**Curso: Fundamentos de Gestión y Estandarización de Datos Ambientales**

**Duración total:** 7 horas (incluye 1 hora de introducción, breaks y cierre).  
**Modalidad:** Taller práctico en Power BI (Power Query para limpieza, modelo de datos para estructura, y documentación integrada).  
**Requisitos previos:** Conocimientos básicos de Excel y Power BI; archivo "Peces 2004 - 2024.xlsx" como dataset de práctica (datos de muestreo de peces en Panamá, con hojas desordenadas: Estaciones, Lista, Compilado, Pivot).  
**Objetivo general:** Aprender a organizar, estandarizar y documentar bases de datos ambientales de forma robusta, desde hojas de Excel desordenadas hasta una base estructurada y referenciada.  
**Materiales:** Power BI Desktop (gratuito), archivo Excel proporcionado.  
**Evaluación:** Práctica final: transformar el dataset en una base documentada y validada.

**Glosario General del Curso**

* **Datos tidy:** Estructura donde cada fila es una observación, cada columna una variable, y cada celda un valor único (según Hadley Wickham).
* **Power Query:** Editor ETL (Extract, Transform, Load) en Power BI para limpiar y transformar datos.
* **Modelo de datos:** Esquema relacional en Power BI con tablas de hechos (mediciones) y dimensiones (referencias).
* **Glosario ambiental:** Tabla de referencia para términos estándar (e.g., "Vertiente" como "P" para Pacífico).
* **Nomenclatura taxonómica:** Nombres científicos validados (e.g., vía GBIF: Global Biodiversity Information Facility).
* **DAX:** Lenguaje de fórmulas en Power BI para medidas de validación.
* **Metadatos:** Descripciones de datos (e.g., origen, formato).

**Buenas Prácticas Transversales**

* Siempre documenta cada paso en Power Query (renombra queries y pasos).
* Usa nombres de columnas descriptivos y en minúsculas (e.g., "especie\_cientifica").
* Valida datos contra fuentes externas offline (archivos descargados) para evitar dependencias de internet.
* Evita datos duplicados: prioriza claves únicas (e.g., códigos de especie).
* Trabaja en copias: no modifiques el Excel original.
* Exporta backups regulares del PBIX.

**Introducción (0.5h)**

* Presentación del dataset: Hojas "Estaciones" (sitios de muestreo), "Lista" (especies con códigos), "Compilado" (datos crudos de abundancia por año/estación), "Pivot" (resumen básico). Problemas comunes: duplicados en especies, formatos inconsistentes, datos no tidy.
* Demo rápida: Abrir Power BI, importar el Excel.
* Objetivo: Motivación para estandarizar datos ambientales (e.g., análisis de biodiversidad en cuencas como el Canal de Panamá).

**Bloque 1: De Excel a Base de Datos Estandarizada (1.5h)**

**Objetivo:** Transformar hojas sueltas en una base de datos unificada y analizable.

**Contenidos**

* Importación masiva.
* Estructura tidy.
* Manejo de errores.
* Uso avanzado de Power Query.

**Tutorial Paso a Paso**

1. **Importar datos:** En Power BI, ve a "Inicio > Obtener datos > Excel". Selecciona todas las hojas del archivo.
2. **Combinar tablas (append):** En Power Query (clic derecho en una query > Editar), selecciona "Compilado" como base. Append "Estaciones" y "Lista" para unir metadata. Renombra la query a "datos\_crudos".
3. **Estructura tidy:** Identifica no-tidy (e.g., columnas por estaciones en "Compilado"). Usa "Despivotar columnas" en las columnas de estaciones (e.g., ACL, B-67) para convertir en filas (variable: "estacion", valor: "abundancia").
4. **Manejo de errores:** Cambia tipos (e.g., "AÑO" a entero, "Date" a fecha ISO yyyy-mm-dd). Usa "Reemplazar valores" para NA (e.g., blanks a "Desconocido"). Rellena ausentes con "Rellenar abajo" en columnas como "Familia".
5. **Combinar (merge):** Merge "datos\_crudos" con "Estaciones" por "Estación" para agregar latitud/longitud.
6. **Resultado:** Tabla maestra "datos\_tidy" con ~440 filas limpias, fechas ISO, valores consistentes. Exporta como CSV para backup.

