse para preparar planes prácticos y documentación clara. Si no puede ver el vídeo incrustado, puede descargar localmente. (MP4 - 26.2 MB)

▶ https://www.youtube.com/watch?v=uhhK6B2Vwls *(YouTube video)*

Ejercicio 1a-c

Para esta actividad, usará una plataforma de tabla virtual. Revise las instrucciones del ejercicio a continuación.

Referencias

- Definiciones de los roles
- Definiciones de las etapas

Para ayudarle a trabajar en los ejercicios de planificación para todos los casos de uso, hemos puesto a su disposición cinco mesas virtuales de juegos de cartas.

Debido a que se trata de espacios compartidos, le pedimos que siga estas normas para que todos los que quieran utilizarlos puedan hacerlo:

1. Acomode todo al finalizar. Si usa una mesa de juego, asegúrese de devolverla a su estado original cuando haya terminado. Esto significa guardar todas las cartas y quitar/eliminar cualquier carta nueva que haya agregado a la baraja. Puede hacer esto haciendo clic en los botones de recuperación debajo de cada baraja de cartas.
2. Si una mesa está ocupada, pruebe con otra. Si observa que una mesa de cartas está siendo utilizada, por favor, use otro enlace de mesa de cartas. Si no puede encontrar una mesa de cartas vacía, avísenos a training@gbif.org para que podamos comprobar las mesas y asegurarnos de que no se hayan dejado desatendidas (o si necesitamos crear una mesa adicional para usted).
3. No deje una mesa desatendida. Por favor, no deje su trabajo sobre una mesa durante largos períodos de tiempo. Si necesita ausentarse durante una hora o dos, es aceptable, pero no deje las cartas desatendidas durante más de 4 horas. Si necesita dejar su configuración por un período prolongado de tiempo, use la plataforma de eLearning para notificar al grupo y a los entrenadores el enlace que está utilizando, por qué debe dejar sus cartas y cuándo regresará para completar su trabajo.
4. Realice capturas de pantalla con frecuencia. No podemos garantizar que sus cartas no sean reiniciadas accidentalmente por alguien del taller o por alguien de su propio proyecto. No puede restablecer las cartas una vez que se hayan retirado y restablecido. Si desea enviar una captura de pantalla con su ejercicio, hágalo.
5. Los entrenadores se reservan el derecho de restablecer las mesas de cartas. Si observamos que una o más mesas contienen cartas que no se han guardado y no se han utilizado durante más de cuatro horas sin ninguna notificación en el Foro de la plataforma eLearning, limpiaremos y restableceremos la mesa para que otros la usen.

Enlaces a la sala de juegos (cada mesa se abrirá en una nueva ventana/pestaña):

AZUL: https://playingcards.io/ndqppx

VERDE: https://playingcards.io/zpwe9j

NARANJA: https://playingcards.io/868jgh

PURPURA: https://playingcards.io/g6khb7

ROJO: https://playingcards.io/3pad4w

AMARILLO: https://playingcards.io/w488nb

Muchas gracias a Jwalant Patel y Eric Ma por encontrar y ayudar a crear las mesas de juego en línea para este ejercicio y a Kate Webbink por su experiencia artística. La plataforma de mesa en línea es proporcionada por https://playingcards.io/.

Ejercicio 1a

Leer USE CASE I (si no lo has hecho ya).

Utilizando las tarjetas, seleccione los objetivos que más se ajusten a los del proyecto y luego, elija las tareas que habría que llevar a cabo para completarlos. A continuación, identifique a las personas/recursos que el proyecto tiene disponibles y asígneles cartas de función según

corresponda. Por último, asigne los grupos de interés y afiliaciones.

1. Revise las cartas de METAS con el grupo, seleccione y diseñe las que se ajusten al caso de uso.
2. Revise las cartas de TAREAS con el grupo.
3. Asigne las cartas de TAREAS a cada una de las cartas de METAS seleccionadas en la descripción del caso de uso.
4. Identifique las instituciones y las personas mencionadas en el caso de uso y tome nota de ellas.
5. Disponga las cartas de AFILIACIONES sobre la mesa.
6. Revise las cartas de PARTES INTERESADAS, identifique cualquiera mencionada en el caso de uso y luego decida a qué AFILIACIÓN pertenecen.
7. Revise las cartas de ROL, identifique los mencionados en el caso de uso y decida a qué grupo de PARTES INTERESADAS pertenecen.
8. Tome notas en las cartas según corresponda.
9. Una vez que las cartas sean asignadas, tome fotografías/capturas de pantalla.
10. Use la hoja de ejercicio para proporcionar sus respuestas.

Pregunta

¿Faltan recursos o metas en las cartas que considere que son fundamentales para la finalización exitosa del proyecto? Anótelos en la hoja de respuestas.

Ejercicio 1b

Utilizando el análisis de las partes interesadas y las metas del ejercicio 1a, desarrolle un flujo de trabajo utilizando las cartas de ETAPA.

1. Vuelva a leer la introducción del caso de uso, si es necesario.
2. Decida de qué TAREAS será responsable cada ROL.
3. Revise las cartas de ETAPA y ordene las TAREAS de forma apropiada.
4. Utilice la hoja de ejercicios, previamente descargada, para proporcionar sus respuestas.

Preguntas

- ¿Hay cuellos de botella evidentes en el flujo de trabajo? Por ejemplo: ¿Hay demasiadas tareas para un rol/recurso en particular?
- ¿Qué cuestiones cree que serían importantes para la movilización exitosa de datos desde el punto de vista de cada parte interesada/rol? Por ejemplo: ¿Cuáles son los entregables? Pensando en el cronograma general del proyecto, ¿son realistas?
- Consolide las notas y priorice por orden de importancia.
- Si tiene tiempo puede explorar diferentes combinaciones, ya que diferentes escenarios son posibles en diferentes contextos, o incluso intentar documentar la situación de su propio proyecto.

Ejercicio 1c

Este ejercicio debe utilizarse cuando el curso se imparte virtualmente o in situ como una actividad grupal.

Después de los ejercicios, el presentador de cada grupo:

1. Presentará las partes interesadas y/o las tareas que faltan y explicará por qué se agregaron.
2. Destacará los dos problemas/temas más críticos que el grupo ha identificado.

Posibles puntos de discusión entre grupos:

- ¿Qué similitudes y diferencias se pueden ver en los flujos?
- ¿Hubo problemas comunes que surgieron en los grupos?

Revisión

Comprueba los conceptos aprendidos en esta sección.

1. ¿Cuál es el orden de los cinco Grupos de Procesos PMBoK?

   - ☐ Planear, Iniciar, Monitorear y Controlar, Ejecutar, Cerrar
   - ☐ Iniciar, Planear, Ejecutar, Monitorear y Controlar, Cerrar
   - ☐ Iniciar, Planear, Ejecutar, Cerrar, Monitorear y Controlar
   - ☐ Iniciar, Planear, Monitorear y Controlar, Ejecutar, Cerrar

2. ¿Cuáles son los tipos de entregables? (múltiples respuestas correctas)

   - ☐ Declarado
   - ☐ Implícito
   - ☐ Estimado
   - ☐ Directo
   - ☐ Indirecto
   - ☐ Suposiciones

3. ¿Qué es un cuello de botella?

   - ☐ un bloqueo que retrasa el desarrollo o el progreso
   - ☐ un espacio donde falta algo o alguien
   - ☐ un problema o situación que impide que alguien haga algo, o que hace que algo sea imposible.

4. ¿Cuáles son los ejemplos de las tareas de movilización? (múltiples respuestas correctas)

   - ☐ Afiliación
   - ☐ Publicación
   - ☐ Imágenes

- ☐ Georreferenciación
- ☐ Mayor conciencia pública

Captura de datos

En este módulo, aprenderá sobre el concepto de estándares, en particular, el estándar Darwin Core y sus componentes. También aprenderá los tipos de datos primarios sobre biodiversidad y la mejor manera de compartir esa información a través de GBIF. Por último, revisará los principios de calidad de los datos en el contexto de la captura de datos y aprenderá sobre la calidad y coherencia de los datos (especialmente en temas como georreferenciación, fechas, nombres y verificación de taxones).

Estándares y Darwin Core

En este video (15:37), aprenderá cómo se interactúa con los estándares diariamente. Luego, se le presentarán Biodiversity Information Standards, incluido el Estándar Darwin Core que continuará utilizando a lo largo de este curso. Si no puede ver el video incrustado, puede descargarlo localmente. (MP4 - 27 MB)

▶ https://www.youtube.com/watch?v=S02PJHPsRAs *(YouTube video)*

Orígenes y tipos de datos

En este video (10:45), revisará los datos primarios de biodiversidad que se pueden compartir en GBIF. Si no puede ver el video incrustado, puede descargalo localmente. (MP4 - 19 MB)

▶ https://www.youtube.com/watch?v=wKeOveydjsw *(YouTube video)*

Preguntas

- ¿Su tipo de datos es diferente de lo que se pensó originalmente?
- ¿Con qué tipo de datos trabajas?
- ¿Cómo publicaría sus datos a GBIF (usando qué núcleo y/o extensión)?

Captura, procesamiento y calidad de datos

En este video (09:11), explorará los principios de calidad de datos aplicados a la captura de datos, específicamente al capturar datos de etiquetas de colecciónes biológicas, cuadernos de campo, hojas de cálculo, etc. Si no puede ver el vídeo incrustado, puede descargarlo localmente. (MP4 - 19 MB)

▶ https://www.youtube.com/watch?v=QkDJIkmwBMA *(YouTube video)*

Ejercicio 2

Para esta actividad, completará un ejercicio simulando la captura de datos. Empezará a trabajar con los términos Darwin Core y tomará decisiones sobre los datos que son necesarias para su organización/proyecto, también considerará cuáles de esos datos serán compartidos más adelante durante la publicación.

Leer USE CASE I (si no lo has hecho ya).

Imagine que usted es la persona encargada de transcribir los datos encontrados en los ejemplares del herbario.

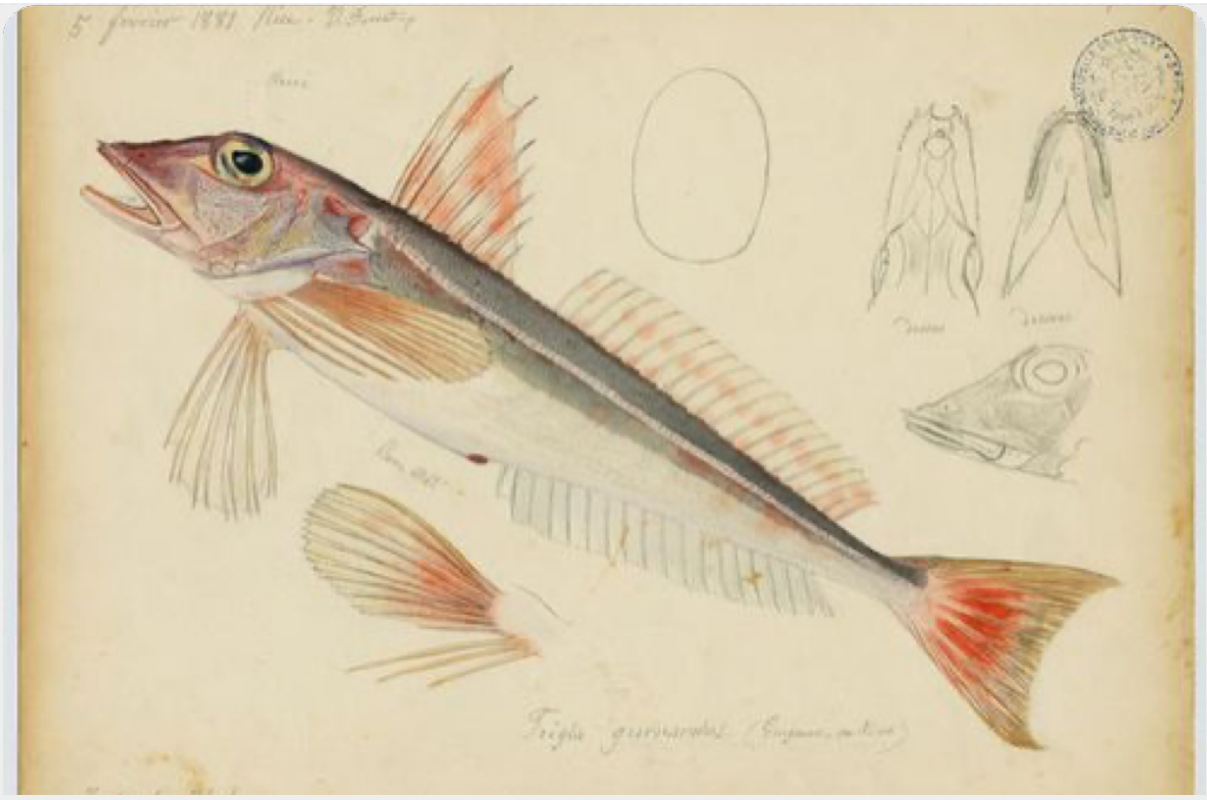
1. Descargue lUC1-2-base-material.zip. (34.4 MB). Hay 10 imágenes. Dos imágenes por muestra, para un total de cinco muestras. Los materiales de herbario están en español (los datos pueden llegar a usted en varios medios y en otros idiomas distintos al suyo), pero debe poder reconocer los datos contenidos en los campos de las etiquetas. Recuerde utilizar las dos imágenes por registro para recopilar la información.
2. Descargue la plantilla de la hoja de cálculo: UC1-2-occurrence-template.xlsx (57.3 KB) para transcribir la información que se encuentra en cada una de las imágenes de los cinco ejemplares.
3. Utilice la hoja de ejercicios, previamente descargada, para proporcionar sus respuestas.

es posible que tenga que añadir campos a la hoja de cálculo, ya que puede capturar más información de las etiquetas de la que estaba prevista en la plantilla.

Revisión

Comprueba los conceptos aprendidos en esta sección.

1. ¿Qué tipo(s) de conjuntos de datos elegiría para una colección de ictiología?



Eutrigla gurnardus (Linnaeus, 1758) | *Muséum d'histoire naturelle de Nice*

- ☐ registro biológico
- ☐ lista de especies
- ☐ evento de muestreo

2. ¿Qué tipo(s) de conjunto de datos elegiría(s) para una lista de especies invasoras?



Jacinto de agua (Eichhornia crassipes) observada en Bourail, Nueva Caledonia, donde es una especie introducida e invasora de acuerdo a GRIIS. Foto por geŕard (2016) licenciada bajo CC BY-SA 2.0

- ☐ registro biológico
- ☐ lista de especies
- ☐ evento de muestreo

3. ¿Qué tipo(s) de conjuntos de datos elegiría para la flora y fauna de un estudio de impacto ambiental?

   Los estudios de evaluación de impacto ambiental son realizados por expertos para evaluar la biodiversidad y los biotopos de una zona determinada, antes, durante y después de que se vea afectada por las actividades humanas (obras de carreteras, turbinas eólicas, minería, construcción de edificios, etc.).



Entomólogo cazando mariposas *por Matthieu Gauvain (CC-BY-SA)*

- ☐ registro biológico
- ☐ lista de especies
- ☐ evento de muestreo

4. ¿Qué tipo(s) de conjunto de datos elegiría(s) para los datos de seguimiento de aves?

   Los datos de rastrei de aves se registran utilizando dispositivos específicos, como rastreadores GPS puestos en aves vivas, lo que permite a los científicos rastrear sus rutas migratorias o sitios de reproducción.



Buitre leonado *observado en la Reserva Natural de Gamla por גיזונית - MinoZig (CC0)*

- ☐ registro biológico
- ☐ lista de especies
- ☐ evento de muestreo

5. ¿Qué tipo(s) de conjuntos de datos elegiría para los datos de trampas de insectos?



Trampa para insectos *por miheco (CC-BY-SA)*

- ☐ registro biológico

- ☐ lista de especies
- ☐ evento de muestreo

6. ¿Qué tipo(s) de conjuntos de datos elegiría para los datos de gestión de parques nacionales?

   Los datos adquiridos en el contexto de la gestión de áreas protegidas (como parques nacionales, pero también reservas naturales más pequeñas) pueden ser diversos y tener diferentes orígenes: estudios botánicos, rastreo de animales marcados, observaciones de guardabosques y guardias, e incluso datos de 'ciencia ciudadana' o datos inferidos de imágenes compartidas en las redes sociales.



elefantes de Sri Lankan *observado por pen_ash.*

- ☐ registro biológico
- ☐ lista de especies
- ☐ evento de muestreo

7. ¿Qué tipo(s) de conjuntos de datos elegiría para un bioblitz de ciencia ciudadana?

   Los datos de ciencia ciudadana a menudo se recopilan a través de días temáticos de trabajo de campo conocidos como "bioblitz". Los voluntarios generalmente se reúnen en un área determinada y pasan el día tratando de observar e identificar tantas especies como puedan en esta área.

   Los datos de cada participante se capturan y fusionan en la herramienta de captura o gestión de datos del programa de ciencia ciudadana.



Buscando aves con personal del parque *por el Servicio de Parques Nacionales de EE. UU. (reutilización autorizada en la búsqueda de imágenes de Google)*

- ☐ registro biológico
- ☐ lista de especies
- ☐ evento de muestreo

8. ¿Qué tipo(s) de connjuntos de datos elegiría(s) para una lista regional de especies?



Rinoceronte negro *observado en el zoológico de Magdeburg en Alemania por Mani300*

- ☐ registro biológico
- ☐ lista de especies
- ☐ evento de muestreo

Gestión de datos

En este módulo revisará los principales conceptos, herramientas relacionadas y buenas prácticas para la gestión de datos, en particular, la limpieza y estandarización de datos.

