



**The University of Reading
Statistical Services Centre**

**Biometrics Advisory and
Support Service to DFID**



LINEAMIENTOS PARA ARCHIVO DE DATOS LECCIONES DE UN ESTUDIO DE CASO

1999



**UNIVERSIDAD
NACIONAL
COSTA RICA**



CONTENIDO

Presentación.....	3
1. Introducción.....	5
2. ¿Por qué conservar los datos de un proyecto?.....	5
3. ¿Qué debe ser archivado?.....	6
4. ¿Qué es un buen archivo de datos?.....	6
5. Medios de diseminación.....	7
6. Ejemplo de una actividad de archivo de datos exitosa.....	8
7. ¿Qué comprende el archivo?.....	9
8. Información adicional.....	9

Presentación

La presente publicación es parte de una serie de guías biométricas producidas originalmente en idioma inglés, por el Centro de Servicios Estadísticos (SSC) de la Universidad de Reading, Reino Unido, para el Departamento para el Desarrollo Internacional (DFID) del Gobierno. La serie está dirigida a personal técnico involucrado en la presentación y ejecución de proyectos en recursos naturales, con el propósito de ayudar a los investigadores a identificar sus necesidades estadísticas y biométricas.

El SSC y el DFID aprobaron la solicitud de la Universidad Nacional (UNA, Costa Rica) para proceder a la traducción, edición, publicación y distribución de las guías en idioma castellano, con el propósito de extender sus beneficios a proyectos e investigadores de países hispanohablantes, fundamentalmente del área de Centro América y el Caribe.

La traducción y edición de la versión castellana estuvo a cargo del Dr. Jorge Camacho-Sandoval, Escuela de Ciencias Agrarias, Universidad Nacional; los Drs. Gilberto Páez y Christoph Kleinn del Centro Agronómico Tropical de Investigación y Enseñanza (CATIE, Costa Rica) realizaron la revisión técnica de la versión castellana. La aprobación final estuvo a cargo del M.Sc. Carlos Barahona del SSC. Además el CATIE ha financiado la traducción, publicación y distribución de las cinco primeras guías.

La UNA y el CATIE agradecen al SSC y el DFID por haber acogido su iniciativa para la difusión de las guías en castellano y por haber ofrecido todas las facilidades para ello.

UNIVERSIDAD NACIONAL

CATIE

1. Introducción

Una parte integral de muchos proyectos de investigación es la recolección de datos de encuestas o experimentos, a un considerable costo en tiempo y esfuerzo. Se debe tener gran cuidado para producir datos de buena calidad, tanto en la etapa de recolección como en la de introducción en la computadora, pero usualmente se hace poco énfasis en asegurarse que los datos estén disponibles para otros usuarios, en forma tal que permita que los datos sean realmente comprendidos y usados correctamente en subsecuentes estudios. Idealmente la creación de un archivo tiene que estar integrada con el trabajo en progreso de un proyecto, más que ser una idea tardía en el momento en que el equipo de investigación se ha dispersado.

Esta breve guía pretende hacer conciencia sobre la importancia de conservar los datos de un proyecto, discutir las características que hacen un buen archivo de datos y ofrecer un ejemplo de un archivo exitoso.

2. ¿Por qué conservar los datos de un proyecto?

En el pasado, los registros generalmente disponibles de un proyecto habían sido a menudo solamente en forma de publicaciones, en donde se disponía de espacio limitado para resumir los aspectos clave, relevantes para el sesgo específico de los artículos relacionados. Los recursos modernos de computación y almacenaje de datos significan que ahora no hay razón por la cual información mucho más detallada no pueda ser conservada y fácilmente reproducida, de forma que pueda ser accesible y utilizada por otros.

Como parte del apoyo a ciertos proyectos, los proponentes pueden argüir que la información cuantitativa producida podría ser de interés en el largo plazo o en un amplio rango de situaciones y que ciertamente, producir resultados solamente de valor efímero no debe recomendar a un proyecto. Esto implica la obligación del equipo del proyecto de documentar y archivar los datos recolectados en el transcurso del trabajo.

Dado un proyecto importante, tal registro es potencialmente valioso para usuarios secundarios y posteriores investigadores, si se les da la oportunidad de extraer información de forma que haga su propio trabajo más efectivo. Garantiza este valor añadido y fortalece la posible financiación del proyecto inicial.

3. ¿Que debe ser archivado?

Hay tres tipos principales de información que requiere ser registrada con precisión:

- Los datos principales del proyecto en sí mismos, no solamente cuadros resúmenes.
- El registro de cómo y por qué los datos fueron recolectados y que representan.
- La documentación sobre los archivos de computadora que permitirán la recuperación posterior de los datos.

Para que sean útiles más allá del plazo de ejecución del proyecto, los archivos necesitan tener una forma organizada, en casi todos los casos computarizada.

Los archivos deben tener copias de seguridad, con una versión maestra almacenada de forma segura y debe tener un sistema, desarrollado durante el proyecto, para que usuarios autorizados puedan hacer copias parciales o totales posteriormente.

Los principios generales del control de calidad de datos se aplican en todas las etapas, es decir, durante la definición y desarrollo de los datos que serán colectados, durante la recolección de datos y durante la creación de los archivos de datos computarizados definitivos.

4. ¿Que es un buen archivo de datos?

