



GUÍA DE CURADURÍA DE REGISTROS BIOLÓGICOS EN **BIOMODELOS**

biomodelos.humboldt.org.co

Versión 1 · 2016

La primera etapa al desarrollar los modelos de distribución de especies involucra la agregación y estandarización de registros biológicos, los cuales deben ser verificados taxonómica y geográficamente. Estas verificaciones emplean herramientas automatizadas para examinar la consistencia geográfica de los datos y la existencia de los nombres científicos en bases de datos globales, pero requieren de expertos para verificar la calidad de los mismos, particularmente en su componente taxonómico. En este documento encontrarás una guía para usar las herramientas disponibles en BioModelos para la curaduría colaborativa de los registros así como de los recursos que puedes utilizar para curar individualmente tus propios conjuntos de datos.

¿Qué registros usa BioModelos?

La base de datos de BioModelos agrega los registros provenientes de bases de datos abiertas (e.g. SiB, GBIF, eBird, SpeciesLink, VertNet, Xenocanto) y de proyectos privados que aún no han publicado sus conjuntos de datos en alguna de estas plataformas. Todos estos registros descargados manualmente son estructurados de tal forma que contengan como mínimo los campos "ID_LBA", "ID del registro" (de acuerdo a la base de datos de la cual fue descargado), "nombre de la especie", "país", "departamento", "municipio", "localidad", "fecha", "latitud" y "longitud". Esta base de datos, que cuenta con cerca de 13'000.000 de registros biológicos, se somete a una serie de filtros de acuerdo a los siguientes criterios:

- 1. Validación taxonómica: este proceso utiliza los valores del campo "nombre científico" reportados en cada uno de los registros. Este proceso se hace a través de la base de datos MySQL de Catalogue of Life en su versión anual del 2014, que hace reconocimiento de nombres científicos hasta el nivel de especie. Esta base de datos genera un reconocimiento taxonómico utilizando el género y epíteto específico. Para poder hacer la validación se genera una partición de texto separando estos dos elementos y se hace una depuración para remover calificadores taxonómicos, autorías científicas y demás. Una vez establecido el género y epíteto específico la base de datos Catalogue of Life 2014 retorna campos de la taxonomía superior (reino, filo, clase y demás), la validez actual del nombre y el nombre científico vigente. Los resultados de este proceso permiten corroborar el nombre original, el nombre depurado, el estado taxonómico del nombre (aceptado, sinónimo, etc.), el nombre aceptado actualmente y la fuente taxonómica (base de datos de especialistas) que aportó ese nombre científico a Catalogue of Life 2014. En caso de no poder corroborar el nombre científico sometido, los campos quedan vacíos y marcados con una etiqueta para su revisión.
- 2. Validación geográfica: en esta parte el primer paso es descartar los registros que no poseen latitud y longitud en grados decimales. Posteriormente se evalúa la consistencia de los valores en los campo "país", "departamento" y "municipio", para lo cual se hace una superposición espacial entre los puntos de los registros y las capas de municipios, departamentos y cascos urbanos, que provienen de la cartografía oficial del país (descargables en el SIGOT). Con esta unión se verifica que la información que tiene originalmente el registro en los campos mencionados corresponda con su ubicación espacial.
- 3. Eliminación de duplicados: se consideran duplicados aquellos registros de una misma especie que se ubican en una celda de aproximadamente 1km2. De estos registros sólo se mantiene uno y a los demás se les asigna una etiqueta para no ser usados en el modelamiento. También en caso de encontrarse más de un registro con la misma institución y número de catálogo, se mantiene sólo el registro que tenga el mayor número de campos documentados.

¿Quiénes participan en la curaduría colaborativa de registros?

Esta actividad es asignada en primera instancia por el moderador del grupo donde se encuentra la especie de interés a uno o varios expertos registrados en la plataforma, durante la construcción de las agendas de modelamiento de los grupos. Sin embargo, cualquier usuario que considere que puede contribuir a esta actividad, puede registrarse en la plataforma, unirse a su grupo de interés y comunicarse con el moderador para aportar en la curaduría de registros.

¿Cómo se hace la curaduría colaborativa de datos a través de BioModelos?

BioModelos cuenta con un visor geográfico que proporciona algunas herramientas para facilitar la curaduría colaborativa de los registros biológicos de especies. Para acceder a ellas, debes estar registrado en la plataforma. Una vez inicies tu sesión, en la sección modelos podrás buscar tu especie de interés. Al dar click en "ir" aparecerá por defecto la sección "Modelos y comentarios", donde podrás seleccionar alguna de las hipótesis disponibles de distribución para la especie. Si no deseas ver los modelos e ir directamente a los registros, puedes cerrar esta sección (Figura 1). Los registros que se visualizan allí corresponden a los de la base de datos de BioModelos (Figura 2).



Figura 1. Cerrar la sección "Modelos y comentarios" que aparece por defecto para ver directamente los registros de la especie.



Figura 2. Visualización de los registros biológicos de la especie Artibeus lituratus.

Al dar click en alguno de los registros se desplegará una caja con información base que considera los atributos de latitud, longitud, ID (dependiendo de la fuente de descarga del registro), especie original, localidad, municipio, departamento, altitud, fecha, institución, colector y evidencia (Figura 3).



Figura 3. Caja desplegada con la información asociada para un registro biológico de la especie *Artibeus lituratus*. Los puntos verdes y amarillos corresponden a conjuntos de registros muy cercanos en su ubicación. Al dar click sobre ellos se hará zoom hasta visualizar los registros individuales (marcadores azules).