Principios de la gestión de datos

En este video (09:49), usted revisará un importante conjunto de principios necesarios para mejorar los datos a través de los procesos de limpieza de datos. Si no puede ver el vídeo incrustado, puede descargalo localmente. (MP4 - 16.6 MB)

▶ https://www.youtube.com/watch?v=4ijm1cJeVHE *(YouTube video)*

Herramientas de gestión de datos

En este video (06:42), aprenderá sobre una variedad de herramientas que puede utilizar para mejorar la calidad de sus datos. Si no puede ver el vídeo incrustado, puede descargalo localmente. (MP4 - 10.3 MB)

▶ https://www.youtube.com/watch?v=Ru3vWiYU3gw *(YouTube video)*

Ejercicio 3a-c

Para estos ejercicios, realizará verificaciones técnicas y controles de validación de consistencia, mejorará los datos con diferentes herramientas y aprenderá a usar OpenRefine.

Leer USE CASE I (si no lo has hecho ya).

Su institución es parte de la “Asociación Global de Poales (GPA)”. Esta asociación ha obtenido fondos para publicar una flora actualizada sobre el grupo y ha solicitado a su herbario que participe y proporcione los registros de alta calidad que pueda tener sobre este orden de plantas. El pedido está bien representado en su colección, por lo que cree que podría contribuir sustancialmente a este esfuerzo.

Ejercicio 3a

Controles de validación

En este ejercicio nos centraremos en errores técnicos y realizaremos una comprobación básica de validación para identificar errores técnicos. Consulte Comprobaciones de validación para obtener información sobre los tipos de errores.

1. Descargue UC1-3ab-data-cleaning.csv. (207.5 KB)
2. Importe el archivo CSV en Excel usando el asistente de Excel. Consulte este Excel-tips-EN.pdf (PDF, 7 MB) para obtener instrucciones de importación para su sistema operativo (Windows, Mac, Linux).
3. Encuentre y corrija los errores manualmente.
4. Utilice la hoja de ejercicios, previamente descargada, para proporcionar sus respuestas.

Ejercicio 3b

Otras herramientas de gestión de datos

La asociación GPA le ha proporcionado una lista de elementos de calidad de datos que debe verificar:

- Todos los nombres de las plantas (nombre completo) están escritos correctamente
- Todos los nombres de las plantas pertenecen al pedido
- Todos los registros tienen coordenadas
- Todas las coordenadas están dentro del país indicado y convertidas a formato decimal
- Todas las fechas están en la columna correspondiente y en el formato AAAA-MM-DD

Las tres categorías de errores son:

- Errores de nomenclatura
- Errores de formato
- Errores geográficos / valores atípicos

1. Consulte Herramientas útiles para completar el ejercicio. No está limitado a estas herramientas, puede utilizar cualquier herramienta que desee.
2. Utilice el mismo archivo del ejercicio anterior.
3. Haga la corrección SOLO para la familia Eriocaulaceae (por lo que puede querer filtrar los datos)
4. Corrija los errores encontrados en el conjunto de datos usado en el ejercicio 3a (ejercicio anterior), usando las herramientas de su elección, y documente los cambios que realiza en la hoja de ejercicio.
5. Corrija todo el archivo si tiene tiempo.
6. Utilice la hoja de ejercicios, previamente descargada, para proporcionar sus respuestas.

Ejercicio 3c



En este vídeo (03:27) aprenderá sobre OpenRefine. Puede utilizar OpenRefine para estandarizar y mejorar la calidad de sus datos. Si no puede ver el vídeo incrustado, puede descargalo localmente. (MP4 - 3.8 MB)

▶ https://www.youtube.com/watch?v=_YFw_bfwc3Y *(YouTube video)*

OpenRefine

En este ejercicio utilizamos OpenRefine para mejorar la calidad de un conjunto de datos utilizando las funcionalidades predeterminadas, los servicios web existentes y las expresiones regulares.

1. Descargue UC1-3c-open-refine.csv. (207.5 KB)
2. Descargue y complete los ejercicios disponibles en OpenRefine-Exercise3c-EN.pdf. (PDF, 1.1 MB) También disponibles en Frances y Español.
3. Utilice la hoja de ejercicios, previamente descargada, para proporcionar sus respuestas.

Consejos para el ejercicio

Comprobaciones de validación

Errores técnicos Relativamente sencillos, a menudo automatizables, **comprobación de la integridad de los datos**. Pueden indicar exportaciones incorrectas, mapeo de datos, desplazamientos de campos (p. ej., desplazamiento de una columna a la derecha) o falta de datos en la fuente.

- **Completitud**: Si todos los datos y metadatos están disponibles: ¿están presentes todos los campos?, ¿están completos todos los campos?
- **Límites**: Por ejemplo, son días dados en el rango 1-31 (dependiendo del mes)
- **Tipo de datos**: Por ejemplo, ¿el campo Fecha contiene una fecha o un número?
- **Formato de datos**: Por ejemplo, ¿las fechas se proporcionan como 01/01/2010 o 01/ene/10?

Errores de consistencia

Aplicación de reglas del mundo real a los datos. Pueden indicar la entrada incorrecta de datos a partir de registros antiguos, errores de transcripción o procesamiento posterior. Algunas son complejas de implementar y **requieren conjuntos de datos de referencia para su comprobación**. Por ejemplo, una lista de colectores conocidos y hábitos de colecta. Estas reglas se pueden recopilar de los usuarios y analistas de datos.

- **Taxonómico**: Por ejemplo, si se ha identificado a nivel de especie, ¿se ha proporcionado un nombre científico binomial y entradas en los campos de género y especie?
- **Aceptación general**: ¿Son coherentes las fechas de recopilación, identificación, actualización y digitalización?
- **Valores atípicos**: Detecte los valores atípicos, pero recuerde que no todos los valores atípicos son necesariamente errores. Por ejemplo, compárelos con un rango de especies conocido, o con un rango ambiental conocido (pero recuerde que los valores atípicos pueden ser identificaciones incorrectas, en lugar de coordenadas incorrectas).
- **Geográfico**: ¿Están las coordenadas dentro de la localidad o región identificada? Por ejemplo, ¿hay registros terrestres en el mar o registros marinos en tierra?
- **Patrones de colecta**: ¿El detalle del registro coincide con los patrones de colecta conocidos de la organización o el colector? ¿Parece que se han creado registros después de la muerte de un colector (¿podría ser un colector diferente con un nombre similar)? Por ejemplo, ¿se atribuyen registros de mamíferos a un grupo de observación de aves?
- **Exactitud y precisión**: Por ejemplo, ¿hay algún registro georreferenciado que indique una precisión o exactitud muy alta en un período de colecta anterior al GPS (o previo a los GPS exactos)?
- **Métodos de coleta**: Los diferentes métodos de muestreo (por ejemplo, transectos y levantamientos de área) tienen características particulares. ¿Los registros son consistentes con el método proporcionado?

Herramientas útiles

- **GBIF Name Parser**: https://www.gbif.org/tools/name-parser
- **Global Names Resolver**: http://resolver.globalnames.org
- **Catalogue of Life name match**: https://data.catalogueoflife.org/tools/name-match
- **TNRS**: https://tnrs.biendata.org/
- **WoRMS**: https://www.marinespecies.org/aphia.php?p=match
- **InfoXY**: http://splink.cria.org.br/infoxy?criaLANG=en
- **Calculadora de georreferenciación**: http://georeferencing.org/georefcalculator/gc.html
- **Convertidor de coordenadas Canadensys**: http://data.canadensys.net/tools/coordinates
- **Análisis de fechas de Canadensys**: http://data.canadensys.net/tools/dates
- **Google Maps**: https://maps.google.com/

Revisión



Comprueba los conceptos aprendidos en esta sección.

1. ¿Por qué es mejor limpiar sus datos?
   - ☐ para hacerlos tan aptos para el uso ("fit for use") como sea posible
   - ☐ para lograr sus objetivos de calidad de datos
   - ☐ los datos deben ser limpiados por los usuarios. no por los proveedores
2. ¿Cómo debe organizar su flujo de trabajo de limpieza de datos?
   - ☐ trabajar solo, usted conoce mejor sus datos
   - ☐ pregunte a sus colegas experimentados
   - ☐ trabajar a nivel institucional para armonizar los flujos de trabajo de calidad de los datos
3. Qué es mejor:
   - ☐ evitar que se generen errores
   - ☐ corregir los errores tan pronto como los encuentre en su base de datos u hoja de cálculo
   - ☐ no limpiar los errores, sino documentarlos sobre la marcha, para que las personas que reutilizan sus datos sepan dónde están
4. ¿Quién es responsable por la calidad de los datos?
   - ☐ La (s) persona (s) que registran los datos en el campo
   - ☐ Los transcriptores de datos
   - ☐ El administrador de la base de datos
   - ☐ Todas las personas involucradas en la gestión de datos.
   - ☐ Las personas que utilizan sus datos
   - ☐ GBIF

5. ¿Qué herramientas se pueden utilizar para limpiar sus datos?

- ☐ Excel y otras herramientas de gestión de hojas de cálculo
- ☐ OpenRefine
- ☐ Su software de base de datos
- ☐ Herramientas en línea como Scientific Names Resolver o Google Maps

Publicación de datos

En este módulo, aprenderá sobre conceptos de publicación de datos, incluyendo IPT, núcleos y extensiones, y la importancia de las licencias, metadatos, campos obligatorios y alojamiento de conjuntos de datos.

Conceptos de publicación de datos

En este video (11:45), aprenderá sobre los conceptos de publicación de datos y recibirá una introducción a la Herramienta de Publicación Integrada (IPT). Si no puede ver el vídeo incrustado, puede descargarlo localmente. (MP4 - 20 MB)

▶ https://www.youtube.com/watch?v=b900d9ukjSQ *(YouTube video)*

Descripción general del IPT

En este vídeo (06:56), recibirá una visión general de la interfaz de publicación de datos IPT. Si no puede ver el vídeo incrustado, puede descargarlo it locally. (MP4 - 8.7 MB)

▶ https://www.youtube.com/watch?v=gHXsaN_JWeI *(YouTube video)*

Instalaciones IPT de capacitación

Si aún no se le ha proporcionado un usuario, póngase en contacto con training@gbif.org y se le proporcionará un nombre de usuario y una contraseña en el IPT del curso.

https://training-ipt-a.gbif.org/

https://training-ipt-b.gbif.org/

https://training-ipt-c.gbif.org/

Demostración del IPT

En este video (24:16), aprenderá a publicar un conjunto de datos de registros biológicos usando un IPT. Si no puede ver el vídeo incrustado, puede descargarlo localmente. (MP4 - 52.6 MB)

▶ https://www.youtube.com/watch?v=eDH9IoTrMVE *(YouTube video)*

Ejercicio 4

En este ejercicio, publicará un conjunto de datos de registros biológicos usando el IPT.

Leer USE CASE I (si no lo has hecho ya).

Publicación de datos

Después de limpiar los datos del conjunto de datos de Poales, el equipo considera que publicar los datos en línea a través de la red GBIF podría ser una buena manera de hacer visible este esfuerzo. Se le ha solicitado que lidere ese trabajo de publicación, basándose en el conjunto de datos.

1. En este punto, necesita una cuenta en uno de los IPT del curso. Si aún no se le ha proporcionado un usuario, comuníquese con training@gbif.org y se le proporcionará un nombre de usuario y una contraseña en uno de los IPT del curso.
2. Descargue UC1-4-poales-publishing.csv. (233.5 KB)
3. Utilice la instalación IPT asignada y publique este archivo.
4. Utilice la hoja de ejercicios, previamente descargada, para proporcionar sus respuestas.

Revisión

Comprueba los conceptos aprendidos en esta sección.

1. ¿Qué significa la publicación de datos en el contexto de GBIF?
   - ☐ Exportar un archivo csv de sus datos limpios que puede compartir con sus colegas
   - ☐ Escribir un artículo que describa sus datos, y el/los protocolo(s) que utilizó para recolectarlos, capturarlos y limpiarlos
   - ☐ Hacer públicamente accesibles sus datos de biodiversidad en un formato estandarizado
2. ¿Qué es un IPT?
   - ☐ una herramienta que lo ayuda a administrar y corregir sus datos
   - ☐ una herramienta que le ayuda a publicar sus datos en GBIF
   - ☐ una herramienta que le ayuda a producir una publicación de datos (Data Paper)
3. ¿Qué licencias y exenciones de Creative Commons son recomendadas por GBIF para la publicación de datos?
   - ☐ CC-BY, CC-BY-SA y CC-BY-ND
   - ☐ CC0, CC-BY y CC-BY-NC
   - ☐ CC0, CC-BY y CC-BY-SA
4. ¿Cuáles son los tres Núcleos (Cores) entre los que puedes elegir un recurso IPT?
   - ☐ Metadata Core, Occurrence Core, Multimedia Core
   - ☐ Taxon Core, Collection Core, MeasurementOrFact Core

- ☐ Occurrence Core, Taxon Core, Event Core

5. ¿Cuántos archivos de Extensiones puede tener un conjunto de datos?

- ☐ cero
- ☐ una
- ☐ tantos como sea necesario

Evaluación y certificación

En este módulo, revisará los criterios que se utilizarán para evaluar las tareas y conseguir la certificación.

Una vez completado el curso y la evaluación de de las asignaciones (por parte de capacitadores y mentores) de forma exitosa, los participantes tienen la oportunidad de recibir una certificación oficial en forma de Open Badge.



Un puntaje general de 2.5-2.9 otorga una insignia BÁSICA de Movilización de Datos de Biodiversidad



Un puntaje general de 3.0-4.0 otorga una insignia AVANZADA de Movilización de Datos de Biodiversidad

Los participantes deben enviar (en inglés) el Caso de Uso II (elegir entre las dos opciones) y el Caso de Uso III, cada caso de uso se califica según las rúbricas educativas del curso. Las rúbricas definen las habilidades y los niveles de desempeño sobre los que se construyen los objetivos de aprendizaje del curso.

Revise las rúbricas para asegurarse de comprender las habilidades que se evaluarán para la certificación.

Rúbrica de planificación

Planificación

Habilidades	Desempeño básico 1	Desempeño intermedio 2	Desempeño avanzado 3	Desempeño excepcional 4
A. Conocimiento de los diferentes elementos que forman parte de un plan de movilización de datos de biodiversidad	Muestra comprensión de algunos roles y tareas requeridos en un plan de movilización sólido, pero no es capaz de diferenciar quién realiza cada tarea en el plan o confunde los roles.	Comprende muchos de los roles y tareas necesarios en un plan de movilización, pero aún no tiene vínculos e interacciones entre ellos.	Comprende las tareas y roles clave necesarios en un plan de movilización y cómo interactúan.	Puede identificar roles y tareas adicionales que pueden ser necesarios para situaciones específicas
B. Capacidad para aplicar los diferentes elementos de un plan de movilización de datos de biodiversidad a un contexto institucional dado (p. ej. el propio)	Puede ver cómo algunas de las tareas y roles se relacionan con un contexto institucional dado, pero tiene dificultades cuando el contexto cambia aunque sea levemente de las referencias genéricas utilizadas.	Puede ver cómo la mayoría de las tareas y roles se pueden aplicar a un contexto institucional dado, pero aún tiene vacíos cuando la situación es diferente a las referencias genéricas utilizadas.	Puede identificar todas las tareas y roles relevantes y cómo se pueden traducir a un contexto institucional dado, incluso si la situación no es idéntica a la de las referencias genéricas utilizadas.	Saber aplicar creativamente los diferentes elementos de un plan de movilización, incorporando elementos nuevos que no se encuentran en las referencias genéricas utilizadas.
C. Capacidad para escribir/redactar un documento de planificación claro	Puede recopilar materia prima que sería la base de un plan (p. ej. como puntos clave), pero tiene dificultades para traducir esto en una narrativa.	Puede construir un plan de trabajo básico, pero carece de coherencia entre sus elementos.	Puede redactar un documento coherente y bien articulado que incluya todos los elementos relevantes de un plan de trabajo.	Puede redactar planes concisos, con resúmenes efectivos y niveles de detalle cada vez mayores.

Habilidades	Desempeño básico 1	Desempeño intermedio 2	Desempeño avanzado 3	Desempeño excepcional 4
D. Capacidad para evaluar un plan de movilización de datos sobre biodiversidad	Puede identificar algunos elementos de un plan de movilización genérico en un ejemplo dado, pero tiene dificultades para relacionarlos con tareas y roles.	Puede identificar la mayoría de los elementos de un plan de movilización y relacionarlos con tareas y roles cuando son idénticos a las referencias utilizadas. Tiene dificultades para evaluar la viabilidad de algunos elementos del plan.	Puede reconocer todos los elementos existentes en un plan de movilización y puede identificar brechas, esfuerzos duplicados e inconsistencias. Puede evaluar la calidad de los componentes individuales. Puede evaluar los puntos potenciales de éxito y los puntos débiles del plan en su conjunto.	Puede sugerir soluciones para solucionar problemas identificados en el plan evaluado.

Rúbrica de captura de datos

Captura de datos

Habilidades	Desempeño básico 1	Desempeño intermedio 2	Desempeño avanzado 3	Desempeño excepcional 4
A. Capacidad para identificar el tipo de datos digitales que se pueden extraer de una fuente de datos de biodiversidad (es decir, que se pueden publicar utilizando la red GBIF)	Puede identificar solo los tipos de datos más evidentes de fuentes comunes de datos de biodiversidad (p. ej., registros biológicos de especímenes de colecciones de historia natural). Muestra poca comprensión del potencial de la publicación en línea con GBIF.	Puede identificar correctamente, con frecuencia, al menos un tipo de datos digitales que se pueden extraer de fuentes comunes de datos. Tiene dificultad para identificar cuáles se pueden publicar actualmente usando GBIF.	Puede identificar siempre uno (o más) tipos de datos digitales que pueden extraerse de fuentes de datos comunes. Puede identificar cuál de esos tipos puede publicarse actualmente usando GBIF.	Puede identificar siempre uno o más tipos de datos digitales que pueden extraerse de fuentes de datos comunes y no comunes. Puede identificar cuál de esos tipos puede ser publicado actualmente usando GBIF y cuáles están en discusión. Puede identificar los núcleos de datos y las extensiones utilizadas para publicar esos tipos de datos.

