Autora: Andrea Arriaga-Madrigal

Modificación por: Andrea Arriaga-Madrigal

Fecha de última modificación: 2020-05-05

**Monitoreo Caribe: Protocolo de manejo y análisis de datos**

Este documento explica la información que debe contener cada documento y cada carpeta y la forma estandarizada de nombrar cada uno de los elementos para facilitar el manejo y análisis de la información de forma sistematizada.

**Generalidades del manejo de información**

Para poder trazar los cambios efectuados en los documentos o datos (información muy sensible) siempre se debe registrar quien modificó el documento y de ser pertinente, la fecha. Cualquier documento que contenga datos de campo, debe ser modificado generando una copia, a la cual se le agrega al final del nombre del archivo, la persona que realiza las modificaciones. Se nombran con la fecha, el tipo de documento y el código de buzo que modificó el documento, cada elemento separado por un guion bajo (e.g. 2019-10-04\_peces\_anarma). Con respecto a los scripts de r y los protocolos, no es necesario guardar copias nuevas, si las modificaciones son menores. En este caso, se debe registrar al inicio de los documentos, la autoría de la modificación y la fecha.

Para facilitar la agrupación de los datos de campo y el análisis, se concatenan las bases de datos a partir de una columna de referencia, que casi siempre es un **código** (e.g. codigo de inmersión que es el identificador único de cada buceo). Los códigos solo funcionan si están escritos exactamente igual en todas las bases de datos, de manera que, cuando se introduce un código en una casilla, lo ideal es copiarla/pegarla en lugar de digitarla manualmente.

Todas las hojas de datos tienen una primera hoja con los **metadatos**, que abarcan toda la información requerida para comprender qué significan las columnas, sus unidades, especificaciones y aclaraciones de cómo se llenan. Esto permite que la hoja de ingreso posea únicamente datos numéricos o categóricos, sin necesitar aclaraciones que generen errores a la hora de agrupar y analizar la información

**Especificaciones de la digitación de textos**

* **No usar caracteres especiales.** Las tildes, “ñ”, diéresis u otros caracteres del español que no se comparta con el idioma inglés deben omitirse. Los nombres propios de sitios, localidades, monitores, instituciones y demás se escriben sin caracteres especiales.
* Usar **minúsculas** siempre como primera opción. La excepción son los datos extraídos de otras bases de datos que ya estén con mayúsculas, que se estén copiando/pegando o al referirse a especies o taxones, cuyos nombres en la literatura se encuentran en mayúscula. En este caso, deben permanecer de esa forma para poder descargar la información de repositorios en línea. Todos los codigos (de buzo, inmersiones y sitios), nombres de sustratos que no son especies y demás categorías ingresadas en campo, en minúscula.
* Es preferible usar **caracteres que no sean variables** e.g. es mejor usar 0 y 1 en lugar de si y no por todas las variantes en la escritura (sí, SI, si, Si o no, NO, No).

**Nombramiento de columnas y hojas de datos:** Se nombran en minúscula, sin espacios, paréntesis, comas ni puntos. Las palabras se dividen usando guion bajo (\_). No centrar ninguna celda ni cambiar el formato de ninguna forma. Al ingresar datos en campo, las especificaciones de unidades o variantes se aclaran en la hoja de metadatos.

**Fechas:** El formato en todos los casos va a ser “aaaa-mm-dd” (i.e. año con cuatro números, mes y día con dos dígitos, separados por guion –)

**Sitios:** El nombre de los sitios lo establecen los investigadores. Cada sitio comprende un conjunto de profundidades, coordenadas y ambientes. Deben evitarse los siguientes errores comunes:

* Generar varios nombres para un sitio (e.g. Sandalo 1 y Sandalo 2; La Viuda Noreste y La Viuda Sur). No importa que el muestreo sea en una coordenada distinta, ambiente diferente o a una profundidad nueva, si es parte del mismo sitio, se le asigna el nombre del sitio que se encuentra en la base de datos de sitios, que maneja el líder de gira.
* Si un sitio se llama Playa Chiquita, no omitir “Playa” y simplificarlo a Chiquita. Debe escribirse el nombre igual a como está en la base de datos de sitios, incluyendo artículos y mayúsculas. (e.g. Roca La Viuda).