Reporte de error

Si al ver todos los registros de la especie, consideras que alguno está por fuera de la distribución de la especie o que la información asociada está incorrecta, puedes dar click en el botón "reportar error" que se encuentra al final de la caja desplegada para el registro (Figura 3). Allí encontrarás las siguientes categorías de error que podrás seleccionar (Figura 4):

- · Georreferenciación dudosa: la localidad del registro está mal georeferenciada.
- **Outlier geográfico:** aplica para observaciones en las que la identificación es correcta pero que corresponden a eventos extremos que pueden sesgar un modelo.
- Individuo introducido: el registro corresponde a un individuo en cautiverio o sembrado.
- Identificación dudosa: se sospecha que la identificación es incorrecta. Esta sospecha surge por ejemplo cuando un registro está considerablemente por fuera del rango de distribución de la especie (y no es un error de georeferenciación) o cuando la evidencia del registro (e.g. observación) no es suficiente para identificar la especie.
- **Nombre desactualizado:** nombre científico desactualizado. Se puede ingresar el nombre actual en el campo de comentarios.



• Otro: otros problemas con el registro que lo invalidan para su uso en modelamiento.

Figura 4. Reporte de error para un registro biológico de Artibeus lituratus.

Una vez selecciones una o varias categoría de error, podrás realizar comentarios adicionales en la caja de comentarios para ayudar a mejorar la calidad de los registros de acuerdo a los errores reportados. Por último debes guardar los reportes realizados. Estos serán almacenados en la base de datos de BioModelos y los registros asociados no serán tomados para el modelamiento de la distribución a menos que se editen manualmente los valores relacionados con el tipo de error reportado.

Agregar un registro

Si deseas agregar un registro para la especie de tu interés, puedes hacerlo de la siguiente manera: selecciona alguno de los modelos que aparecen en la sección "Modelos y comentarios" para habilitar la herramienta "Editar" (Figura 5); al desplegarla aparece la opción "Dibujar marcador", la cual te permite desplazarte por el visor geográfico y ubicar un nuevo registro de la especie, teniendo como guía la información de latitud y longitud (Figura 6).

Para subir un conjunto de datos (e.g. más de 10 registros biológicos) nuevo para una especie, puedes escribirnos a biomodelos@humboldt.org.co. Todos los permisos de visualización de los registros biológicos son determinados por el autor o autores del conjunto de datos. En caso de que desees dar acceso abierto al conjunto de datos, recomendamos publicarlos a través del SiB Colombia.



Figura 5. Activación de la herramienta "Editar" (esquina superior izquierda) para el modelo híbrido de la especie Artibeus lituratus.



Figura 6. Herramienta "Dibujar marcador" para agregar registros biológicos de la especie Artibeus lituratus.

¿Cómo puedo curar mis propios conjuntos de datos?

Los expertos pueden trabajar directamente con sus propios conjuntos de datos (construidos a partir de investigaciones previas en campo) o pueden solicitar los registros biológicos de la especie o especies de interés que se encuentren en la base de datos de BioModelos.

Es importante tener en cuenta que para poder llevar un seguimiento histórico de los cambios realizados en los registros por parte de los expertos, es necesario que las modificaciones se realicen teniendo en cuenta el campo "ID_LBA", que corresponde a un identificador interno que maneja el laboratorio de biogeografía aplicada para el uso de los registros en el modelamiento de las distribuciones de especies. En este sentido, si solicitaste la base de datos de tus especies de interés a BioModelos, mantén este campo sin modificaciones. Si deseas añadir nuevos registros (bien sean tuyos o de otros investigadores que dieron su aprobación para usarlos en BioModelos), deja este campo en blanco.

Una vez cuentes con los registros de la especie, puedes empezar la actividad de curaduría de los registros biológicos con tus propias herramientas. Algunas herramientas para la limpieza de datos pueden ser consultadas en lo siguientes documentos:

Categorías taxonómicas con OpenRefine: Una guía de referencia Georreferenciación de localidades: Una guía de referencia para colecciones biológicas Verificación geográfica de registros biológicos

Adicionalmente, se recomienda contar con conocimientos básicos en Sistemas de Información Geográfica (SIG) para visualizar la ubicación de los registros biológicos.

¿Cuál será el manejo de las bases de datos curadas por los expertos?

Los conjuntos de datos que han sido sometidos a un proceso de curaduría por parte de los expertos serán almacenados en la base de datos de BioModelos. Allí, los registros para los cuales se haya reportado algún tipo de error no serán eliminados de la base de datos, sino que se les asignará una etiqueta para no ser usado en el modelamiento de la distribución de la especie y hacer una revisión posterior para corregir dicho error. Los reportes de error que se hagan a los registros del SiB Colombia serán reportados a los publicadores para contribuir al mejoramiento en la calidad de los datos abiertos sobre la diversidad biológica de Colombia.

Definiciones

Moderador: usuario que administra las actividades de un grupo de expertos en BioModelos.

Agenda de modelamiento: listado de especies y fechas en las que un grupo propone ejecutar las distintas etapas del desarrollo de un modelo de distribución en BioModelos.

Experto: un experto en BioModelos es aquel que puede contribuir a la curación de registros o mejoramiento de un modelo de distribución de una o más especies, en virtud de su conocimiento sobre la distribución y ecología de las mismas.

¿Preguntas?

Escríbenos a biomodelos@humboldt.org.co.