NOTA DE PRENSA

@mncn_csic

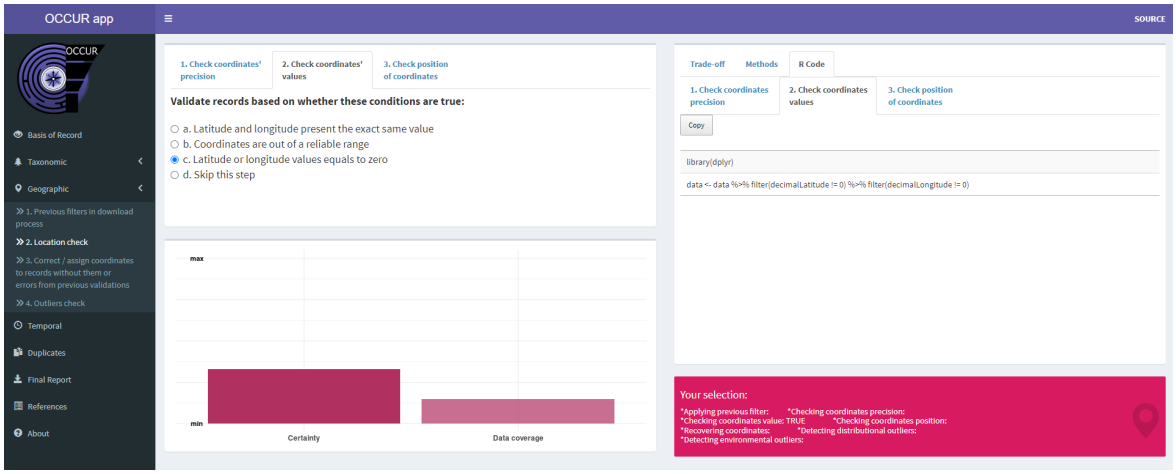
www.mncn.csic.es

La herramienta ya está disponible para la comunidad científica

Desarrollan OCCUR, una aplicación para la depuración de registros biológicos de especies

- La herramienta permite establecer los criterios para filtrar, validar y homogeneizar los datos adecuándolos a cada estudio científico
- Para el desarrollo de modelos predictivos y de distribución de especies se utiliza información de diferentes repositorios de datos

Madrid, 29 de mayo de 2024 La materia prima de la ciencia son los datos. Es a partir de ellos como se establecen hipótesis y se llega a conclusiones, de ahí la importancia de que sean de la mayor calidad posible. Para optimizar los registros que se utilizan en biología ambiental, un equipo de investigadores del Museo Nacional de Ciencias Naturales (MNCN-CSIC) ha desarrollado la aplicación OCCUR, que permite conocer y establecer los criterios para filtrar, homogeneizar y validar los datos que utilizan los equipos que trabajan en temas que van desde la distribución de especies a las predicciones de lo que ocurrirá con los seres vivos a partir de los cambios ambientales que estamos viviendo. La descripción y el método de desarrollo de la aplicación, que está a disposición de los grupos de investigación interesados, se ha publicado en la revista *Methods in Ecology and Evolution*.



Interface de la aplicación

Los registros de presencia de biodiversidad son datos sobre las especies animales y vegetales. Se puede tratar de la observación de un espécimen o de marcas de su rastro en un determinado lugar y momento. Además de la localización, estos registros pueden incluir datos ambientales o mediciones concretas del ejemplar. La recopilación

Comunicación MNCN
Tel.: 914 111 328 / Ext. 443450/3536
Móvil: +34 699 047 365
c-e: comunicacion@mncn.csic.es
c/ José Gutiérrez Abascal, 2.
28006, Madrid

de esta información recabada por parte de instituciones, personal investigador e incluso personas que participan en proyectos de ciencia ciudadana, ha crecido exponencialmente en las últimas décadas. Los datos se almacenan en repositorios de acceso libre o portales de datos que luego son utilizados para estudios científicos de diferentes ámbitos. El repositorio más importante a nivel mundial es el 'Global Biodiversity Information Facility' (GBIF) que cuenta con casi 3.000 millones de registros disponibles. "Gracias a estos repositorios de datos masivos podemos llevar a cabo mejores aproximaciones para evaluar ciertos aspectos del estado de la biodiversidad reduciendo los costes y recursos que implican los muestreo", expone la investigadora del MNCN Cristina Ronquillo. "Sin embargo, para utilizar estos recursos correctamente es necesario tomar unas medidas determinadas porque el criterio de cada colector es diferente. Esa fue la motivación principal para sacar adelante esta herramienta. Nuestro interés es promover una serie de buenas prácticas y protocolos en los procesos de limpieza y preparación de datos tal y como se haría en cualquier otra fase del análisis científico", continúa.