**Organización de las carpetas**

Hay una carpeta para cada tipo de información. De forma que, al subir nuevos datos, deben ser distribuidos en las carpetas correspondientes y no hacer nuevas que generen duplicados. Solo en algunos casos muy particulares se van a generar nuevas carpetas, siempre después de la aprobación de los encargados de la organización del proyecto. A continuación, se desglosan las carpetas existentes y el tipo de información que contienen.

**Permisos:** Información de gestión del proyecto, solo es modificado por los investigadores

**Literatura de soporte:** Artículos, libros, protocolos que sean de utilidad para el desarrollo del monitoreo.

**Datos:** Contiene la información necesaria para las giras, tanto para los asistentes como para el/la líder de gira (Fig. 1) y los espacios adecuados para almacenar la información generada en cada salida al campo.

**Respaldo fotos/videos:** Material audiovisual recopilado en las giras, tanto de los muestreos como del resto de actividades.

**field\_leader:** Documentos que deben ser actualizados en cada gira por el líder (i.e. base taxonómica de peces, sésiles y móviles, base de submarinistas, inmersiones y sitios). La carpeta **machotes\_asistentes** incluye los machotes actualizados que deben ser entregados a los asistentes para ingresar los datos en campo (i.e. peces, móviles, sésiles y sésiles enfermedades). Cada documento que se modifique debe ser guardado como una versión nueva, con la fecha actualizada. Usar el formato “aaaa-mm-dd\_tipo de documento\_persona que modifica”.

**data\_raw:** Contiene las hojas de datos tomadas en campo y los respaldos fotográficos de las hojas. Cada hoja contiene los datos de un tipo de censo, de un submarinista de una gira.

En la carpeta **gps** se encuentran los puntos marcados con el gps en campo.

La carpeta **Fotos hojas campo** contiene respaldos fotográficos de las hojas de campo. Se guardan todas las fotos de las hojas de campo de la gira en un solo documento pdf llamado “datos\_aaaa-mm”. Dentro del pdf, deben agruparse las fotografías por día y por submarinista, cada hoja debe estar rotulada con el día, submarinista y datos de la inmersión.

En la carpeta llamada **photoquad** se encuentran los documentos que genera photoquad en formato “.cellcount” con el porcentaje de cobertura de cada especie por cuadrícula. También hay un documento llamado “protocolo\_photoquad”, donde se desglosa los pasos para realizar el análisis en este software.

**integrate\_caribe:** Es un script en r, que se debe correr al inicio en que se va a trabajar con cualquier otro script del proyecto. Llama los paquetes generales que se necesitan para la manipulación de información y generación de resultados y establece el directorio de trabajo. También especifica algunos aspectos de formato para los gráficos o mapas que se generan.

**creation\_code:** Códigos en R que toman los datos crudos, los limpia, los concatena y generan nuevos documentos con los datos agrupados para su futuro análisis. Cada código está diseñado para ser corrido por cualquier persona. Para esto, cada código debe contener las librerías requeridas y todas las líneas de código para que hagan su función en cualquier computadora. Los datos trabajados que se generan con cada código se guardan en la carpeta **datos\_trabajados.** Las hojas de datos se deben nombrar como “tipo de datos\_localidad\_especificaciones” e.g. “sesiles\_Playa Chiquita”

**intermediate\_data:** Hojas de datos que fueron generadas con los códigos de creación. Cada vez que se corre el código de creación, se actualizan. Si se necesitan modificar manualmente, hacer una copia con otro nombre.