Habilidades	Desempeño básico 1	Desempeño intermedio 2	Desempeño avanzado 3	Desempeño excepcional 4
B. Capacidad para extraer información relevante de una fuente de datos de biodiversidad en estructuras de datos simples (p. ej., hojas de cálculo) que siguen los estándares internacionales.	Solo puede extraer grandes fragmentos de información obvia (por ejemplo, toda la información geográfica como una sola unidad) que son evidentes en la fuente de datos. Muestra poco conocimiento de los estándares actuales para el registro de datos de biodiversidad.	Puede recuperar varios elementos de información de la fuente de datos (pero no todos) y puede desglosarlos en partes significativas. Muestra algunos conocimientos básicos de los estándares más comunes (p. ej., DwC) y los campos de datos más utilizados en esos estándares.	Puede identificar toda la información valiosa en una fuente de datos y extraer los elementos obligatorios en una estructura de datos estándar (p. ej., una hoja de cálculo basada en Simple DwC). Puede identificar la información que falta e inferir de la información existente (p. ej., deducir el nombre de un país a partir de una provincia).	Puede identificar toda la información valiosa en una fuente de datos compleja y dividirla en partes significativas que luego se traducen directamente en estándares internacionales. Puede identificar la información crítica que falta en la fuente e inferirla de los datos existentes o de información adicional sobre la fuente (metadatos).

Habilidades	Desempeño básico 1	Desempeño intermedio 2	Desempeño avanzado 3	Desempeño excepcional 4
C. Capacidad para comprender y aplicar los principios básicos de la calidad de los datos al proceso de captura de datos.	Muestra una comprensión limitada de cómo la aplicación de principios simples de calidad de datos puede tener un gran impacto en el producto final, evitando una limpieza adicional necesaria después.	Conoce algunos de los principios más genéricos de la calidad de los datos (p. ej., evitar errores ortográficos) pero muestra un conocimiento limitado sobre cómo aplicar principios más específicos al proceso de captura de datos.	Conoce todos los principios básicos de la calidad de los datos y cómo aplicarlos de forma sencilla al proceso de captura de datos. Utiliza formatos de manera consistente durante el proceso de captura de datos (p. ej., en fechas, nombres de países). Documenta todos los procedimientos y cambios relacionados con la calidad de los datos de una manera sencilla.	Demuestra un buen conocimiento de todos los principios comunes de la calidad de los datos y cómo utilizarlos para mejorar el proceso de captura de datos. Utiliza formatos de datos de forma coherente y puede utilizar nomenclaturas, listas de referencias o funciones específicas del software para mejorar la calidad del original. Documenta claramente todos los cambios y decisiones tomadas en relación con la calidad de los datos.

Rúbrica de gestión de datos

Gestión de datos

Habilidades	Desempeño basal 1	Desempeño en desarrollo 2	Desempeño logrado 3	Desempeño excepcional 4
A. Capacidad para evaluar la calidad (es decir, identificar problemas y sus tipos) de un conjunto de datos de biodiversidad.	Solo utiliza controles visuales para analizar la calidad. No se puede diferenciar entre tipos de errores. Puede detectar valores faltantes en los campos obligatorios y graves inconsistencias de datos.	Solo puede utilizar técnicas muy básicas (por ejemplo, clasificación) para analizar la calidad de los datos. Puede detectar discrepancias entre los nombres de campo y el contenido. Puede identificar errores técnicos de manera consistente, pero solo los errores de coherencia más típicos en un conjunto de datos.	Puede utilizar herramientas y técnicas específicas para evaluar la calidad. Reconoce el nivel mínimo de desagregación / normalización necesario para el uso común y la publicación. Puede identificar consistentemente errores técnicos y la mayoría de los errores de coherencia en un conjunto de datos.	Utiliza un enfoque sistemático para el análisis de conjuntos de datos que cubre todos los dominios de datos principales. Puede identificar consistentemente errores técnicos y de coherencia en un conjunto de datos. Puede utilizar otras fuentes de datos (por ejemplo, metadatos u otros conjuntos de datos) para identificar o inferir errores de coherencia en un conjunto de datos.
B. Capacidad para realizar corrección de formato de datos.	Solo se pueden realizar correcciones manualmente en las tablas. Muestra conocimientos genéricos sobre el uso de tipos de formato en datos digitales (por ejemplo, fechas, cadenas, números)	Puede identificar al menos una herramienta específica para corregir automáticamente errores de formato, pero solo puede usarla en casos específicos. De lo contrario, utiliza mecanismos simples (por ejemplo, "buscar y reemplazar") para resolver problemas.	Puede utilizar al menos una herramienta para corregir automáticamente los errores de formato.	Puede utilizar funciones avanzadas de más de una herramienta para corregir errores de formato.

Habilidades	Desempeño basal 1	Desempeño en desarrollo 2	Desempeño logrado 3	Desempeño excepcional 4
C. Capacidad para realizar corrección de datos nomenclaturales.	Solo se pueden realizar correcciones manualmente en las tablas. Solo utiliza el conocimiento personal de grupos taxonómicos conocidos.	Puede identificar al menos una herramienta específica para corregir automáticamente errores de nomenclatura, pero solo puede usarla en casos específicos. De lo contrario, utiliza mecanismos simples (por ejemplo, "buscar y reemplazar") para resolver problemas.	Puede utilizar al menos una herramienta para corregir automáticamente errores de nomenclatura. Puede encontrar y utilizar información de nomenclatura de referencia adecuada para los grupos taxonómicos con los que trabaja habitualmente.	Puede utilizar más de una herramienta para corregir errores de nomenclatura. Puede encontrar y utilizar información de nomenclatura de referencia adecuada para grupos taxonómicos fuera de sus áreas de especialización.
D. Capacidad para realizar corrección de datos geográficos.	Solo se pueden realizar correcciones manualmente en las tablas. Solo utiliza el conocimiento personal de áreas geográficas conocidas.	Puede identificar al menos una herramienta específica para mapear y / o corregir automáticamente errores en la información geográfica, pero solo puede usarla en casos específicos. De lo contrario, utiliza mecanismos simples (por ejemplo, "buscar y reemplazar") para resolver problemas.	Puede utilizar al menos una herramienta para mapear y / o corregir automáticamente errores en la información geográfica. Puede encontrar y utilizar información geográfica de referencia adecuada en un formato adecuado para las áreas con las que trabaja habitualmente.	Puede utilizar más de una herramienta para mapear y / o corregir automáticamente errores en la información geográfica. Puede encontrar y utilizar información geográfica de referencia en un formato adecuado para áreas fuera de sus áreas de especialización.

Habilidades	Desempeño basal 1	Desempeño en desarrollo 2	Desempeño logrado 3	Desempeño excepcional 4
E. Capacidad para utilizar software específico (por ejemplo, OpenRefine) como herramienta para la limpieza de datos.	Puede identificar al menos una herramienta de limpieza de datos. Puede identificar las características principales de una herramienta de limpieza de datos (por ejemplo, OpenRefine).	Puede identificar varias herramientas de limpieza de datos. Puede usar una o algunas de las características básicas del software de limpieza de datos para limpiar un conjunto de datos (por ejemplo, crear un proyecto OpenRefine, usar facetado, filtrado, agrupamiento o reconciliación).	Puede utilizar todas las funciones básicas de un software de limpieza de datos para limpiar un conjunto de datos (por ejemplo, en OpenRefine: facetado, filtrado, agrupación, reconciliación).	Puede usar las funciones avanzadas de uno o más paquetes de software de limpieza de datos para limpiar conjuntos de datos (por ejemplo, en OpenRefine: usar API, expresiones regulares, Google Refine Expression Language).
F. Capacidad para documentar procedimientos de transformación de datos.	Rara vez describe los cambios realizados durante la conservación, el formato o la transformación de los datos.	Describe los cambios realizados la mayor parte del tiempo. No describe los cambios de manera consistente o completa (por ejemplo, describe el cambio, pero no el autor).	Recuerda siempre describir los cambios realizados. Siempre describe los cambios de forma coherente, de modo que todas las ediciones del mismo tipo se puedan identificar fácilmente.	Puede describir con precisión y coherencia los cambios realizados de forma repetible.

Rúbrica de publicación de datos

Publicación de datos

Habilidades	Desempeño básico 1	Desempeño intermedio 2	Desempeño avanzado 3	Desempeño excepcional 4
A. Conocimiento sobre estándares de datos de información sobre biodiversidad (IBD).	Muestra conocimiento limitado o nulo sobre los estándares de datos IBD y cuáles son aceptados por GBIF.	Puede identificar los estándares IBD y sabe cuáles son aceptados por GBIF, pero no sabe dónde encontrar información sobre cómo utilizarlos. No puede identificar qué términos son obligatorios.	Conoce los estándares IBD aceptados por GBIF. Puede encontrar una lista de los núcleos y extensiones de datos aceptados. Publica conjuntos de datos de acuerdo con los estándares GBIF requeridos y / o recomendados para términos de datos y metadatos y sabe cómo encontrar las definiciones de los términos.	Demuestra comprensión sobre las características y limitaciones de los distintos estándares IBD.
B. Capacidad para analizar la idoneidad de un conjunto de datos de biodiversidad para su publicación a través de GBIF.	Muestra un conocimiento limitado o nulo de los criterios formales que debe cumplir un conjunto de datos para poder publicarse a través de GBIF.	Conoce los criterios formales que debe cumplir un conjunto de datos para poder publicarse a través de GBIF, pero no puede evaluar si un conjunto de datos determinado los cumple.	Puede evaluar correctamente si un conjunto de datos se puede publicar actualmente a través de GBIF. Puede asignar al menos un tipo de datos válido (= núcleo) a un conjunto de datos según la descripción proporcionada por el titular de los datos y después de haber analizado el conjunto de datos.	Puede identificar más de una opción de publicación para un conjunto de datos (cuando sea posible).
C. Uso de IPT: capacidad para producir/analizar metadatos de alta calidad.	Muestra un conocimiento limitado o nulo sobre las características de los buenos metadatos.	Conoce las características de los buenos metadatos, pero tiene dificultades para reconocerlos.	Conoce las características de los buenos metadatos y cómo reconocerlos. Puede producir recomendaciones sobre cómo mejorar los metadatos existentes.	Conoce las características de los metadatos de alta calidad y cómo producirlos.

Habilidades	Desempeño básico 1	Desempeño intermedio 2	Desempeño avanzado 3	Desempeño excepcional 4
D. Uso de IPT: capacidad para cargar/conectar datos y asignarlos a núcleos y extensiones existentes.	Puede cargar conjuntos de datos de un solo archivo en un IPT, pero no puede asignarlos a ningún núcleo.	Solo puede cargar conjuntos de datos de un solo archivo en un IPT y asignarlos a un solo tipo de núcleo sin extensiones.	Puede cargar varios archivos en un IPT como parte de un único conjunto de datos y asignarlos correctamente a un núcleo y al menos una extensión. Puede utilizar la función de valor constante del IPT.	Puede cargar varios archivos en un IPT como parte de un único conjunto de datos y asignarlos correctamente a un núcleo y a varias extensiones. Puede utilizar la función de traducción de datos del IPT.
E. Uso del IPT: capacidad de usar la herramienta para publicar y registrar conjuntos de datos.	Puede ver un conjunto de datos publicado y los metadatos asociados en un IPT. Puede descargar un archivo DwC-A de un IPT. Puede navegar por un conjunto de datos registrado desde el IPT al portal de GBIF.	Puede actualizar un conjunto de datos existente y publicado cargando un nuevo archivo de origen. Puede volver a publicar el archivo sin errores.	Puede publicar y registrar correctamente un nuevo conjunto de datos. Puede comprender y actuar sobre la publicación de mensajes de error en el IPT.	Muestra comprensión del control de versiones de conjuntos de datos en el IPT.

Caso de uso II: especies invasoras

Familiarícese con el escenario del caso de uso.

Para el caso de uso II puede elegir entre dos escenarios:

- Lista de especies invasoras
- Muestreo de Lepidoptera

Se calificará su elección para el caso de uso II.

Escenario

Seguimiento de especies invasoras



Leucaena leucocephala (Lam.) de Wit *observada en Hawaii por Sharon Grant (licenciada bajo* CC-BY-NC 4.0*)*

Esta narrativa se desarrolló como una base para ejercicios prácticos en el curso de movilización de datos de biodiversidad y el concepto y el contenido del ejercicio fueron desarrollados por Sharon Grant, John Wieczorek, David Bloom y Laura Anne Russell.

Es un escenario ficticio basado en un conjunto de datos reales y está destinado únicamente para fines educativos. El conjunto de datos original se atribuye a Simpson A (2016). Big Island Invasive Species Committee - Pest Reports - 2005-2010. Version 4.1. United States Geological Survey. Occurrence Dataset accedido a través de GBIF.org el 2017-07-13.

Descripción

El Hawaii Invasive Species Council (HISC) recibió una subvención federal para colaborar con las escuelas secundarias como parte del plan de estudios de todo el estado de Hawái sobre especies invasoras para aumentar el conocimiento local sobre las especies invasoras, aumentar la recopilación de datos y producir listas de verificación anotadas para áreas con reportes escasos. Se ha contratado a un director de proyecto a tiempo completo para supervisar el proyecto. Todos los fondos y asignaciones son administrados por el Asociado fiscal de HISC.

El Gerente del Invasive Species Committee (ISC) de cada isla recibió una subvención secundaria para establecer programas educativos locales y recopilar datos. Los programas formaron a estudiantes de secundaria para que se convirtieran en mentores de estudiantes y facilitaran la recopilación de imágenes y datos por parte de los miembros de la comunidad local. El Big Island Invasive Species Committee (BIISC) recibió una subvención secundaria adicional para ampliar su base de datos central para acomodar los datos de cada ISC, proporcionar a las escuelas participantes sus propios sitios web y mantener un único portal de datos con capacidad de búsqueda para servir a los esfuerzos de investigación gubernamentales, públicos y académicos.

Se seleccionaron dos escuelas en cada isla porque estaban ubicadas en áreas donde el conocimiento

y la documentación de la evaluación de especies invasoras era deficiente o inexistente. Los profesores trabajaron con su asociado local del Invasive Species Council (ISC) para crear material didáctico en el que se detallen 21 especies de plantas invasoras importantes, incluyendo cómo identificar las etapas de vida de cada especie y los métodos de control más eficaces.

La Escuela de Graduados de la Universidad de Hawái en Maui (UHM) imparte un curso de extensión comunitaria. Cuatro estudiantes de botánica de la Universidad, como parte de la evaluación final del curso, están validando las identificaciones de las imágenes y descripciones enviadas por cada escuela secundaria a su Consejo de Especies Invasoras (ISC) local.

Recopilación de datos

Los estudiantes de cada escuela secundaria organizaron una serie de estudios comunitarios de un día de duración en sus vecindarios. Los participantes, guiados por técnicos de detección temprana y mentores de estudiantes del ISC local, visitaron varios lugares donde se les entregaron guías fotográficas y se les asignó una ruta a seguir durante los eventos de recopilación. A lo largo de cada ruta, se les asignó la tarea de identificar las especies objetivo y tomar de 1 a 3 fotos de ellas utilizando teléfonos móviles con GPS.

Los detalles que describen cada observación de las 21 especies invasoras de interés, se registraron utilizando un formulario de recopilación de datos digitales durante cada evento de recopilación comunitaria. Los participantes subieron las imágenes capturadas en teléfonos móviles y se les animó a hacer clic en sus ubicaciones utilizando un mapa de Google, incrustado en el formulario, para asignar latitud y longitud a cada observación. El diseño del formulario se basó en el formulario de notificación de plagas de HISC.

REPORTER INFORMATION

Report Number: T07222017172139

First Name:

Last Name:

Email: e.g. pestreport@643pest.com

Phone: e.g. 808-123-4567 or 8081234567

PEST SIGHTING INFORMATION

NOTE: Asterisk [*] and red label color indicate a required field.

*Date of Pest Sighting:

*Pest Name:

*Pest Description:
(Plant: size; flower, foliage, or fruit color / scent / orientation; habitat)
(Insect / animal: size; color; plant / host found on or nearby; habitat)

*Island of Pest Sighting:
(Please choose an island before entering information into the Location field).

--Select Island--

Location of Pest Sighting:
(Street address, cross streets, city, mile marker, place name or general area)

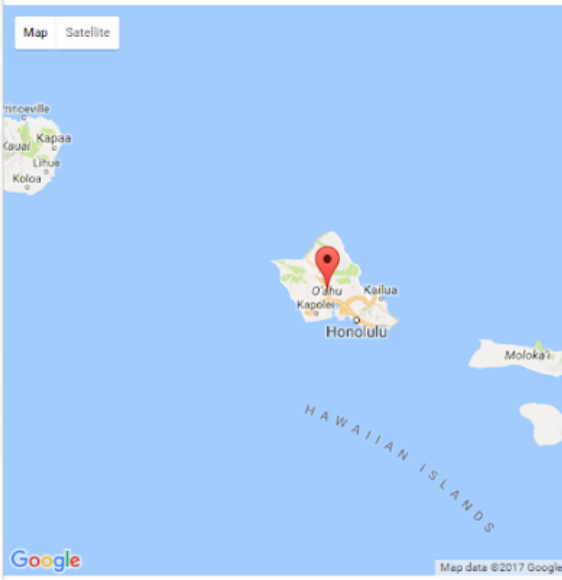
Additional Comments:

Image Upload (*Allowed File Types: .jpg, .png, .gif | Upload up to 3 images*):

Drag & drop files here ...

Browse ...

PEST SIGHTING LOCATION DETAILS



Directions:

1. The map will have automatically moved to the island chosen from the **Island** drop-down list above. If an island was not chosen or the incorrect island was chosen, please go back and make a new selection. Choosing an island is a **mandatory** step.

2. Use the **map tools** at left (Zoom In, Zoom Out, Pan) for help finding and zooming in to a desired area or location. Use the **Geo-location Search** box below to search for and zoom in to a specific geo-location (address, city, place name) in Hawaii.

3. To pinpoint the **exact location of the pest sighting**: a) click on and drag the red map marker to a particular location; or b) click on any particular location on the map to move the red map marker to that location. The **coordinates (Latitide, Longitude)** submitted with the pest report are shown in the boxes below and reflect the final position of the red map marker.

Geo-location Search:

Enter an address, city, place name...