Cuando se utilizan datos de estos repositorios es fundamental tener en cuenta las limitaciones de los registros debido a los sesgos e imprecisiones que incluyen, así como su estandarización y armonización. Por ejemplo, si todos los registros están identificados a nivel de especie, si incluyen el autor del nombre científico, si las medidas están estandarizadas o detectar si el registro ha sido georreferenciado con las coordenadas de una ciudad cercana y o del lugar de muestreo. También es importante revisar los criterios para descargar o seleccionar determinados registros en función del estudio. "En nuestro trabajo diario con ecólogos descubrimos que había cierto desconocimiento de un gran número de criterios y herramientas que podían implementarse para trabajar con esta información. En este sentido es importante enfocar el proceso de depuración de datos en seleccionar aquellos que sean útiles para responder a la pregunta que nos hemos planteado y no en obtener los mejores datos", aclara Joaquín Hortal, del MNCN.

La aplicación OCCUR pone a disposición de la comunidad científica una herramienta fácil de usar que permite a quienes investigan conocer los caminos posibles a la hora de tratar los datos para valorar cuáles incluir en función de la cantidad y calidad de los registros disponibles. Esta aplicación sintetiza los criterios y métodos de procesamiento de registros propuestos por 25 trabajos anteriores, agrupándolos en cinco módulos: tipo de naturaleza del registro, taxonomía, geografía, información temporal y detección de duplicados. Al final, OCCUR genera un informe en el que aparecen los pasos seleccionados en cada caso, lo que facilita el desarrollo de los análisis y la inclusión, organización y escritura de los métodos en el artículo científico que describa el estudio en cuestión. "Cabe destacar que, en aquellos pasos en que ha sido posible, 'OCCUR' también proporciona código en el lenguaje estadístico R para ser incluido en los análisis de cada usuario", puntualiza Ronquillo.

OCCUR, el caso práctico de los musgos

La utilidad de OCCUR se ha comprobado en un trabajo recientemente publicado en la revista *Ecology and Evolution*, que analizaba los más de 9 millones de registros de musgos disponibles para la región templada del hemisferio norte. Los resultados de este trabajo destacan que los diversos métodos de procesamiento de registros mostraron diferencias notables en la diversidad de especies observadas en

Comunicación MNCN
Tel.: 914 111 328 / Ext. 443450/3536
Móvil: 699 047 365
c-e: comunicacion@mncn.csic.es
c/ José Gutiérrez Abascal, 2.
28006, Madrid

determinadas áreas de Europa y norte América, y con ello variaciones en las relaciones entre clima y biodiversidad medidas a partir de estos datos masivos. “Esto tiene consecuencias importantes, ya que significa que la calidad de los datos de partida que utilizamos para calibrar los modelos del impacto del cambio global puede alterar sus predicciones, lo que pone en evidencia la necesidad de realizar un trabajo minucioso con el procesamiento de los datos masivos de biodiversidad, que puedan replicar otros investigadores en el futuro", concluye Joaquín Hortal.

Ronquillo, C., Stropp, J., y Hortal, J. (2024). OCCUR Shiny application: A user-friendly guide for curating species occurrence records. *Methods in Ecology and Evolution*, 15, 816–823. https://doi.org/10.1111/2041-210X.14271.

Ronquillo, C., Stropp, J., Medina, N. G., & Hortal, J. (2023). Exploring the impact of data curation criteria on the observed geographical distribution of mosses. *Ecology and Evolution*, 13, e10786. https://doi.org/10.1002/ece3.10786

Comunicación MNCN
Tel.: 914 111 328 / Ext. 443450/3536
Móvil: 699 047 365
c-e: comunicacion@mncn.csic.es
c/ José Gutiérrez Abascal, 2.
28006, Madrid