



Colombia 50% Mar
INVAMAR

Instituto de Investigaciones Marinas y Costeras "José Benito Vives de Andrés"
Vinculado al Ministerio de Ambiente y Desarrollo Sostenible



Control de Calidad para los Datos de Biodiversidad

- *Principios y herramientas*



Qué significa Calidad de los Datos?

El grado en el cuál los datos cumple con los requisitos que los hacen confiables para ser usados para un propósito en particular.

- Sin errores
- Disponibles
- Verificables
- Compartidos con estándares
- Relevantes para la construcción de conocimiento y responder a las preguntas que se planteen.

La Calidad de los Datos puede verse afectada:

- Cuando se recogen los datos en campo
- Cuando se digitalizan
- Cuando se documentan
- Al almacenarse
- Durante su procesamiento
- Al colocarlos en un contexto particular
- Al presentarlos



Principios Básicos

La responsabilidad por la calidad de los datos debe asignarse a aquellos que los generan y esta debe hacerse incluso desde antes de que sean generados.

Obtener datos de calidad es un compromiso de todo el grupo de trabajo

Para ello las organizaciones deben establecer claramente procedimientos, estándares y herramientas para la documentación de los conjuntos de datos durante todo el ciclo completo de la gestión de los datos.

Principios básicos

- Piense en sus datos como activos de uso a largo plazo
- Establecer estrategias que PREVENGAN que se presenten errores en los datos: Políticas institucionales, estándares.
- Comparta las experiencias en la gestión de los datos
- Apóyese en herramientas tecnológicas que minimicen los esfuerzos orientados a corregir o detectar errores en los datos
- Si administra varios conjuntos de datos priorice las tareas orientadas a verificar la calidad de los datos, asigne mayor prioridad a los conjuntos de datos que son mas importantes para los usuarios
- Aplique técnicas de aseguramiento de la calidad
- Minimice la duplicación de los datos o redundancia (ej. Estaciones)
- Capacite a las personas que gestionan los datos y concientícelos de la importancia de la calidad de los datos.

Procedimiento

- Explore los datos
- Seleccione posibles errores
- Utilice la herramienta adecuada para corregir el tipo de error
- Aplique las herramientas
- Valide los resultados
- Documente los cambios

Definiciones

- Exactitud/ Precisión
- Incertidumbre
- Error al azar o error sistemático



Herramientas

OBIS ofrece herramientas que verifican la calidad de los datos online a fin de que los usuarios revisen sus datos sin necesidad de que estos se encuentren en el Sistema y herramientas incorporadas a la base de datos que se aplican a los datos que ingresan al sistema o están integrados a él.




Banderas de Calidad en OBIS

- La bandera de calidad indica el grado en el cual un registro se acerca a un objetivo de calidad en particular
- OBIS aplica 18 chequeos de calidad a partir de los cuales asigna un valor de calidad a un registro, cada uno de los chequeos tiene solamente dos estados cumple/no cumple.

Validación taxonómica





WoRMS Taxon match

You can use the WoRMS Taxon Match Tool ([credits](#)) to automatically match your species list or taxon list with WoRMS. After matching, the tool will return your file with the AphiaID's, valid names, authorities, WoRMS classification and/or any other output you selected. [\[View manual\]](#)
For performance reasons, the limit is set to 5,000 rows. You can send larger files to info@marinespecies.org and we will return the results to you as soon as possible.

File *
Allowed filetypes: Plain text [TXT], Comma Separated [CSV] & Excel Sheet [XLS, XLSX]

Row delimiter First row contains column names

Column delimiter

Match authority

Match upto Higher taxa only possible if a full classification is given in additional columns

Limit to taxa belonging to

Output AphiaID LSID TSN ScientificName Authority Accepted name Classification Qualitystatus
 Taxon status Environment Citation

- RSS
- Add provider
- @WRMarineSpecies

Validación Geográfica

La descripción de una localidad debe tener el número de elementos necesarios para localizar con la mayor precisión posible el lugar de la captura o el avistamiento.

La validación geográfica incluye chequeos para:

- Latitud y longitud diferentes de cero
- Latitud en el rango válido de -90 a 90
- Longitud entre 180 y -180
- Coordenadas caen en un área marina
- Validación de la profundidad a partir de la Carta Batimétrica General de los Océanos (GEBCO) con un rango de precisión de 100m
- Datos que caen dentro de un área geográfica claramente delimitada por el Sistema

Algunas inconsistencias se validan por procesos en lotes que no están disponibles online. Los errores en los datos son reportados a quienes proveen o figuran como custodios de los datos

Validación de otras variables

- Fechas
- Datos de investigadores
- Datos de instituciones
- Datos descriptivos: Habitat, status de conservación, caracteres morfológicos, caracteres fisiológicos
- Numero de individuos
- Comportamientos
- Tamaños, peso
- Sexo
- Unidades de medida
- Etc.

Validación de otras variables

En general se recomiendan listas de valores codificados, validación de máximos y mínimos plausibles considerando las unidades de medida y la cobertura temporal de las colectas.

- Hojas de cálculo: Excel
- OpenRefine (Google)
- R (scripts)

Muchos de estos errores se pueden controlar mediante la aplicación de herramientas electrónicas para la captura de datos.



Fuentes

Vandepitte L., 2015. Quality control of biodiversity data: tolos & techiques

Otras fuentes

Guía de utilización de OpenRefine por SiB Colombia.

<http://www.gbif.org/resource/81272>

Taller de calidad de datos: mejorando los datos primarios sobre biodiversidad

<https://goo.gl/itKZXQ> (google drive)

Leen Vandepitte, Samuel Bosch, Lennert Tyberghein. et al. 2015. Fishing for data and sorting the catch: assessing the data quality, completeness and fitness for use of data in marine biogeographic databases. doi: 10.1093/database/bau125.