Marker Coordinates:

Latitude

Longitude

Descripción de datos digitales

Una base de datos, creada y alojada por el Departamento de computación de la UHM, mantiene las imágenes y los datos del formulario en línea, pero estos datos no son de acceso público. Los datos se exportaron como archivos de valores separados por comas (.csv) y se entregaron a los cuatro estudiantes de posgrado de la UHM para su validación taxonómica utilizando las imágenes y descripciones enviadas. El analista SIG de BIISC utilizó las coordenadas de Google Maps y los datos EXIF de la imagen para verificar la calidad de las observaciones y agregar las georreferencias que faltaban. Los estudiantes mentores renombraron todos los archivos de imágenes para que coincidieran con el número de observación para su posterior referencia cruzada en el BIISC.

Hoja de ejercicios de Invasoras

Descargar exercise sheet. (MS Word, 342 KB)

Ejercicio 1

Planificación

Usted es el responsable local del ISC y, debido al éxito de los planes de estudio y los estudios comunitarios, otras 10 escuelas de su isla quieren poner en marcha sus propios proyectos al año siguiente. Le gustaría darles cabida, pero su financiación del ISC vencerá a fines de este año. HISC ha indicado que verán favorablemente una pequeña solicitud de subvención para ampliar sus programas el año siguiente y el BIISC ha ofrecido su apoyo.

Ejercicio 1a

Analizar las implicaciones financieras de la ampliación del número de escuelas

1. Evalúe las siguientes opciones para ampliar el número de escuelas participantes. Solo puede seleccionar DOS de estas opciones, por lo que debe elegir sabiamente.
2. Utilice la hoja de ejercicios para proporcionar sus respuestas.

Opciones

1. Pagar extra a los pasantes de verano para trabajar en el ISC local para coordinar los estudios.
2. Ofrecer apoyo financiero a BIISC para crear sitios web para cada escuela nueva.
3. Ofrecer una compensación económica a los estudiantes de posgrado. No podrá pagarles a los cuatro el equivalente a un salario regular, pero podría cubrir los costos de los puestos a medio tiempo de dos de ellos.
4. Contratar a una empresa de software para que construya una base de datos que pueda introducir automáticamente los datos directamente desde el formulario en línea. El sistema incluirá una interfaz de administración que permita manipular los datos y exportarlos en formato csv.
5. Financiar cuatro actividades de divulgación pública (p. ej., un BioBlitz) para promover la sensibilización en las comunidades y aumentar la participación de los voluntarios.
6. Preparar y llevar a cabo un curso de formación reutilizable para los profesores de los centros educativos con el fin de enseñarles a preparar los datos para su envío al BIISC.

Ejercicio 1b

Asignar roles

El nuevo proyecto cuenta con las siguientes personas disponibles para el procesamiento y la movilización de datos.

1. ¿Asignar roles para maximizar la eficiencia del procesamiento y la transformación de datos para producir datos de la más alta calidad de la manera más eficiente posible?
2. Utilice la hoja de ejercicios para proporcionar sus respuestas.

Roles

- Analista SIG de BIISC: uso avanzado de computadoras, herramientas de análisis de datos y SIG.
- Administrador de ISC: buenas habilidades informáticas.
- Asociado de divulgación del ISC: buenas habilidades de identificación de campo, uso básico de computadoras. Experto en redes sociales.
- Estudiante mentor 1: conocimiento taxonómico básico. Uso básico de computadoras.
- Estudiante mentor 2: conocimiento taxonómico básico. Uso básico de computadoras.
- Estudiante de botánica 1: conocimiento taxonómico avanzado. Habilidades de programación.
- Estudiante de botánica 2: conocimiento taxonómico avanzado.
- Estudiante de botánica 3: conocimiento taxonómico avanzado.
- Estudiante de Bbtánica 4: conocimiento taxonómico avanzado.

Ejercicio 2

Captura de datos

El BIISC está planeando poner a disposición del público todos los datos del proyecto mediante la publicación de conjuntos de datos en GBIF. Como Asociado de divulgación de BIISC, usted debe identificar los campos de Darwin Core relevantes para acomodar los datos del formulario en línea. Se ha dado cuenta de que los estudiantes de postgrado que realizan las validaciones agregaron datos adicionales que describen especies y ubicaciones al formulario de datos. Para acomodar estos datos, debe ampliar la estructura de datos para agregar los datos del formulario en línea con la taxonomía y las georreferencias agregadas.

1. Descargue UC2-IS-2-ForCapture.csv. (7.1 MB)
2. Usando el conjunto de datos descargado, genere una hoja de cálculo como ejemplo de la estructura de datos extendida y de los campos que ha identificado como relevantes para Darwin Core.
3. Use la hoja de ejercicios para proporcionar sus respuestas y envíe la hoja de cálculo.

Ejercicio 3

Gestión de datos

Durante el verano, los pasantes de la oficina principal de HISC crearon listas de especies a partir de los datos de registros biológicos originales que se recopilaron y ampliaron del formulario en línea. Al asumir el papel de líder del proyecto HSC, ahora debe realizar controles de calidad finales antes de la

publicación.

1. Descargue UC2-IS-3-ForCleaning.xlsx. (156 MB)
2. Evalúe el conjunto de datos e identifique qué tipos de errores están presentes.
3. Identifique posibles formas de corregir esos problemas y realice las correcciones para la mayor cantidad de errores que pueda.
4. Use la hoja de ejercicios para proporcionar sus respuestas y envíe la hoja de cálculo.

Ejercicio 4

Publicación de datos

El HISC ahora está listo para publicar los datos de la lista de especies y los registros biológicos asociados en GBIF. Para este ejercicio, asumirá el papel de líder del proyecto. Su responsabilidad es: publicar los datos de la lista de especies limpia y los registros biológicos asociados en línea a través de la red GBIF.

1. Descargue UC2-IS-4-ForPublication.xlsx. (99 KB)
2. Utilice la instalación IPT previamente proporcionada para publicar el conjunto de datos dado.
3. Utilice la hoja de ejercicios para proporcionar sus respuestas y el enlace al conjunto de datos publicado.

Caso de uso II: avistamiento de lepidópteros



Familiarícese con el escenario del caso de uso.

Para el caso de uso II puede elegir entre dos escenarios:

- Lista de especies invasoras
- Muestreo de Lepidoptera

Se calificará su elección para el caso de uso II.

Escenario

Muestreo de lepidópteros en todos los países



Papilio machaon Linnaeus, 1758 *observada en Israel por רניו רמוע (licenciada bajo* CC-BY-NC 4.0*)*

Esta narrativa fue desarrollada como base para ejercicios prácticos en el curso de movilización de datos de biodiversidad y el concepto y contenido del ejercicio fue desarrollado por Alberto González-Talaván, basado en trabajos previos de Alberto González-Talaván, Danny Vélez, Larissa Smirnova, Laura Russell, Mélianie Raymond y Nicolas Noé. Es un escenario ficticio y solo tiene fines educativos.

Descripción

La Red Internacional de Aficionados de Mariposas (IBAN) ha proporcionado un marco para que los grupos nacionales de observación de aficionados capturen datos sobre registros biológicos de mariposas (Lepidoptera) desde 2009. Una extensa red de observadores aficionados utiliza un protocolo estándar basado en las caminatas de Pollard para capturar esta información en hojas de papel que envían a su oficina nacional. Algunas de estas oficinas digitalizan esta información en hojas de cálculo, pero otras no cuentan con los recursos humanos para hacerlo y envían los registros en papel al IBAN para su procesamiento. IBAN produce un informe anual basado en los avistamientos proporcionados por estos miembros nacionales, con mapas de distribución actualizados y análisis de tendencias poblacionales para algunas especies clave.

La sede central de IBAN está compuesta principalmente por voluntarios. Con la creciente popularidad de la ciencia ciudadana y el interés general en las mariposas como un grupo carismático de organismos, cada año se reciben más y más datos y las hojas de datos en papel sin digitalizar se apilan rápidamente. El comité directivo de IBAN está tratando de identificar un flujo de trabajo más eficiente y ágil para la creación de datos digitales porque les gustaría comenzar a publicar estos datos en línea con regularidad. También les gustaría comenzar a procesar imágenes digitales que sus voluntarios ya están capturando con dispositivos móviles como teléfonos y tabletas. Su objetivo final es dar a conocer la red y fortalecer las colaboraciones con los gobiernos locales y regionales para influir en las políticas de conservación de los lepidópteros en los países involucrados.

Actualmente, no existe un acuerdo formal entre IBAN y los aficionados que capturan datos, que contemple, por ejemplo, las formas en que se pueden usar los datos. El comité directivo tiene algunas preocupaciones de que cuando comiencen a publicar los datos en línea, tendrán que formalizar este acuerdo.

Recopilación de datos

El protocolo recomendado (la caminata de Pollard) se basa en transectos que oscilan entre 300 y 600 m de longitud, divididos en tramos de 50 m. Cada transecto debe cubrir un solo tipo de hábitat.

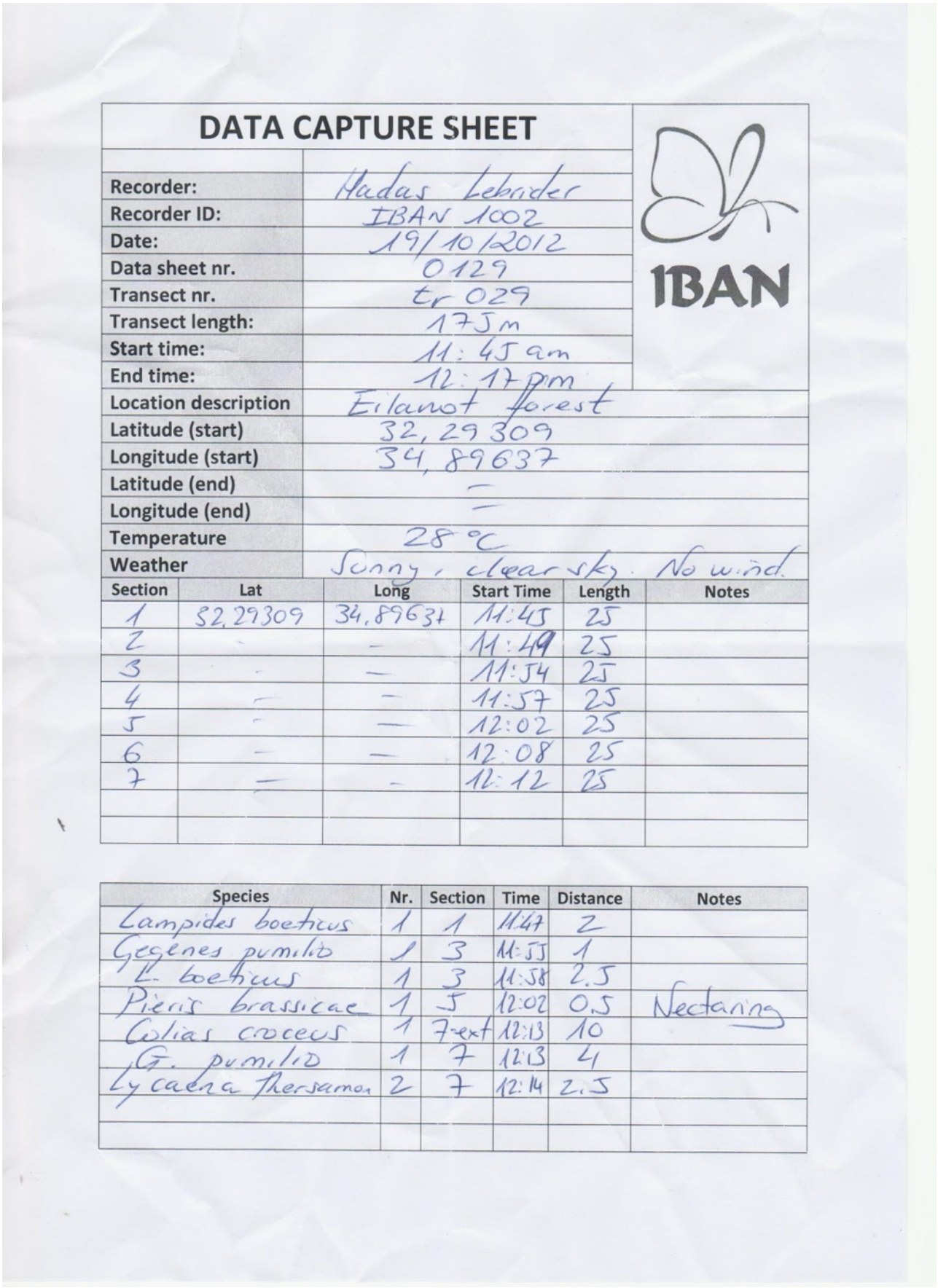
En cada visita, los caminantes del transecto deben contar todas las especies de lepidópteros que puedan ver en un radio de 5 m de la línea del transecto. También deben registrar los comportamientos especiales (puesta de huevos o libación de néctar), así como la etapa de desarrollo (p. ej., larvas o huevos).

En la mayoría de los países, estos esfuerzos de muestreo se realizan una vez cada dos semanas desde principios de octubre hasta finales de junio.

Existen medidas de control de calidad: cada registro notificado se marca como "Pendiente de aprobación". El estado del registro solo se cambia a "Aprobado" después de la verificación por parte de un experto taxonómico designado. Las especies detectadas fuera de su temporada o área de distribución regular se marcan para verificación adicional.

La hora del día y las condiciones climáticas se registran al comienzo del transecto. A lo largo del transecto, se cuenta el número de individuos de cada especie vista. Las especies no identificadas se cuentan y registran por familia o como un complejo predefinido de dos o tres especies similares. Las mariposas que se ven fuera del rango de 5 metros se registran como "Extra+el número de la sección más cercana" (p. ej., 5-extra). También se registra la hora de finalización del transecto.

Ejemplo de captura de datos análoga



DATA CAPTURE SHEET

IBAN

Recorder:	Hadas Lebiush
Recorder ID:	IBAN 1002
Date:	19/10/2012
Data sheet nr.	0129
Transect nr.	tr 029
Transect length:	175 m
Start time:	11:45 am
End time:	12:17 pm
Location description	Eilanot forest
Latitude (start)	32,29309
Longitude (start)	34,89637
Latitude (end)	–
Longitude (end)	–
Temperature	28 °C
Weather	Sunny, clear sky. No wind.

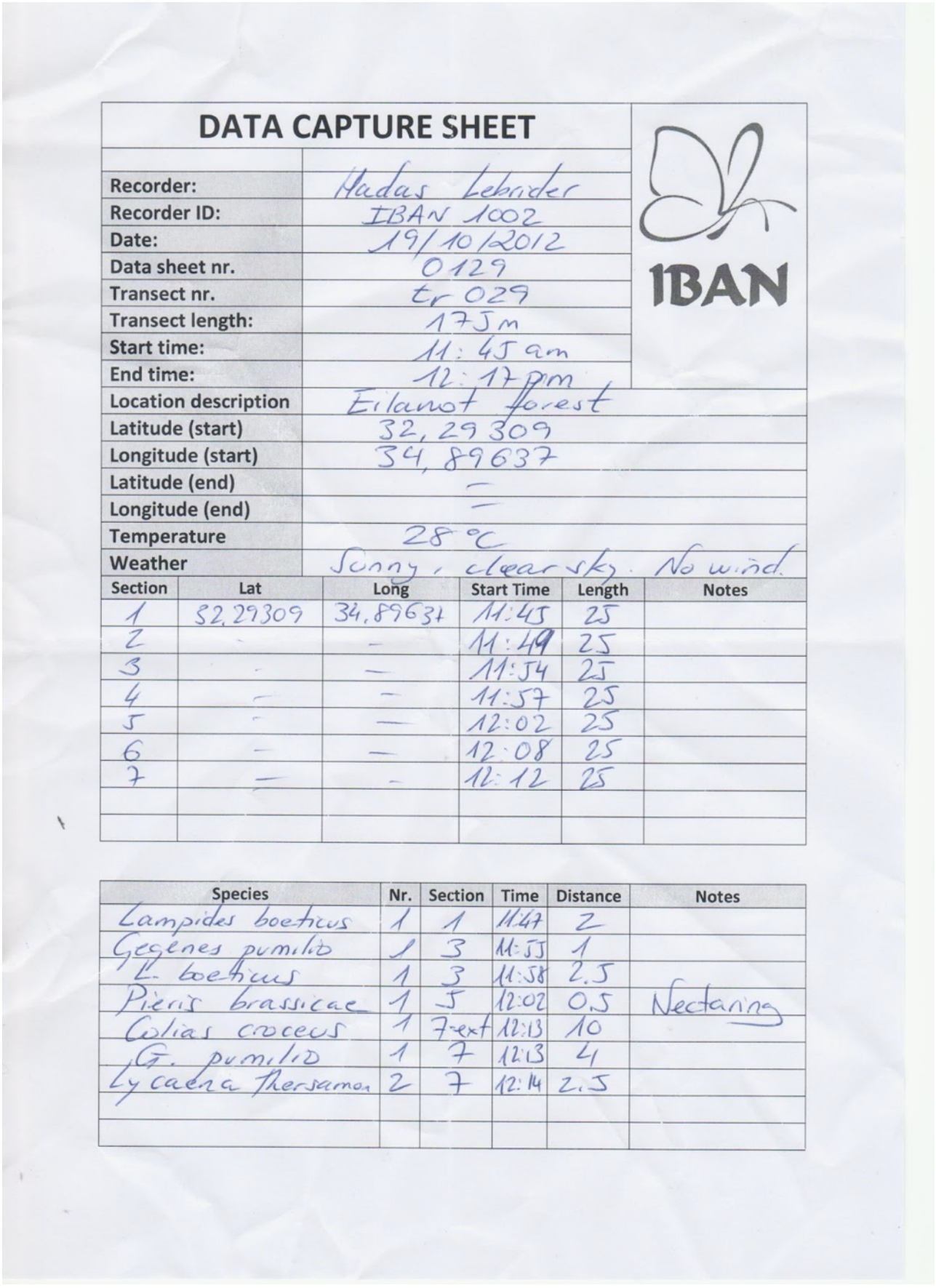
Section	Lat	Long	Start Time	Length	Notes
1	32,29309	34,89637	11:45	25	
2	–	–	11:49	25	
3	–	–	11:54	25	
4	–	–	11:57	25	
5	–	–	12:02	25	
6	–	–	12:08	25	
7	–	–	12:12	25	

Species	Nr.	Section	Time	Distance	Notes
Lampides boeticus	1	1	11:47	2	
Gegenes pumilio	1	3	11:55	1	
L. boeticus	1	3	11:58	2.5	
Pieris brassicae	1	5	12:02	0.5	Nectaring
Colias croceus	1	7-ext	12:13	10	
G. pumilio	1	7	12:13	4	
Lycaena thersamon	2	7	12:14	2.5	

Descripción de datos digitales

Algunas oficinas nacionales utilizan grupos de voluntarios para digitalizar los registros que están en papel y producir hojas de cálculo digitales. Las hojas de cálculo son muy simples e incluyen tres hojas de datos. Una captura la información relacionada con los esfuerzos de muestreo, la segunda las

condiciones climáticas y la tercera las especies encontradas y el número de individuos observados por el aficionado.