**Buenas Prácticas Específicas**

* Filtra filas irrelevantes temprano (e.g., headers como "row1").
* Usa "Eliminar columnas" para redundancias (e.g., "CCODE" si ya tienes "PCODE").
* Prueba con muestras pequeñas antes de procesar todo.

**Ejercicio Práctico (30 min)**

Transforma "Compilado" en tidy y merge con "Estaciones". Verifica: ¿Cuántas observaciones por año? (Usa "Agrupar por" en Power Query).

**Bloque 2: Estandarización con Glosarios Ambientales (1h)**

**Objetivo:** Unificar términos en campos clave mediante diccionarios.

**Contenidos**

* Creación/importación de glosarios.
* Transformaciones con merge.
* Detección de atípicos.
* Sustitución estándar.

**Tutorial Paso a Paso**

1. **Crear glosario:** En Power Query, crea nueva tabla "glosario\_ambiental" con columnas: Término Original, Estándar (e.g., "P" → "Pacífico", "C" → "Caribe", "P-C" → "Ambas"). Importa desde "Lista" (extrae únicas de "Vertiente").
2. **Merge para estandarizar:** Merge "datos\_tidy" con "glosario\_ambiental" por "Vertiente". Expande columna "Estándar" y elimina original. Repite para "Ambiente" (e.g., "M" → "Marino", "D" → "Dulce").
3. **Detección de atípicos:** Usa "Agrupar por" para listar únicas en "Vertiente"; filtra variantes (e.g., "P " con espacio). Reemplaza con "Reemplazar valores" (e.g., "Manglar" → "Manglares" a "Manglar Estándar").
4. **Resultado:** Tabla "datos\_estandarizados" con campos controlados (e.g., solo "Pacífico", "Caribe").

**Buenas Prácticas Específicas**

* Mantén glosarios en tablas separadas para reutilización.
* Usa fuzzy matching en merge para errores de escritura (opción avanzada).
* Documenta cambios en una columna "fuente\_original".

**Ejercicio Práctico (20 min)**

Estandariza "Ambiente" y "V. Económico" usando glosario de "Lista". Verifica consistencia.

**Bloque 3: Normalización de Nombres de Especies (1h)**

**Objetivo:** Establecer nomenclatura científica única y evitar duplicados.

**Contenidos**

* Integración con catálogo externo.
* Relaciones 1:N.
* Detección de variantes.
* Unión con claves taxonómicas.

**Tutorial Paso a Paso**

1. **Obtener catálogo externo:** Usa un archivo descargado de GBIF (simula: crea tabla "catalogo\_especies" con datos de "Lista", agregando taxonomía: Reino=Animalia, Filo=Chordata, etc. Para real, busca "GBIF species list Panama fish" y descarga CSV).
2. **Homologar nombres:** Merge "datos\_estandarizados" con "catalogo\_especies" por "Especies" (usa "CODE" como clave). Expande para agregar "especie\_cientifica" validada. Maneja NA con "Reemplazar valores" a "Especie no identificada".
3. **Detectar variantes:** Agrupa por "Especies" para contar únicas; filtra truncados (e.g., "A. SP" → "Ariopsis sp"). Usa relaciones 1:N en modelo de datos (especie vernácula → científica).
4. **Unión taxonómica:** Agrega columnas de taxonomía (e.g., Familia de "Lista").
5. **Resultado:** Campo "especie\_cientifica" limpio, sin duplicados (e.g., "Cichla monoculus" unificado), trazable a GBIF.

**Tutorial Paso a Paso (Continuación, Integrando Datos Externos)**

Basado en la búsqueda, descarga el CSV de GBIF Backbone Taxonomy (enlace: <https://hosted-datasets.gbif.org/datasets/backbone/>) o el dataset de peces de Panamá (e.g., <https://www.gbif.org/dataset/3ea542dc-885d-4c8a-91a3-64d0a5d78301>). Importa a Power BI como "catalogo\_gbif". Filtra por "Panama" y "fish". Merge con tu tabla por coincidencia aproximada en "Especies" (usa "Fuzzy merge" si disponible en versión pro).

**Buenas Prácticas Específicas**

* Prioriza nombres científicos validados para evitar sesgos locales.
* Maneja sinonimias (e.g., nombres obsoletos) con una columna "sinonimo".
* Actualiza el catálogo anualmente.

**Ejercicio Práctico (20 min)**

Normaliza "Especies" en "Compilado" contra "Lista" (simulando GBIF). Elimina duplicados como "Cichla monoculus" repetido.