**analysis\_code:** Código en R para analizar los datos trabajados. Al inicio del código se llama los datos trabajados y con estos se generan gráficos, se calculan índices o se realiza cualquier tipo de análisis. Todas las figuras que se generan se guardan en la carpeta **figuras**.

**figures:** Los mapas, gráficos, esquemas, que sean generados con los códigos de análisis.

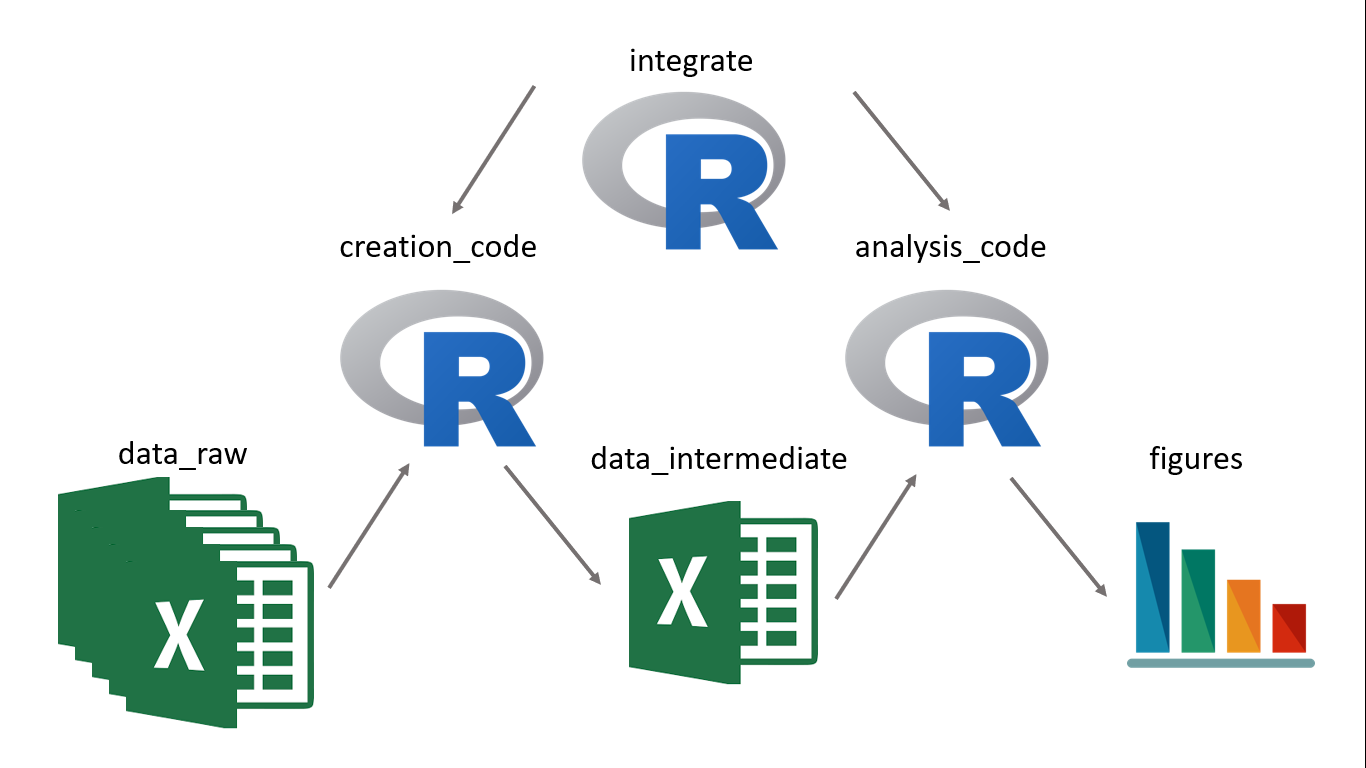


Fig. 1. Esquema que representa el papel de cada carpeta y su papel en el procesamiento de los datos generados en campo.