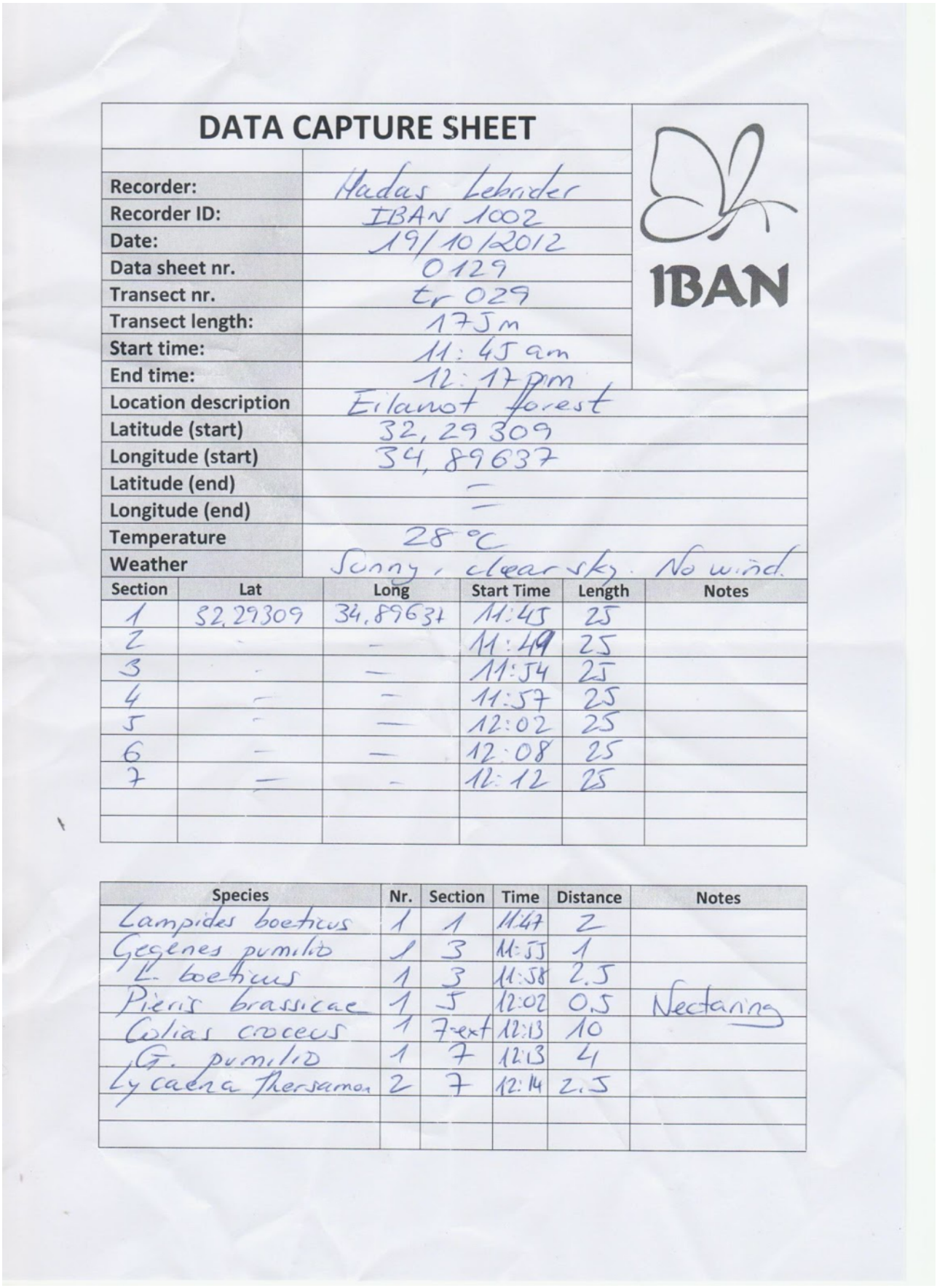
DATA CAPTURE SHEET

IBAN

Recorder:	Hadas Lebrider
Recorder ID:	IBAN 1002
Date:	19/10/2012
Data sheet nr.	0129
Transect nr.	tr 029
Transect length:	175 m
Start time:	11:45 am
End time:	12:17 pm
Location description	Eilanot forest
Latitude (start)	32,29309
Longitude (start)	34,89637
Latitude (end)	–
Longitude (end)	–
Temperature	28 °C
Weather	Sunny, clear sky. No wind.

Section	Lat	Long	Start Time	Length	Notes
1	32,29309	34,89637	11:45	25	
2	–	–	11:49	25	
3	–	–	11:54	25	
4	–	–	11:57	25	
5	–	–	12:02	25	
6	–	–	12:08	25	
7	–	–	12:12	25	

Species	Nr.	Section	Time	Distance	Notes
Lampides boeticus	1	1	11:47	2	
Gegenes pumilio	1	3	11:55	1	
L. boeticus	1	3	11:58	2.5	
Pieris brassicae	1	5	12:02	0.5	Nectaring
Colias croceus	1	7-ext	12:13	10	
G. pumilio	1	7	12:13	4	
Lycaena thersamon	2	7	12:14	2.5	



DATA CAPTURE SHEET

IBAN

Recorder:	Hadas Lebiush
Recorder ID:	IBAN 1002
Date:	19/10/2012
Data sheet nr.	0129
Transect nr.	tr 029
Transect length:	175 m
Start time:	11:45 am
End time:	12:17 pm
Location description	Eilanot forest
Latitude (start)	32,29309
Longitude (start)	34,89637
Latitude (end)	–
Longitude (end)	–
Temperature	28 °C
Weather	Sunny, clear sky. No wind.

Section	Lat	Long	Start Time	Length	Notes
1	32,29309	34,89637	11:45	25	
2	–	–	11:49	25	
3	–	–	11:54	25	
4	–	–	11:57	25	
5	–	–	12:02	25	
6	–	–	12:08	25	
7	–	–	12:12	25	

Species	Nr.	Section	Time	Distance	Notes
Lampides boeticus	1	1	11:47	2	
Gegenes pumilio	1	3	11:55	1	
L. boeticus	1	3	11:58	2.5	
Pieris brassicae	1	5	12:02	0.5	Nectaring
Colias croceus	1	7-ext	12:13	10	
G. pumilio	1	7	12:13	4	
Lycaena thersamon	2	7	12:14	2.5	

Hoja de ejercicios Lepidoptera

Descargue la hoja de ejercicios. (MS Word, 342 KB)

Ejercicio 1

Planificación

El volumen de datos análogos (registros en papel) que llegan a la sede del IBAN pronto superará su capacidad de digitalización, por lo que el comité directivo ha decidido reconsiderar el enfoque actual de esta área de su trabajo, que ha crecido sin control durante los últimos años. Hasta la fecha, así es como se ha organizado el trabajo:

- Los registros en papel llegan por correo. El secretario abre los paquetes y recopila los registros.
- Hay cinco voluntarios con habilidades informáticas básicas que utilizan dos computadoras compartidas para digitalizar los registros en papel. Estos voluntarios también son científicos ciudadanos, por lo que están familiarizados con la taxonomía del orden Lepidoptera y con las especies que se encuentran en el país donde se encuentra la sede del IBAN.
- Los digitalizadores van y vienen siempre que tienen tiempo, por lo que generalmente verifican la disponibilidad de computadoras por teléfono. A veces hay conflictos de tiempo y algunos tienen que irse a casa porque las dos computadoras están ocupadas y, a veces, ninguna de las dos computadoras se utiliza.
- Cuando digitalizan, generalmente toman un registro de papel a la vez de la pila y lo digitalizan (si pueden). Los problemas más comunes que ocurren son los siguientes:
  - el digitalizador no conoce la especie observada (se producen errores ortográficos),
  - el digitalizador no conoce el área donde se ha realizado el muestreo,
  - el digitalizador no puede leer la escritura a mano o el idioma en el que están escritos algunos de los comentarios.
- Un solo experto taxonómico obtiene todas las tablas digitalizadas y produce el informe y los mapas de distribución basados en ellas. Normalmente, necesita descartar alrededor del 15% de los datos digitalizados debido a inconsistencias, errores ortográficos u otros errores que no tiene tiempo de verificar.

Ejercicio 1a

Analizar el componente financiero de su nuevo plan de digitalización

El comité directivo está analizando las siguientes opciones para su nuevo plan de digitalización, todas las cuales tienen implicaciones financieras en su ya reducido presupuesto. Saben que solo pueden implementar DOS de estas opciones, por lo que deben elegir sabiamente. Use la hoja de ejercicios para brindar una recomendación sobre las dos opciones que deben seleccionar y explique por qué las eligió.

1. Opción 1: comprar tres computadoras más, para que todos los digitalizadores puedan trabajar simultáneamente.
2. Opción 2: ofrecer apoyo financiero a las oficinas nacionales para comprar escáneres de superficie plana y enviar/compartir los registros electrónicamente en lugar de por correo.
3. Opción 3: ofrecer una compensación económica a los digitalizadores. No podrán pagarles a los cinco el equivalente a un salario regular, pero podrían cubrir los costos de los puestos de medio tiempo para tres de los voluntarios.
4. Opción 4: Adquiera el software de digitalización de la biodiversidad existente en inglés, que viene con verificación de entrada taxonómica y ayudas integradas para corregir la información geográfica.

5. Opción 5: contratar a una empresa de desarrollo de software para desarrollar software de digitalización personalizado. Por el mismo precio del software comercial, los desarrolladores proporcionarán una solución en el idioma local, que coincidirá perfectamente con el esquema de datos original y también proporcionarán un portal de datos web para exponer los resultados del esfuerzo de digitalización.
6. Opción 6: Organizar un curso para que los cinco digitalizadores mejoren sus habilidades en taxonomía, el uso de computadoras y estándares de informática de la biodiversidad.

Ejercicio 1b

Asignar roles

Estos son los recursos humanos disponibles para este esfuerzo de digitalización. ¿Cómo asignaría los roles para maximizar la eficiencia del proceso de digitalización y producir datos de la mejor calidad posible? Utilice la hoja de ejercicios para proporcionar sus respuestas.

1. Asistente administrativo. Sin conocimientos taxonómicos. Uso básico de la computadora. Puede leer 3 idiomas.
2. Voluntario 1. Conocimientos taxonómicos básicos. Uso básico de la computadora.
3. Voluntario 2. Conocimientos taxonómicos básicos. Uso básico de la computadora.
4. Voluntario 3. Conocimientos taxonómicos básicos. Uso básico de la computadora. Puede leer 3 idiomas.
5. Voluntario 4. Conocimientos taxonómicos básicos. Uso básico de la computadora. Puede leer 3 idiomas.
6. Voluntario 5. Conocimientos taxonómicos básicos. Uso avanzado de computadoras (incluyendo herramientas de análisis de datos y SIG).
7. Experto taxonómico. Conocimientos taxonómicos avanzados. Uso avanzado de computadoras (incluyendo herramientas de análisis de datos y SIG).

Ejercicio 2

Captura de datos

Imagine que es uno de los voluntarios que digitalizan los registros en papel recibidos en la sede del IBAN. Ha recibido dos registros en papel.

1. Descargar los registros 1 y 2 UC2-LS-2-ForCapture.zip. (943 KB)
2. ¿Qué estructura de datos usaría para reflejar los datos en estos registros?
3. Cree una hoja de cálculo utilizando esta estructura y los datos de los registros.
4. Use the exercise sheet to provide your answers and submit the spreadsheet created in the previous step.

Ejercicio 3

Gestión de datos

Tomando el rol de uno de los voluntarios con habilidades informáticas avanzadas, imagine que se le ha asignado la responsabilidad de los asuntos de calidad de los datos. Su tarea principal es reducir la cantidad de datos que se descartan actualmente (alrededor de un 15%) antes de procesarlos debido a

errores e inconsistencias. Ha recibido un conjunto de datos como producto bruto del esfuerzo de digitalización.

1. Descargue los datos UC2-LS-3-ForCleaning.xlsx. (44 KB)
2. Evalúe el conjunto de datos e identifique qué tipos de errores están presentes.
3. Identifique posibles formas de corregir esos problemas y realice esas correcciones para el mayor número de errores que pueda.
4. Use la hoja de ejercicios para proporcionar sus respuestas y envíe la hoja de cálculo.

Ejercicio 4

Publicación de datos

Para este ejercicio asumirá el rol de experto taxonómico que colabora con IBAN en su sede. Algunas de sus responsabilidades anteriores (redactar el informe anual y producir los mapas de distribución base) se han transferido a los voluntarios, y ahora se le ha dado una nueva responsabilidad: publicar los datos limpios en línea a través de la red GBIF. El voluntario a cargo de la calidad de los datos ha proporcionado un conjunto de datos para su publicación.

1. Descargue los datos UC2-LS-4-ForPublication.xlsx. (58 KB)
2. Utilice la instalación IPT previamente proporcionada para publicar el conjunto de datos dado.
3. Utilice la hoja de ejercicios para proporcionar sus respuestas y el enlace al conjunto de datos publicado.

Caso de uso III: aves de la literatura

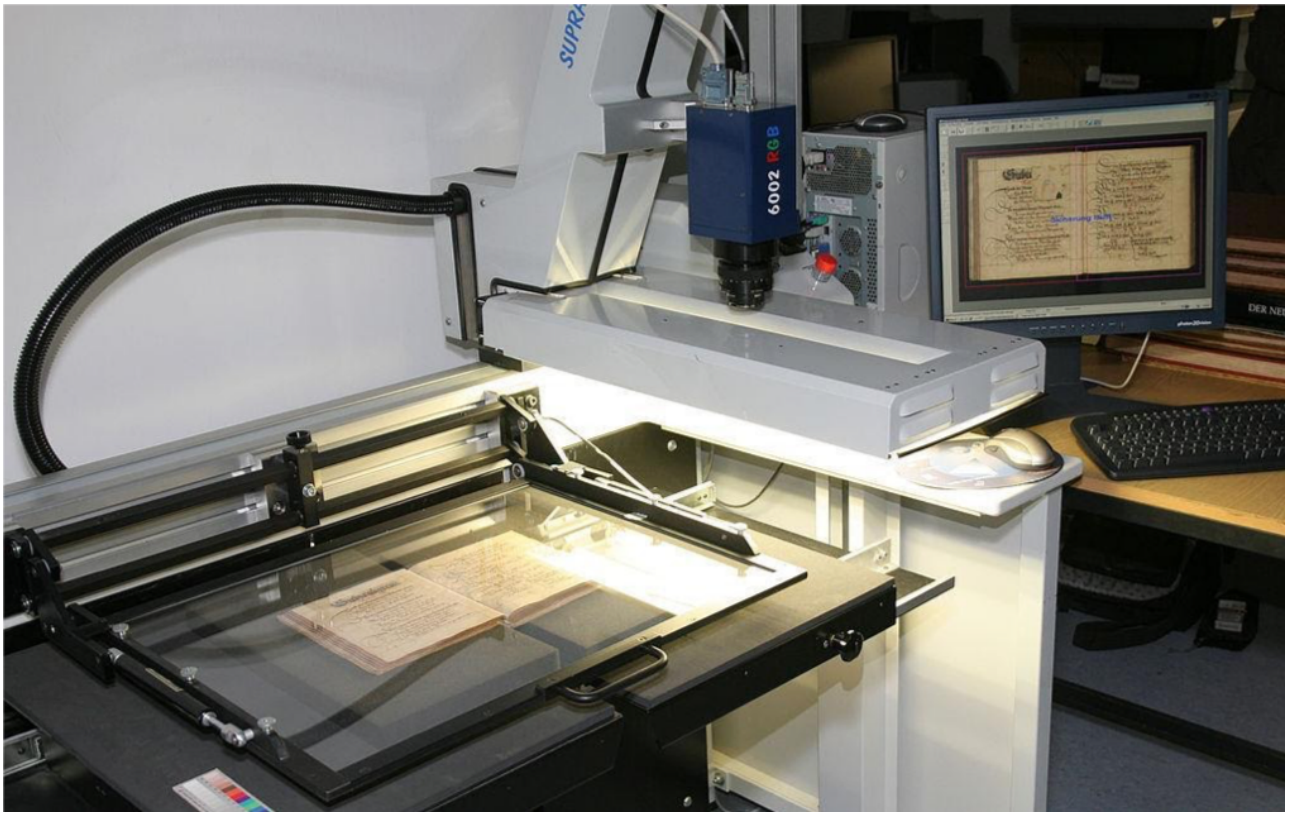


Familiarícese con el escenario del caso de uso.

El caso de uso III será calificado.

Escenario

Proyecto de movilización de datos de la literatura "Aves caídas en los faros daneses, 1883–1939"



Escáner de alta resolución *para el proyecto de digitalización de libros de Heiko Hornig (licenciado bajo* CC BY-SA 2.5*)*

Esta narrativa fue desarrollada como base para ejercicios prácticos del curso de movilización de datos de biodiversidad y el concepto y contenido del ejercicio fue desarrollado por Alberto González-Talaván, Andrea Hahn, Laura Russell y Sharon Grant. Se basa en una adaptación previa de Alberto González-Talaván, Danny Vélez, Larissa Smirnova, Laura Russell, Mélianie Raymond y Nicolas Noé.