**Bloque 4: Construcción de Base de Datos en Power BI (1h)**

**Objetivo:** Modelar la base para futura migración a SQL.

**Contenidos**

* Separación por tablas.
* Relaciones (modelo estrella).
* Verificación de claves.
* Configuración de formatos.

**Tutorial Paso a Paso**

1. **Separar tablas:** En Power Query, crea: "hechos" (de "Compilado": abundancias, fechas, estaciones); "dimensiones" (e.g., "dim\_especies" de "Lista", "dim\_sitios" de "Estaciones").
2. **Relaciones:** En "Modelo" vista, conecta "hechos" a "dim\_especies" por "CODE" (1:N), a "dim\_sitios" por "Station ID" (1:N). Crea estrella: hechos en centro.
3. **Verificación:** Usa "Administrar relaciones" para chequear duplicados (e.g., claves únicas en dimensiones).
4. **Formatos:** En "Modelo", formatea "Date" como yyyy-mm-dd, "abundancia" como entero, "Vertiente" como categoría.
5. **Resultado:** Base robusta, lista para SQL export (usa "Exportar > Datos" a CSV, luego importa a SQL tools).

**Buenas Prácticas Específicas**

* Usa modelo estrella para eficiencia en consultas.
* Evita ciclos en relaciones.
* Prueba integridad: ¿Hay huérfanos? (e.g., estaciones sin metadata).

**Ejercicio Práctico (20 min)**

Modela relaciones entre "Compilado" y "Estaciones". Verifica claves únicas en "Station ID".

**Bloque 5: Documentación Integrada del Sistema (1h)**

**Objetivo:** Registrar decisiones para interoperabilidad.

**Contenidos**

* Metadatos en Power BI.
* Log de transformación.
* Exportación de esquema.
* Página de documentación.

**Tutorial Paso a Paso**

1. **Metadatos:** En "Modelo" vista, selecciona columna > "Propiedades" > agrega descripción (e.g., "Especies: Nombre científico validado vía GBIF"). Configura categorías.
2. **Log de transformación:** En Power Query, copia pasos aplicados (vista "Pasos aplicados") a nueva tabla "log\_transformaciones".
3. **Exportación:** Usa "Archivo > Exportar > Plantilla" para esquema; genera diccionario automático con script DAX o herramienta externa.
4. **Página de documentación:** En "Informe" vista, crea página "Docs" con tablas visuales: glosario, log, esquema.
5. **Resultado:** PBIX autocontenible con metadatos (e.g., exportable a Darwin Core para biodiversidad).

**Buenas Prácticas Específicas**

* Usa estándares como Dublin Core para metadatos.
* Incluye versión y fecha en logs.
* Haz docs accesibles (e.g., exporta a PDF).

**Ejercicio Práctico (20 min)**

Documenta "datos\_tidy" con metadatos y log. Crea página de informe.

**Bloque 6: Validación y Checklist Final (0.5h)**

**Objetivo:** Validar calidad y consistencia.

**Contenidos**

* Medidas DAX.
* Revisión de inconsistencias.
* Checklist.

**Tutorial Paso a Paso**

1. **Medidas DAX:** Crea: Inconsistentes = COUNTROWS(FILTER(hechos, ISBLANK(especie\_cientifica))) para NA críticos. Duplicados = SUMMARIZE(hechos, [CODE], "Count", COUNT()) con filtro >1.
2. **Revisión:** En informe, visualiza: tabla de NA, gráficos de duplicados, claves huérfanas (e.g., merge left outer para detectar).
3. **Checklist:**
   * ¿Datos tidy? (Sí/No)
   * ¿Campos estandarizados? (Contra glosario)
   * ¿Especies normalizadas? (0 duplicados)
   * ¿Relaciones válidas? (Sin errores)
   * ¿Documentado? (Metadatos completos)
   * ¿NA manejados? (<5% críticos)
4. **Resultado:** Validación completa; hoja de ruta: "Mantener glosario actualizado, migrar a SQL para escalabilidad".

**Buenas Prácticas Específicas**

* Automatiza validaciones con DAX para reusar.
* Registra issues en log.

**Ejercicio Práctico (10 min)**

Ejecuta DAX para validar "Compilado". Completa checklist.

**Cierre (0.5h)**

* Resumen: Aplicación a datos ambientales reales (e.g., monitoreo de peces en Panamá).
* Recursos adicionales: Documentación Power BI, guías GBIF.
* Q&A y entrega de certificado.