

Validación automatizada de datos de biodiversidad

Autores: F. Muñoz¹, S. Kremer¹, J. Marquez¹ y S. Orellana¹.

Afiliaciones:

1. Cienciambiental Consultores S.A.

Abstract

La gestión de datos de biodiversidad ha representado un desafío en los proyectos de evaluación ambiental. En este contexto, como consultora ambiental hemos trabajado en el diseño de un sistema integral de gestión de datos que otorgue calidad y trazabilidad a cada registro. La propuesta contempla diversas herramientas y procesos, de los cuales los validadores automáticos constituyen una parte. Este sistema se estructura en tres etapas principales: (1) levantamiento de datos en terreno, (2) procesamiento en gabinete y (3) almacenamiento final en un set de datos histórico desde el cual es posible generar distintos productos. En la primera etapa, el registro de datos se realiza mediante un formato estandarizado que facilita su posterior procesamiento y compilación. Este formato, común entre consultoras ambientales, cuenta con campos bien definidos y acotados, además de vocabularios controlados para algunos de ellos. Sin embargo, al ser completado por los especialistas en terreno, es en esta etapa donde se concentran la mayoría de los problemas, como errores de digitación, inconsistencias entre campos o registros incompletos que comprometen la calidad de la información final. Para abordar esta problemática, hemos desarrollado un sistema de validación de formularios de terreno utilizando R, principalmente con los paquetes `{tidyverse}`, `{readxl}` y `{janitor}`. Estos validadores están orientados a datos de fauna, flora y limnología, y se implementan como scripts que aplican controles de calidad sobre cada formulario recibido, incluyendo la detección de valores faltantes, revisión de formatos de fecha y coordenadas, y comprobación de coherencia interna y entre campos, además de validaciones específicas según el método. Estos procedimientos permiten identificar errores de manera inmediata y estandarizar los datos antes de su incorporación a las siguientes etapas del proceso. Asimismo, los códigos de validación estarán disponibles en una aplicación Shiny, desde la cual cualquier especialista podrá verificar si sus datos se encuentran correctamente tabulados y conformes al estándar, obteniendo reportes automáticos de errores u observaciones. Durante la charla se abordará la relevancia del proceso de estandarización de datos en formatos predefinidos, elemento fundamental tanto para el análisis de datos como para el desarrollo de aplicaciones y soluciones basadas en código. También se discutirá el diseño e implementación de la aplicación Shiny y los desafíos enfrentados, tanto desde el punto de vista computacional como desde la gestión de la información. En última instancia, la implementación de estos validadores busca mejorar la eficiencia

en la detección de errores, reducir la necesidad de revisiones manuales extensas y otorgar autonomía a los colaboradores, quienes pueden corregir y validar sus propios datos mientras se familiarizan con los estándares propuestos. De esta manera, se desplaza el control de calidad hacia etapas tempranas del flujo de trabajo, reduciendo la propagación de errores y asegurando una mayor consistencia en las etapas posteriores.

Palabras clave: Gestión de datos de biodiversidad; Estandarización; Validación de formularios; Aplicación Shiny; Calidad y trazabilidad