Es un escenario ficticio basado en un proyecto y en un conjunto de datos reales y está destinado a fines educativos únicamente. Se atribuyen el proyecto original y el conjunto de datos original al nodo GBIF danés, DanBIF.

Descripción

El Museo de Historia Natural de Dinamarca (NHM-DK) es un centro de investigación asociado con la Universidad de Copenhague. Su biblioteca es miembro de la asociación nacional de bibliotecas que recientemente recibió fondos estatales para poner en línea los recursos que poseen sus miembros. El NHM-DK desea comenzar a digitalizar los cuadernos de campo, las publicaciones de revistas y los libros que se encuentran en su biblioteca, algunos de los cuales tienen un valor histórico significativo.

Después de una breve consulta con sus socios habituales, NHM-DK recibió una sugerencia del director de la oficina de gestión del Parque Nacional Nordjylland. Les gustaría que se digitalizara el contenido de una recopilación de literatura clásica particular para un proyecto que están ejecutando: "Aves en los faros daneses, 1883-1939" (en danés, "Fuglene ved de danske Fyr, 1883-1939"). Quieren utilizar cualquier dato de registros biológicos consignado en los libros de dos faros (Lodbjerg Fyr y Hanstholm Fyr) para un proyecto de exhibición en el lugar.

El NHM-DK ha iniciado conversaciones con el nodo nacional de GBIF, DanBIF, sobre la movilización de la información contenida en estos volúmenes, es decir, para preservar sus contenidos para el futuro y proporcionar acceso en línea para todos. Con la participación de DanBIF, existe la intención de publicar y registrar los datos extraídos resultantes en GBIF. Como GBIF requiere que se aplique una licencia con todos los datos publicados, el museo ha decidido publicar los datos con una licencia

Creative Commons que permite el uso de datos con atribución (CC-BY).

Los servicios informáticos necesarios son proporcionados por la Unidad de Tecnología de la Universidad de Copenhagen, como para todos los proyectos digitales de los museos.

El subdirector del HHM-DK, que está coordinando este trabajo, ha elaborado un esquema general del mismo:

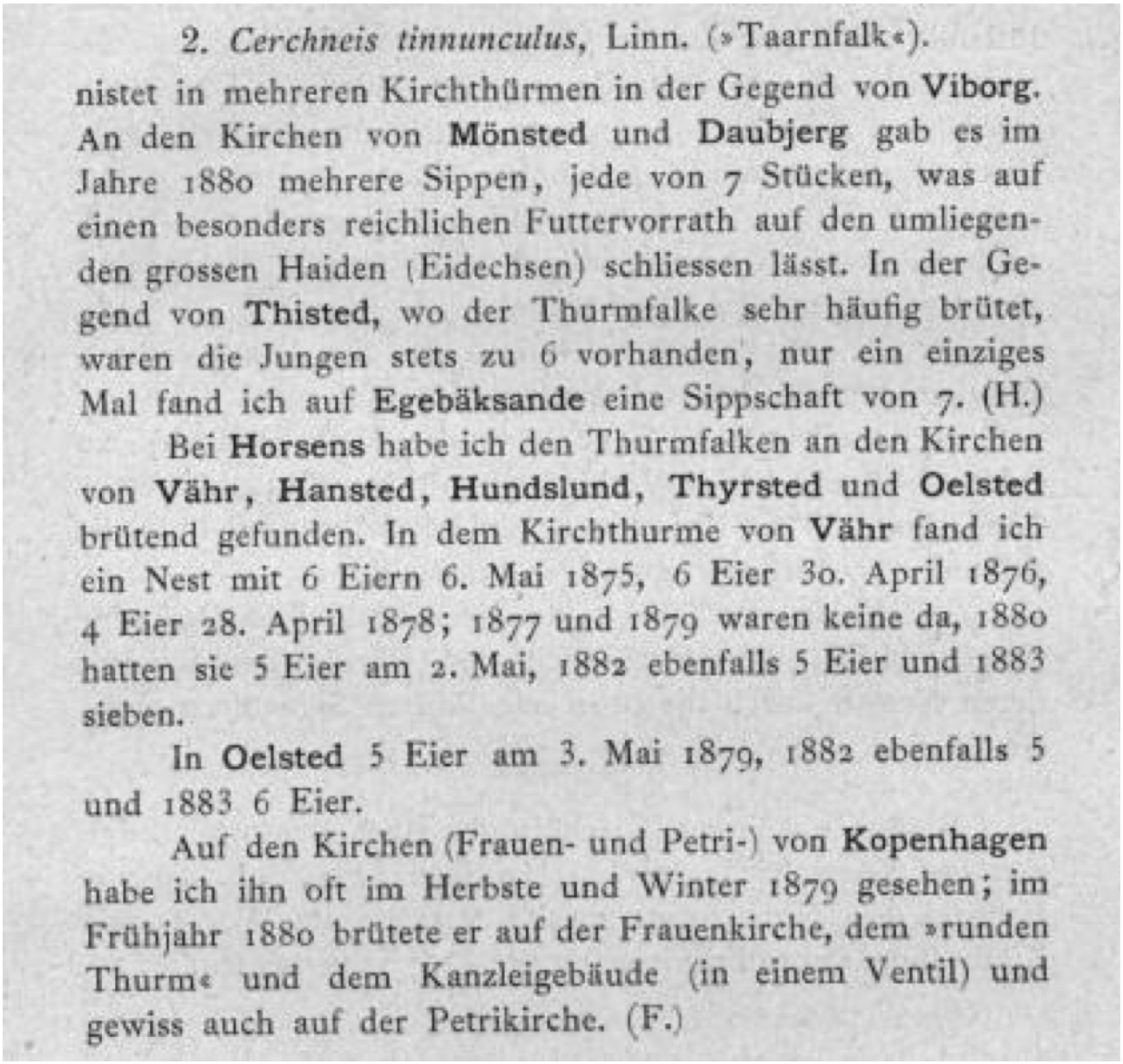
1. El museo llevará a cabo la digitalización de la literatura mediante dos miembros del personal de la biblioteca, capacitados en el uso del escáner de la biblioteca, para escanear volúmenes delicados. También extraerán el texto de los escaneos mediante un software de reconocimiento óptico de caracteres (OCR).
2. Tres voluntarios de la Sociedad Ornitológica de Copenhague (COS), que colaboran regularmente con el museo y están familiarizados con las aves de la región, han sido reclutados para ayudar y completarán la transferencia de datos de los PDF escaneados al formato de hoja de cálculo. Deberán ir al museo y utilizar las computadoras disponibles en la biblioteca para acceder a los archivos almacenados en la intranet del museo (red privada).
3. El curador de ornitología del Departamento de Aves del NHM-DK dirigirá el equipo responsable de la verificación taxonómica, la curación, la limpieza, el formato y la transformación de los datos, y supervisará la introducción de metadatos para el conjunto de datos publicado. El equipo incluye un investigador colaborador de Suecia y dos estudiantes posdoctorales. Han sido seleccionados para esta tarea porque están acostumbrados a trabajar con datos sobre biodiversidad. Todos ellos utilizarán sus propios computadores de trabajo.
4. El administrador de nodo de DanBIF se asegurará de que la institución esté debidamente registrada en GBIF como publicador de datos, y de que el director adjunto y el curador de ornitología tengan las credenciales adecuadas y el acceso a la instancia de IPT de DanBIF para cargar y publicar los datos.

Recopilación de datos originales

En el período 1883-1939, había 45 faros y buques faro en funcionamiento en Dinamarca. Estos faros fueron utilizados por varias especies de aves durante las noches del período de migración de aves desde los años 1886 hasta 1939. La presencia y las actividades de estas aves fueron registradas, especialmente por los guardianes de estos faros que también colectaban especímenes que se enviaban al museo de Copenhagen. Estas aves fueron cuidadosamente conservadas y catalogadas por los encargados de la colección en el museo y los especímenes aún se pueden encontrar allí hoy en día. También se documentaron las observaciones de las condiciones meteorológicas durante las noches en las que los cuidadores observaron las aves.

Descripción de datos análogos

Este es un ejemplo de la descripción de una serie de observaciones de especies de uno de los libros (en alemán, excepto el nombre común de la especie que se proporciona en danés).



2. *Cerchneis tinnunculus*, Linn. (»Taarnfalk«).

nistet in mehreren Kirchthürmen in der Gegend von **Viborg**. An den Kirchen von **Mönsted** und **Daubjerg** gab es im Jahre 1880 mehrere Sippen, jede von 7 Stücken, was auf einen besonders reichlichen Futtervorrath auf den umliegenden grossen Haiden (Eidechsen) schliessen lässt. In der Gegend von **Thisted**, wo der Thurmfalke sehr häufig brütet, waren die Jungen stets zu 6 vorhanden, nur ein einziges Mal fand ich auf **Egebäksande** eine Sippschaft von 7. (H.)

Bei **Horsens** habe ich den Thurmfalken an den Kirchen von **Vähr**, **Hansted**, **Hundslund**, **Thyrsted** und **Oelsted** brütend gefunden. In dem Kirchthurme von **Vähr** fand ich ein Nest mit 6 Eiern 6. Mai 1875, 6 Eier 30. April 1876, 4 Eier 28. April 1878; 1877 und 1879 waren keine da, 1880 hatten sie 5 Eier am 2. Mai, 1882 ebenfalls 5 Eier und 1883 sieben.

In **Oelsted** 5 Eier am 3. Mai 1879, 1882 ebenfalls 5 und 1883 6 Eier.

Auf den Kirchen (Frauen- und Petri-) von **Kopenhagen** habe ich ihn oft im Herbste und Winter 1879 gesehen; im Frühjahr 1880 brütete er auf der Frauenkirche, dem »runden Thurm« und dem Kanzleigebäude (in einem Ventil) und gewiss auch auf der Petrikirche. (F.)

Descripción de datos escaneados y traducidos

Este es un ejemplo de la salida escaneada y traducida del ejemplo análogo anterior.

Output of the OCR software	Translation into English
2. *Cerchneis tinnunculus*, Linn. (»Taarnfalk«) nistet in mehreren Kirchthürmen in der Gegend von **Viborg**. An den Kirchen von **Mönsted** und **Daubjerg** gab es im Jahre 1880 mehrere Sippen, jede von 7 Stücken, was auf einen besonders reichlichen Futtervorrath auf den umliegenden grossen Haiden (Eidechsen) schliessen lässt. In der Gegend von **Thisted**, wo der Thurmfalke sehr häufig brütet, waren die Jungen stets zu 6 vorhanden, nur ein einziges Mal fand ich auf **Egebäksande** eine Sippschaft von 7. (H.) Bei **Horsens** habe ich den Thurmfalken an den Kirchen von **Vähr**, **Hansted**, **Hundslund**, **Thyrsted** und **Oelsted** brütend gefunden. In dem Kirchthurme von **Vähr** fand ich ein Nest mit 6 Eiern 6. Mai 1875, 6 Eier 30. April 1876, 4 Eier 28. April 1878; 1877 und 1879 waren keine da, 1880 hatten sie 5 Eier am 2. Mai, 1882 ebenfalls 5 Eier und 1883 sieben. In **Oelsted** 5 Eier am 3. Mai 1879, 1882 ebenfalls 5 und 1883 6 Eier. Auf den Kirchen (Frauen- und Petri-) von **Kopenhagen** habe ich ihn oft im Herbste und Winter 1879 gesehen; im Frühjahr 1880 brütete er auf der Frauenkirche, dem »runden Thurm« und dem Kanzleigebäude (in einem Ventil) und gewiss auch auf der Petrikirche. (F.)	2. *Cerchneis tinnunculus*, Linn. ("Taarnfalk") nests in several steeples around **Viborg**. At the churches of **Mönsted** and **Daubjerg** there were several family groups in 1880, each of 7 individuals, suggesting a particularly abundant source of food on the surrounding heather (lizards). In the area of **Thisted**, where the tower falcon broods very often, young were always present in broods of 6, only once did I find a group of 7 on **Egebäksande** (H.) In **Horsens** I found kestrels brooding on the churches of **Vähr**, **Hansted**, **Hundslund**, **Thyrsted** and **Oelsted**. In the steeple of **Vähr** I found a nest with 6 eggs on 6 May 1875, 6 eggs on 30 April 1876, 4 eggs on 28 April 1878; in 1877 and 1879 there were none, on 2 May 1880 they had 5 eggs, in 1882 also 5 eggs, and in 1883 seven. In **Oelsted** 5 eggs on 3 May 1879, in 1882 also 5, and in 1883 6 eggs. On the churches (Our Lady's and St. Peter's) of **Copenhagen**, I have often seen it in the autumn and winter of 1879; in spring 1880, it brooded on Our Lady's church, the "round tower" and the law firm building (in a valve) and certainly also on St. Peter's Church. (F.)

Descripción de datos digitales

Al estudiar el extracto del libro, los voluntarios de la Sociedad de Ornitología de Copenhague sugieren extraer los siguientes datos del texto escaneado y traducido:

- Nombre científico que aparece en el libro
- Nombre(s) común(es) en danés que aparece(n) en el libro
- Localidad
- Año/mes/día
- Número observado de individuos
- Sexo
- Etapa de la vida
- Comentarios
- URL de la página del libro digitalizado en la que se proporciona el registro biológico

Hoja de ejercicios de aves de la literatura

Descargar la hoja de ejercicios. (MS Word, 342 KB)

Ejercicio 1

Planificación

El equipo debe desarrollar un flujo de trabajo sostenible para digitalizar los recursos bibliográficos, extraer información valiosa de biodiversidad y publicarla en línea a través de GBIF. Necesitan

desarrollar un plan que pueda sostenerse en el futuro una vez que finalice la financiación de la asociación nacional de bibliotecas.

La sección escenario de este caso de uso incluye una breve descripción del flujo de trabajo concebido por el director adjunto. Según el flujo de trabajo y el texto que lo acompaña, complete lo siguiente:

1. Identifique las diferentes partes interesadas que participan en este proyecto
2. Identifique su afiliación y asigne a cada uno de ellos a un grupo de partes interesadas
3. Identifique los roles asociados a ellos y asigne las tareas de las que son responsables actualmente
4. Realice un análisis crítico del flujo de trabajo, identifique riesgos y brechas potenciales, y sugiera formas de mejorar el flujo de trabajo, maximizar la eficiencia del proyecto de digitalización y producir datos de la más alta calidad posible.
5. Utilice la hoja de ejercicios para proporcionar sus respuestas.

Ejercicio 2

Captura de datos

Se han completado los escaneos y el reconocimiento de caracteres (OCR) de los libros. Los datos de registros biológicos ahora deben extraerse de esas fuentes y compilarse en un formato de hoja de cálculo.

Los datos originales estaban en alemán y, para facilitar su uso cuando se publiquen en línea, el director del proyecto desea que estén disponibles en inglés.

> 2. Cerchneis tinnunculus, Linn. ("Taarnfalk") nests in several steeples around Viborg.
>
> At the churches of Mönsted and Daubjerg there were several family groups in 1880, each of 7 individuals, suggesting a particularly abundant source of food on the surrounding heather (lizards). In the area of Thisted, where the tower falcon broods very often, young were always present in broods of 6, only once did I find a group of 7 on Egebäksande (H.)
>
> In Horsens I found kestrels brooding on the churches of Vähr, Hansted, Hundslund, Thyrsted and Oelsted. In the steeple of Vähr I found a nest with 6 eggs on 6 May 1875, 6 eggs on 30 April 1876, 4 eggs on 28 April 1878; in 1877 and 1879 there were none, on 2 May 1880 they had 5 eggs, in 1882 also 5 eggs, and in 1883 seven.
>
> In Oelsted 5 eggs on 3 May 1879, in 1882 also 5, and in 1883 6 eggs.
>
> On the churches (Our Lady's and St. Peter's) of Copenhagen, I have often seen it in the autumn and winter of 1879; in spring 1880, it brooded on Our Lady's church, the "round tower" and the law firm building (in a valve) and certainly also on St. Peter's Church. (F.)

1. Asuma el rol de un voluntario encargado de transformar el texto traducido en una hoja de cálculo como registros biológicos individuales. Los registros biológicos necesitarán que se les asignen números únicos.
2. Cree una hoja de cálculo utilizando los campos de datos enumerados en la Descripción de datos digitales utilizando los datos encontrados en el ejemplo anterior registrados por: Chr. P. Lütken.
3. Use the exercise sheet to provide your answers and submit the spreadsheet created in the previous step.

En los ejemplos utilizados, los registros biológicos individuales no siempre contienen datos para completar todas las columnas de la hoja de cálculo.

Ejercicio 3

Gestión de datos

Los voluntarios de la Sociedad Ornitológica de Copenhague han recopilado los datos en un formato de hoja de cálculo. Al asumir el rol de curador de ornitología en el departamento de aves, se le ha asignado la responsabilidad de los problemas relativos a la calidad de los datos.

A través de la georreferenciación retrospectiva, se han agregado coordenadas al conjunto de datos junto con la localidad, pero no otra geografía superior. Dado que todas las observaciones se realizaron en Dinamarca, el continente y el país se pueden agregar fácilmente. Además, solo se proporcionó el nombre científico. Se puede obtener una taxonomía más alta utilizando herramientas de software como OpenRefine. Usted además es consciente de que hay errores tipográficos cometidos por los digitalizadores.

1. Descargue UC3-DL-3-ForCleaning.zip. (45 KB)
2. Identifique y corrija cualquier año inválido.
3. Verifique y corrija la taxonomía.
4. Verifique que las coordenadas son correctas para las dos localidades dadas. Corrija las que no lo sean. Las coordenadas deben estar en formato decimal.
5. Agregue cualquier dato de los elementos faltantes que puedan derivarse utilizando los datos disponibles
6. Recuerde conservar la información original proporcionada y documentar sus cambios y suposiciones como parte de los registros individuales y los metadatos.
7. Use the exercise sheet to provide your answers and submit the cleaned text file extracted from the step 1.

el conjunto de datos debe contener sólo los años 1883-1939

Ejercicio 4

Publicación de datos

Para este ejercicio, asumirá el rol de la persona responsable de publicar los datos limpios en línea a través de la red GBIF. Se le ha proporcionado un archivo multimedia y un archivo histórico de identificación que debe publicarse junto con las observaciones. El miembro del personal a cargo de la calidad de los datos le ha proporcionado conjuntos de datos limpios para que los publique.

1. Descargue UC3-DL-4-ForPublication.zip. (65 KB)
2. Utilice la instalación IPT previamente proporcionada para publicar el conjunto de datos dado.
3. Utilice la hoja de ejercicios para proporcionar sus respuestas y el enlace al conjunto de datos publicado.

Proyectos finales

Para sus actividades finales, completará y enviará el Caso de uso II y el Caso de uso III para su evaluación.

CASO DE USO II

Hay dos opciones para el CASO DE USO II (Especies invasoras o Muestreo de Lepidoptera). Solo necesita seleccionar uno de ellos como su proyecto.

Archivos requeridos para el envío:

- hoja de ejercicios completa (se acepta documento de MS Word o similar)
- hoja de cálculo de captura de datos (Se acepta MS Excel, csv, txt o similar)
- conjunto de datos limpios/estandarizados (Se acepta MS Excel, csv, txt o similar)

CASO DE USO III

Solo hay una opción para el CASO DE USO III (Aves de la literatura).

Archivos requeridos para el envío:

- hoja de ejercicios completa (se acepta documento de MS Word o similar)
- hoja de cálculo de captura de datos (Se acepta MS Excel, csv, txt o similar)
- conjunto de datos limpios/estandarizados (Se acepta MS Excel, csv, txt o similar)

Incluya el caso de uso y el número del ejercicio junto con su nombre en todos los archivos enviados. Por ejemplo, Russell-UC2-IS-exercise-sheet.docx, Russell-UC2-IS-2.xlsx, Russell-UC2-IS-3.xlsx. **Todos los archivos deben enviarse en inglés**. Póngase en contacto con training@gbif.org si tiene alguna pregunta.

Envío del proyecto

Los proyectos pueden enviarse desde la versión en línea (HTML) del curso.

Evaluación del curso

Complete la evaluación del curso

Documentación clave

Las siguientes referencias proporcionan más detalles sobre los temas cubiertos en este curso. Todos los enlaces se abren en una ventana/pestaña nueva.

Darwin Core

- Darwin Core Terms: A quick reference guide
- Simple DarwinCore
- Darwin Core Questions & Answers
- Darwin Core extensions registered with GBIF

Publicación de datos

- Guía rápida para publicar datos a través de GBIF
- How to publish biodiversity data through GBIF.org
- Conviértete en un publicador de datos con GBIF
- Buenas prácticas para la publicación de datos sobre biodiversidad derivados de evaluaciones de impacto ambiental.
  Secretaría de GBIF & IAIA: Asociación Internacional para la Evaluación de Impactos (2020).
- Guidance for private companies to become data publishers through GBIF: Template document to support the internal authorization process to become a GBIF publisher
  Rui Figueira, Pedro Beja, Cristina Villaverde, Miguel Vega, Katia Cezón, Tainan Messina, Anne-Sophie Archambeau, Rukaya Johaadien, Dag Endresen & Dairo Escobar (2020).
- Publishing DNA-derived data through biodiversity data platforms
  Anders F. Andersson, Andrew Bissett, Anders G. Finstad, Frode Fossøy, Marie Grosjean, Michael Hope, Thomas S. Jeppesen, Urmas Kõljalg, Daniel Lundin, R. Henrik Nilsson, Maria Prager, Cecilie Svenningsen & Dmitry Schigel (2020).
- Tipos de conjuntos de datos soportados por GBIF
- GBIF data quality requirements for publishing
- GBIF data licenses
- Checklist core templates
- Ocurrence core templates
- Sampling event core templates
- Sampling event data best practices
- Compartiendo imágenes, sonidos y videos en GBIF
- Artículos de datos
- Artículos de datos publicados

Publicación de datos: IPT

- The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet
  Robertson et al. (2014)
- To install IPT or not to install IPT
- IPT data hosting centres
- IPT Install / Set up webinar
- Installing the IPT video

- IPT in Practice demonstration video

Digitalización

- iDigBio Digitization Resources
- iDigBio Collections Digitization Workflows
- iDigBio Digitization Workflows and Protocols
- iDigBio specimen image capture guide
- Canadensys 10-step guide to managing images with your biodiversity data

GBIF

- Qué es GBIF
- Plan estratégico
- Hágase miembro
- Science Review
- Establishing an Effective GBIF Participant Node: Concepts and general considerations
  GBIF Secretariat (2019).

Georreferenciación

- Georeferencing Best Practices
  Arthur D. Chapman & John R. Wieczorek (2020).
- Georeferencing Quick Reference Guide
  Paula F. Zermoglio, Arthur D. Chapman, John R. Wieczorek, Maria Celeste Luna & David A. Bloom (2020).
- Georeferencing Calculator Manual
  David A. Bloom, John R. Wieczorek & Paula F. Zermoglio (2020).
- Georeferencing resources

Especies invasoras

- GRIIS - Global Register of Introduced and Invasive Species
- TrIAS - Tracking Invasive Alien Species

Living Atlases

- Living Atlases
- ALA key technical documentation

Misceláneos

- VertNet Guide to opening text files in Excel
- VertNet data licensing guide

OpenRefine

- OpenRefine documentation
- OpenRefine regular expressions
- Guía para la limpieza de datos sobre biodiversidad con OpenRefine
  Paula F. Zermoglio, Camila A. Plata Corredor, John R. Wieczorek, Ricardo Ortiz Gallego & Leonardo Buitrago (2021).
- Using Google Refine and taxonomic databases (EOL, NCBI, uBio, WORMS) to clean messy data
  iPhylo blog post. Rod Page 2012.
- Reconciling author names using Open Refine and VIAF
  iPhylo blog post. Rod Page 2013.
- Validating scientific names with the GBIF Portal web service API
  Guest post was written by Gaurav Vaidya, Victoria Tersigni and Robert Guralnick 2013.
- iDigBio Cleaning data with OpenRefine
  iDigBio 2013.
- Have We Got the Names “Right”?
  Canadensys 2014.
- Cleaning data with OpenRefine
  Desmet and Brosens 2016 TDWG.
- EasyOpen Redlist
  Querying the IUCN Red List, using a species list, OpenRefine, and some pre-written code. Olly Griffin July 2019.

Planificación/Colaboración

- Agile methodology
- What is SCRUM
- SCRUM Framework
- Metodología Kanban
- Scrum Guide
- GitHub

Calidad

- Principles of Data Quality
  Arthur Chapman 2005.
- Principles and Methods of Data Cleaning: Primary Species and Species-Occurrence Data
  Arthur Chapman 2005.
- Be careful with dates in Excel
  DataOne 2014.
- Character encoding for beginners
- MVZ Guide for Recording Localities in Field Notes

Especies sensibles

- Current Best Practices for Generalizing Sensitive Species Occurrence Data
  Arthur D. Chapman 2020.

Taxonomía

- GBIF checklist datasets and data gaps
- GBIF Labs - Analizador de nombres
- GBIF Labs - Coincidencia de especies
- Global Names Resolver
- Marine Name Matching Strategy for taxonomic quality control
- Nomenmatch

Glosario

ALA

Atlas of Living Australia. El nodo australiano de GBIF, quien desarrolló un portal de datos de código abierto ampliamente utilizado dentro de la comunidad de GBIF y sus socios para los portales nacionales de biodiversidad.

API

Interfaz de programación de aplicaciones. Un conjunto de métodos de comunicación claramente definidos entre varios componentes de software.

BID

Información sobre Biodiversidad para el Desarrollo. Un proyecto financiado por la UE y coordinado por GBIF cuyo objetivo es aumentar la capacidad de movilización de datos en las regiones de África, el Caribe y el Pacífico.

BIFA

Fondo para la Diversidad Biológica de Asia.

Licencias CC

Creative Commons. Se trata de una serie de licencias establecidas por la organización Creative Commons que permiten compartir y reutilizar la creatividad y el conocimiento mediante la provisión de herramientas legales gratuitas. Tres de ellas se pueden asignar a conjuntos de datos compartidos por GBIF: CC0, CC BY y CC BY-NC.

Vocabulario controlado

Se trata de un conjunto restringido de términos que se utilizan como valores posibles para un campo determinado. Uno puede pensar en ello como una lista de búsqueda o un menú desplegable para un campo en particular. Por ejemplo, el campo DwC basisOfRecord solo debe contener uno de estos valores: "PreservedSpecimen", "FossilSpecimen", "LivingSpecimen", "HumanObservation", "MachineObservation". Diríamos que la lista de valores es un vocabulario controlado para ese campo.

DwC

Darwin Core es un estándar de datos de biodiversidad, mantenido por TDWG y ampliamente

utilizado dentro de la comunidad y socios de GBIF. Es un conjunto de términos estandarizados (o nombres de campo) y sus definiciones, que se utilizan para compartir información sobre biodiversidad.

DOI

Digital Object Identifier. Identificador persistente o handle que se utiliza para identificar objetos de forma única. Los DOI se utilizan principalmente para identificar información académica, profesional y gubernamental, como artículos de revistas, informes de investigación y conjuntos de datos, y publicaciones oficiales.

DwC-A

Archivo Darwin Core. Un archivo comprimido (zip) que contiene toda la información necesaria para compartir con GBIF, para un recurso en particular. Cada zip contiene tres tipos de archivos:

1. los datos reales, en uno o más archivos de texto: ooccurrence.txt/event.txt/measurmentoffact.txt, etc
2. un archivo de mapeo: rtf.xml
3. un archivo de metadatos (EML): eml.xml Cuando publica utilizando el IPT, crea un Archivo de Darwin Core, que se comparte con GBIF. Además, cuando descarga datos del sitio web de GBIF, también puede elegir un formato DwC-A.

GUID

Identificador Global Único

IPT

Integrated Publishing Toolkit. Es una aplicación web (software) gratuita y de código abierto para publicar datos de biodiversidad. El software en sí se aloja en un servidor (ya sea en su institución o en otro lugar) que debe tener acceso a Internet las 24 horas del día, los 7 días de la semana. Se utiliza para crear y gestionar archivos de Darwin Core que pueden ser compartidos y utilizados por cualquier persona, incluido GBIF.

Préstamo

En el contexto de las colecciones de historia natural, este es el procedimiento de préstamo de especímenes entre instituciones.

LSID

Life Sciences Identifier. Son identificadores persistentes, globales y únicos para objetos biológicos.

Publicación de datos

Con respecto a GBIF, tenemos una definición muy específica de publicación de datos. Se refiere a lograr que los conjuntos de datos de biodiversidad sean accesibles y reconocibles públicamente, en una forma estandarizada, a través de un punto de acceso, generalmente una dirección web (una URL).

Recurso

Un Recurso es el término colectivo utilizado para referirse a un conjunto de datos en particular y sus metadatos una vez que se ha cargado a una instancia IPT.

TDWG

Taxonomic Databases Working Group, ahora renombrado Biodiversity Information Standards.

URN

Número de Recurso Uniforme

UUID

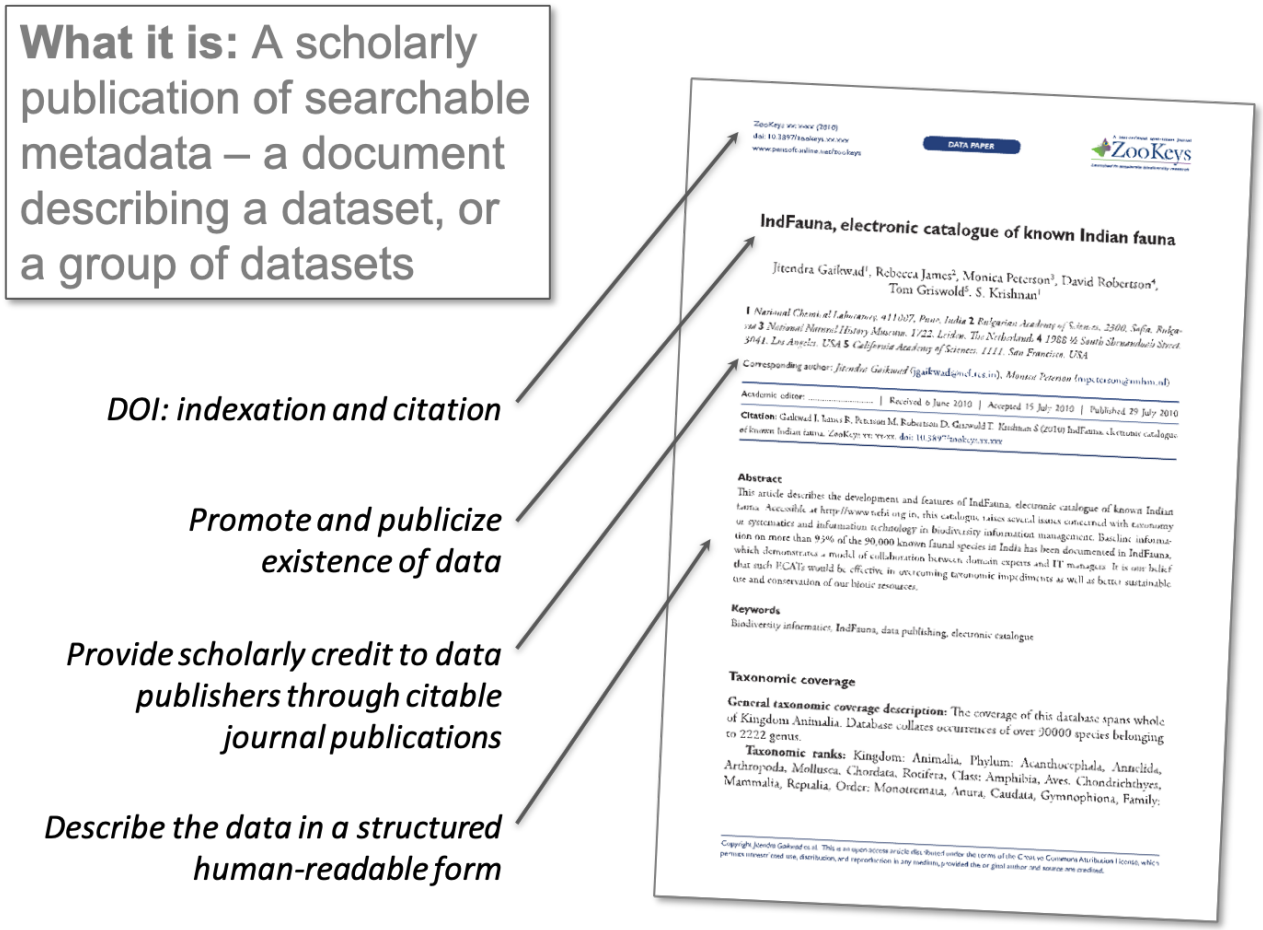
Identificador Universalmente Único

Apéndice: Artículos de datos

Un artículo de datos es un documento revisado por pares que describe un conjunto de datos, publicado en una revista revisada por pares. Preparar, conservar y describir los datos supone un esfuerzo. Los artículos de datos reconocen este esfuerzo por medio de un artículo académico. En este curso no cubrimos cómo crear artículos de datos, sin embargo, como actividad opcional, puede ver este vídeo (51:51) presentado por Lizanne Roxburgh. En este video, aprenderá más sobre la publicación de artículos de datos. Si no puede ver el video incrustado, puede descargarlo localmente. (MP4 - 99,2 MB)

https://vimeo.com/265350948 *(Vimeo video)*

Puede leer más sobre artículos de datos en GBIF.org.





Vea los enlaces a continuación para ver un artículo de datos tal y como aparece en el IPT, en GBIF.org y en la revista científica. Cada uno de ellos tiene un vínculo directo.

- Revista científica: https://doi.org/10.3897/BDJ.2.e4244
- GBIF: https://www.gbif.org/dataset/3c9e2297-bf20-4827-928e-7c7eefd9432c
- IPT: http://130.120.204.55:8080/ipt/resource.do?r=harvestmen_of_french_guiana

Apéndice: Soluciones

Este apéndice contiene las respuestas e información adicional a todos los cuestionarios de revisión. Además, esta sección contiene una solución sugerida para el CASO DE USO I.

Soluciones de la revisión de las bases

Para el enunciado dado, ingrese el término correcto (base de datos, lenguaje de base de datos, programa de base de datos)

- combines and presents functions and features for manipulating data, together in a unified interface
  database program
- structured and organized collection of data and/or information held on a computer

database

- the way by which a human communicates with a computer
  database language

Si abre un archivo de datos y ve lo siguiente, ¿cuál sospecharía que es el problema?

```
tre, ou ne pas tre, c est l la question.
```

- La codificación incorrecta fue usada para abrir el archivo

Para el software dado, ingrese el tipo de software (captura de datos, gestión de datos, limpieza de datos, publicación de datos).

- Integrated Publishing Toolkit (IPT)
  data publishing
- Specify
  data capture AND **data management**
- iNaturalist
  data capture
- OpenRefine
  data cleaning

Para el ejemplo dado, ingrese el tipo de datos correcto (binario, booleano, flotante, entero, entero largo, texto, texto no estructurado)

- 1236975
  long integer
- 01101111
  binary
- We walked 5 miles down the road west from the post office in the center of town. We then went 2 miles north on a dirt path to the river. Then we continued west along the river for another 5 miles.
  unstructured text
- 1024
  integer
- 29.0
  float
- Yes/No
  boolean
- 6 rabbits were observed
  text

¿Cuál de estos términos describe un "nombre de campo/columna"?

- Asignado
- Identificador
- Único

¿Cuál de estos términos describe una "etiqueta de campo"?

- Descriptivo

- Legible
- Interfaz de usuario

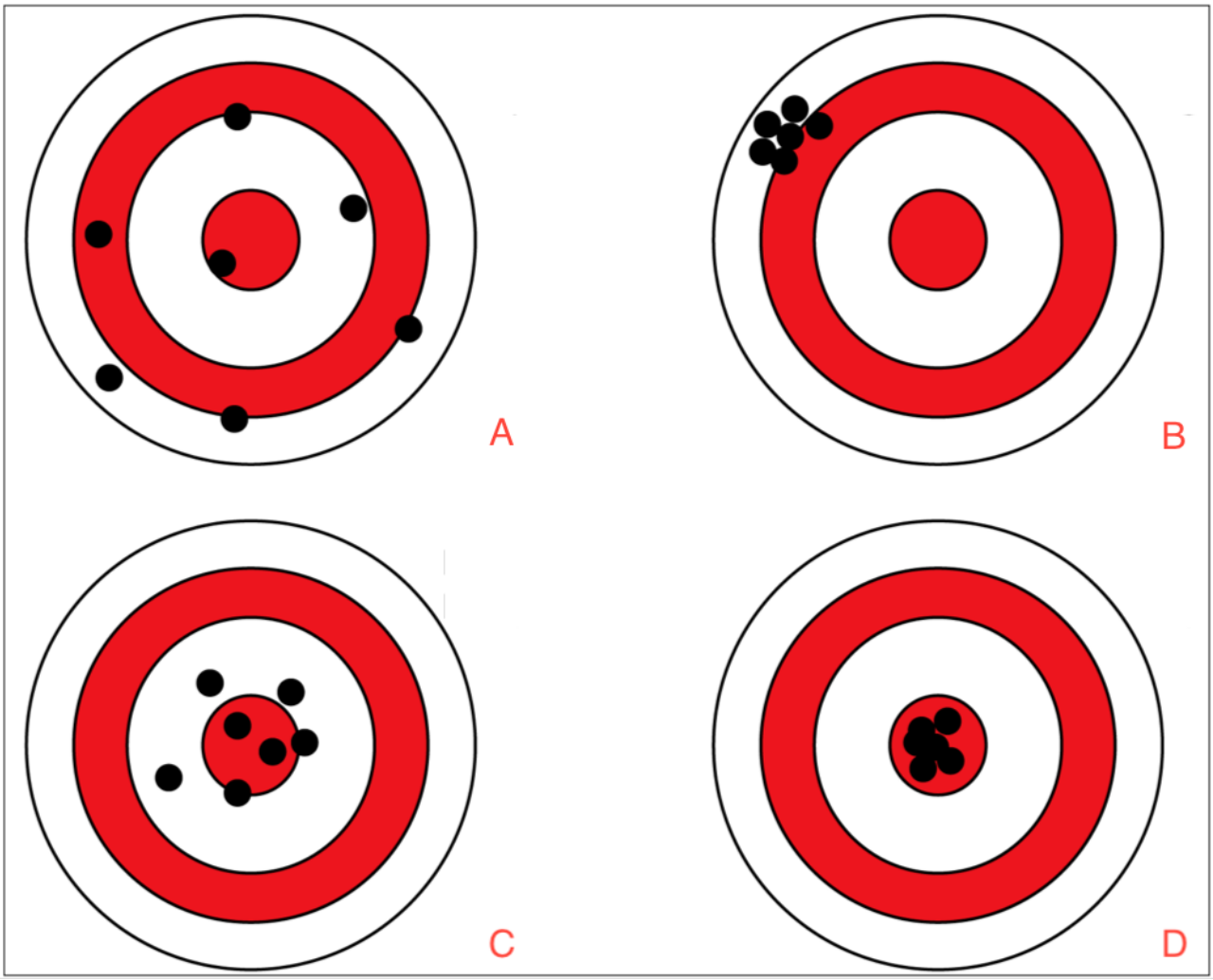
Para cada enunciado, introduzca la estructura correcta (fila, columna, tabla)

- All data refers to a SINGLE concept.
  table
- An attribute has the SAME field/data type for every record.
  column
- Attributes of a record ALWAYS stay together.
  row

¿Quién determina la aptitud para el uso ("fitness for use") de sus datos?

- Los usuarios de los datos para investigación o educación

For the given statements, input the matching image. (A, B, C, D)



- High accuracy, low precision
  C
- Low accuracy, high precision
  B
- High accuracy, high precision
  D

- Low accuracy, low precision
  A

Identifique las relaciones de datos donde el conjunto de datos B debe combinarse con el conjunto de datos A (0:1, 1:0, 1:1, 1:∞, ∞:1, ∞:∞). No se utilizan todas las relaciones.

- Collector field exists in both dataset A and B
  1:1
- Country field only exists in dataset B
  0:1
- Name field exists in dataset A, but dataset B contains First Name and Last Name fields
  1:∞
- ID field exists in both dataset A and B
  1:1
- Elevation exists in dataset A, but not in dataset B
  1:0
- Date exists in dataset A, but Day, Month, and Year are separate fields in dataset B
  1:∞

Los metadatos son importantes porque (seleccione los enunciados VERDADEROS)

- permiten a los usuarios determinar si un conjunto de datos es adecuado para su uso.
- le permiten saber bajo qué términos legales se permite la reutilización de datos.

Soluciones de la revisión de planificación

¿Cuál es el orden de los cinco Grupos de Procesos PMBoK?

- Iniciar, Planear, Ejecutar, Monitoreo y Control, Cerrar

Aprenda más: https://quizlet.com/306742513/1-introduction-pmbok-guide-6th-edition-flash-cards/

¿Cuáles son los tipos de entregables?

- Declarado - SI
- Implícito - SI
- Estimado - NO
- Directo - SI
- Indirecto - SI
- Suposiciones - NO

¿Qué es un cuello de botella?

- un bloqueo que retrasa el desarrollo o el progreso - SÍ un espacio donde falta algo o alguien - NO, ESTO ES UNA BRECHA
- un problema, o situación que impide a alguien hacer algo, o que hace algo imposible. NO, ESTO ES UN BARRERA

¿Cuáles son los ejemplos de las tareas de movilización?

- Afiliación - NO, Este es un Tipo de Recurso
- Publicación - SI
- Imagen - SI
- Georreferenciación - SI
- Mayor conciencia pública - NO, Este es un objetivo implícito.

Soluciones de la revisión de captura de datos

¿Qué tipo(s) de conjuntos de datos elegiría para una colección de ictiología?

- occurrence
  Most of the time, specimens from collection databases are shared as occurrence data. Each occurrence (specimen or group of specimens) has its own unique identifier (sometimes derived from its catalogue number in the source collection) and the Darwin Core fields used to share them within GBIF describe each specimen: scientific name, the date it was collected on the field, who collected and/or identified it, where, etc. Each collection can have more than one specimen from a same species, as long as each specimen is identified by a unique ID.
- checklist
  It is also possible to create and share a taxonomical checklist derived from a collection database; in this case, it is recommended to share the checklist as a taxonomical dataset, with the occurrence (specimen) list associated with it by using the Occurrence core as an extension to the Taxon Core on the GBIF IPT.

¿Qué tipo(s) de conjunto de datos elegiría(s) para una lista de especies invasoras?

- occurrence
  Some data publishers will share occurrence datasets coming from studies or programs tracking specimens from some specific invasive species; when the data focuses on individuals instead of the invasive species, in general, they can be shared as occurrence data.
- checklist
  Invasive species can be tracked and monitored at different scales (regional, national, thematic…); as this type of dataset focuses more on the species and their distribution across a given geographical scope, they are mainly shared as taxonomical datasets within GBIF (see GRIIS search results).

¿Qué tipo(s) de conjuntos de datos elegiría para la flora y fauna de un estudio de impacto ambiental?

- occurrence
  Data are recorded by naturalists on the field and can be shared as simple occurrence datasets.
- sampling event
  They can also be shared as event datasets if standardized protocols (such as vegetation plots, transects, traps…) are used to collect the data.

¿Qué tipo(s) de conjunto de datos elegiría(s) para los datos de seguimiento de aves?

- occurrence
  These data are shared as occurrence datasets: ideally, each bird is identified with its organismID, and each occurrence (GPS ping) has its own occurrenceID, which is useful to track the different GPS locations of the same bird over the scope of the tracking programme or project. (See example)

¿Qué tipo(s) de conjuntos de datos elegiría para los datos de trampas de insectos?

- occurrence
  Although such data can be shared as simple occurrence datasets, it is best if they're shared as event datasets, where the location, identifier and contents of each trap can be better detailed.
- sampling event
  Insect traps (as well as other traps such as pitfall traps, malaise traps…) are typically used in monitoring programmes to check the presence (or absence) of some species and/or assess their specific abundance. Using the "eventID" field to identify each trap allows the users to get all of the specimens collected within each trap. The same logic applies to other field protocols such as transects, plots, remote cameras, etc.: by using the Event Core instead of the Occurrence core, you'll be able to share much more information about the context of the data collection, and allow users to better understand (and even replicate) your work.

¿Qué tipo(s) de conjuntos de datos elegiría para los datos de gestión de parques nacionales?

- occurrence
  record individuals of species
- checklist
  It is important to know how many species are present in the park/reserve perimeter and their conservation status.
- sampling event
  check and track the populations

¿Qué tipo(s) de conjunto de datos elegiría(s) para un bioblitz de ciencias ciudadana?

- occurrence
  Bioblitz datasets are mainly shared as occurrence datasets.
- sampling event
  Depending on the citizen science programme, specific sampling protocols might be used by the volunteers, in which case, the data can be shared as an event dataset.

¿Qué tipo(s) de conjunto de datos elegiría(s) para una listado regional de especies?

- checklist
  Geographical or thematic species lists are often used to share information about the species present in a given area; most of the time, these lists also mention the distribution of each species as well as their conservation status in this area. Regional species lists can give a useful insight into a region's biodiversity and habitats, and need to be shared as taxonomical datasets, with or without associated occurrences.

Soluciones de la revisión de gestión de datos

¿Por qué es mejor limpiar sus datos?

- para hacerlos tan aptos para el uso ("fit for use") como sea posible
- para lograr sus objetivos de calidad de datos

Siempre debe aspirar a administrar y publicar datos con la mayor calidad posible. Esto mejorará su trabajo diario (es más fácil trabajar con datos organizados y limpios), así como el trabajo de los potenciales reutilizadores de sus datos, que necesitan comprenderlos y confiar en su fuente antes de usarlos.

¿Cómo debe organizar su flujo de trabajo de limpieza de datos?

- pregunta a sus colegas experimentados
- trabajar a nivel institucional para armonizar los flujos de trabajo de calidad de datos

No se espera que alguien sepa todo sobre datos de biodiversidad; debe buscar ayuda y consejo de sus colegas u otras personas con conocimientos y asegurarse de que está aplicando las buenas prácticas recomendadas por su institución a medida que limpia sus datos.

Qué es mejor:

- evitar que se generen errores
- corregir los errores tan pronto como los encuentre en su base de datos o hoja de cálculo

La mejor manera de evitar la propagación de errores en sus datos es evitar que ocurran al inicio del proceso de colecta/registro de datos.

Por supuesto, los errores son inevitables, así que debería limpiarlos tan pronto como los encuentre, y documentar el proceso de limpieza.

Si no tiene tiempo o recursos para limpiar adecuadamente sus datos, es mejor esperar a poder hacerlo en lugar de publicar datos erróneos que puedan confundir a la gente.

¿Quién es responsable por la calidad de los datos?

- Todos los involucrados en la gestión de datos

Cada persona involucrada en la gestión del flujo de trabajo de datos es, al menos parcialmente, responsable por su calidad, desde los técnicos de campo hasta los administradores de bases de datos.

Las personas que puedan usar sus datos más adelante, pueden informarle de cualquier error restante en sus datos y deben usarlos de manera responsable para su propia investigación, pero la calidad de los datos iniciales no es su responsabilidad.

GBIF puede realizar comprobaciones automáticas de sus datos (p. ej., detección de valores perdidos, valores geográficos atípicos, nombres científicos desconocidos), pero no debe responsabilizarse de los errores que hayan ocurrido anteriormente en el proceso de gestión de datos.

¿Qué herramientas se pueden utilizar para limpiar sus datos?

- Excel y otras herramientas de gestión de hojas de cálculo
- OpenRefine
- Su software de base de datos
- Herramientas online como el Scientific Names Resolver o Google Maps

Se pueden usar todo tipo de herramientas para limpiar sus datos, pero debe identificar cuáles responderán a sus necesidades en términos de resolución taxonómica, georreferenciación, eliminación de duplicados, etc. Puede encontrar herramientas útiles enumeradas en la sección de gestión de datos.

Soluciones de la revisión de publicación de datos

¿Qué significa la publicación de datos en el contexto de GBIF?

- Hacer públicamente accesibles su(s) conjunto(s) de dato(s) de biodiversidad y se puede hallar en un formato estandarizado

La publicación de datos en GBIF significa hacer públicamente accesibles su(s) conjunto(s) de dato(s) de biodiversidad en un formato estandarizado (la mayoría de las veces, Darwin Core), para que pueda ser descubierto y reutilizado por otras personas.

¿Qué es un IPT?

- una herramienta que lo ayuda a publicar sus datos en GBIF
- una herramienta que le ayuda a producir un artículo de datos (Data Paper)

El IPT (Integrated Publishing Toolkit) es un software codificado en Java que le permite cargar y publicar datos en GBIF. No se debe utilizar como herramienta de gestión o limpieza de datos.

El IPT también puede ayudarle en el proceso de escribir y enviar un documento de datos, gracias al archivo EML que se genera automáticamente cuando se completan los metadatos del recurso de datos.

¿Qué licencias y exenciones de Creative Commons son recomendadas por GBIF para publicación de datos?

- C0, CC-BY y CC-BY-NC

Las licencias de Creative Commons y exenciones recomendadas para publicar su(s) conjunto(s) de datos en GBIF son CC0, CC-BY y CC-BY-NC. Son licencias y/o exenciones ampliamente reconocidas que se ajustan a los requisitos internacionales de datos abiertos para compartir y reutilizar datos.

Tenga en cuenta que sólo debe elegir CC0 o CC-BY licencia/exención para su(s) conjunto(s) de dato(s) relacionados con BID.

¿Cuáles son los tres Núcleos (Cores) entre los que se puede elegir para un recurso IPT

- Occurrence Core, Taxon Core, Event Core

Puedes elegir uno de los tres núcleos siguientes para cada uno de tus recursos IPT: registro biológico, taxón o eventos.

El estándar de Darwin Core también te permite enlazar extensiones a tu núcleo elegido, como SimpleMultimedia o MeasurementOrFact.

Los metadatos se completan en una sección separada del IPT y se comparten utilizando el estándar EML, no el Darwin Core (que se utiliza sólo para datos).

¿Cuántos archivos de extensiones puede tener un conjunto de datos?

- tantos como sea necesario

Una vez que haya elegido un núcleo para su recurso IPT, puede añadirle extensiones de Darwin Core. Puede añadir una o varias extensiones, dependiendo del tipo de núcleo que elija y de qué extensiones son compatibles con el.

Las extensiones no son obligatorias (puede publicar un conjunto de datos sin ninguna extensión) pero pueden ser útiles si quiere compartir información adicional que no podrías mapear con el Core elegido.

Solución sugerida del Caso de Uso I

solución sugerida (PDF 144 KB)

Agradecimientos

Diseño del curso e instrucción

El éxito de este curso depende en gran medida del apoyo brindado a los participantes de la red de mentores de mejora de capacidad de GBIF. Visite la página de GBIF en tutoría de mejora de la capacidad para leer más sobre estas personas y sus contribuciones.

Los siguientes individuos son reconocidos por sus contribuciones significativas al diseño, los materiales y la instrucción del curso:

- Néstor Beltrán*
- David Bloom
- Katia Cezoń*
- Dag Endresen
- Alberto González-Talaván*
- Sharon Grant*
- Marie-Elise Lecoq
- Sophie Pamerlon*
- Nicolas Noe*
- Meĺianie Raymond*
- Laura Anne Russell*
- John Wieczorek
- Paula Zermoglio
- Creadores del plan de estudios

Reconocimiento especial a Arthur Chapman por la reutilización de sus materiales sobre Calidad de Datos.

Traductores

Francés

- Maxime Coupremanne
- Jaures Gbètoho
- Marie Grosjean
- Patricia Mergen
- Sophie Pamerlon
- Andry Jean Marc Rakotomanjaka
- Y. Sabastian Wirsiy

Portugués

- Rui Figueira
- Clara Baringo Fonseca
- Keila Elizabeth Macfadem Juarez
- Tainan Messina

Español

- Leonardo Buitrago
- Victor Chocho
- Camila Plata
- Anabela Plos
- William Ulate
- Paula Zermoglio

Recursos

- La plataforma de mesa en línea es proporcionada por PlayingCards.io. Muchas gracias a Jwalant Patel y Eric Ma por encontrar y ayudar a crear las mesas de juego en línea y a Kate Webbink por su experiencia artística.
- Los iconos usados en este curso son hechos por Freepik desde www.flaticon.com
- OpenRefine
- Integrated Publishing Toolkit

Apoyo en recursos

- Belgium Biodiversity Platform
- GBIF France
- GBIF Noruega
- GBIF España
- SiB Colombia
- The Field Museum
- VertNet

Colofón

Cita sugerida

Curso de movilización de datos sobre biodiversidad de GBIF (2021) GBIF. 12ª edición. Secretaría de GBIF:. https://doi.org/10.35035/ce-c6cr-6w42. [Fecha del curso.]

Colaboradores

El *Curso de movilización de datos sobre biodiversidad de GBIF* se desarrolló originalmente como parte de Información de Biodiversidad para el Desarrollo, un programa financiado por Unión Europea. El plan de estudios original fue creado por Néstor Beltrán, Sharon Grant, Nicolas Nóe, Sophie Pamerlon, Alberto González-Talaván, Mélianie Raymond, Laura Anne Russell y Katia Cezón, con contribuciones adicionales de entrenadores, mentores y estudiantes de GBIF.

Licencia

Curso de Movilización de Datos sobre Biodiversidad de GBIF está licenciado bajo Creative Commons Attribution 4.0 Unported License.

URI persistente

https://doi.org/10.35035/ce-c6cr-6w42

Control del documento

12ª edición, mayo 2021