Varias características determinan la obtención de un buen archivo de datos. En breve:

- Accesibilidad, de manera que los usuarios puedan acceder a la información almacenada mediante programas de computadora ampliamente disponibles.
- Facilidad de uso, para asegurar que (i) la estructura del archivo de datos es simple, de manera que la relación entre los formularios usados en el campo y la información computarizada sea evidente; (ii) halla definiciones claras de las variables almacenadas en el archivo (por ejemplo unidades de medida) y de los códigos usados (nombre de las

variables categóricas, etc.); y (iii) halla consistencia en nombres, códigos, unidades de medida y abreviaturas en todo el archivo.

- Se debe asegurar la fiabilidad con un archivo libre de errores que pueda ser manejado dentro del plazo y presupuesto del proyecto.
- Documentación a saber (i) procedimientos usados para la recolección de datos, incluyendo la metodología y las unidades de muestreo utilizadas, (ii) estructura del archivo, es decir, como están relacionados los diferentes directorios, (iii) una lista de los diferentes directorios de computadora que comprende el archivo, (iv) una lista completa de todas las variables, incluyendo anotaciones sobre como son tratados los valores faltantes, (v) estadísticos resumen que permitan al usuario realizar chequeos cruzados de si la información recuperada se corresponde con la requerida, y (vi) advertencias y comentarios relevantes relativos a cualquier parte de la base de datos.
- Preservación del anonimato o cualquier condición de confidencialidad por la cual las fuentes hicieron disponibles los grupos de datos.
- El archivo debe ser tan completo como sea posible y útil. Debe incluir una copia computarizada de (i) los formularios de campo, (ii) la bitácora de manejo de datos, (iii) descripción de las variables derivadas, y (iv) comentarios y observaciones especiales.

5. Medios de diseminación

Los medios para la diseminación de datos del proyecto tienen que ser considerados en la planificación de cualquier forma de archivo. De esta manera se puede hacer la selección de los aspectos que serán diseminados, los cuales pueden ser claramente selectivos, por ejemplo, en razón de la confidencialidad de algunos datos.

La argumentación hecha en la actualidad es ésta:

- Es más fácil y barato duplicar discos de computadora que fotocopiar cantidad de informes, y
- Es más fácil reutilizar información numérica si es diseminada en forma que pueda ser accesible directamente por la computadora.

En la actualidad es posible diseminar datos en discos compactos. Esto permite incluir datos de SIG (sistemas de información geográfica) y también grandes imágenes con programas específicos incluidos para que el usuario pueda verlas. Actualmente el equipo requerido para escribir discos compactos es algo elevado, pero los precios están cayendo y puede que pronto estén aún proyecto con presupuestos pequeños.

Por el momento parece que objetos como fotos aéreas deben ser almacenadas en un lugar donde puedan ser conservadas de manera segura por un plazo relativamente largo y que tenga la capacidad de copiar negativos y positivos para usuarios autorizados. Por supuesto, detalles de como obtener copias, deben ser parte del archivo de información.

6. Ejemplo de una actividad de archivo de datos exitoso.

El Statistical Services Centre (SSC) en la Universidad de Reading estuvo muy involucrado con los aspectos estadísticos del Estudio Estatal de Utilización de Tierras (ELUS) en Malawi. Como parte de ésta participación, se realizó una propuesta para archivar el gran volumen de información que había sido recolectada. La propuesta fue apoyada por el asesor en recursos naturales en Malawi y el equipo del proyecto ELUS.

Un miembro del personal del SSC, quién tiene considerable experiencia en todos los programas de computadora utilizados, visitó Malawi por tres semanas en Febrero de 1997, para conducir la actividad de archivo. Fue necesaria una semana adicional después de la visita para completar el archivo y su documentación. El tiempo necesario para dicho trabajo depende del tamaño de los cuestionarios considerados y la calidad de los datos disponibles para el consultor en archivo de datos. Mientras que el equipo de ELUS puso mucha atención para asegurarse de que sus principales bases de datos estuvieran libres de errores en la medida de lo posible, advertimos, con base en otras experiencias, que la "limpieza" de datos puede consumir inmensas cantidades de tiempo.

La actividad de archivo consideró un taller de un día de duración con algunos usuarios identificados de la base de datos del proyecto ELUS. El objetivo fundamental fue familiarizar a los participantes con la estructura y organización del archivo y obtener sus puntos de vista sobre como mejorar el procedimiento de archivo.

El archivo exitoso de las bases de datos de ELUS en el lapso de tiempo descrito, fue posible en el tanto que el trabajo fue aprobado, financiado y completado durante el periodo de vigencia del proyecto, el personal del SSC

estaba familiarizado con la estructura de datos de ELUS y por la cooperación del equipo de ELUS para hacer disponible en discos toda la información relevante en el momento apropiado.

7. ¿Que comprende el archivo?

En el caso de ELUS, un portafolios contiene tres discos protegidos contra escritura con archivos comprimidos y los programas necesarios para que éstos puedan ser automáticamente restaurados en un espacio de 15 Mb de un disco duro. Los archivos de datos están todos incluidos por duplicado, en dos formatos comunes, como archivos transportables SPSS y dBase IV, al menos uno de los cuales debe ser accesible a los usuarios por muchos años. Para los archivos de texto se usó Word 6, incluyendo la descripción completa de los esquemas de muestreo. Hay, por escrito, en diez páginas, una introducción y un informe resumido del muestreo.

Luego hay varios cientos de páginas que dan detalles, para cada cuestionario usado, cada archivo y cada variable. Para cada archivo la descripción incluye el número de casos, el número de variables, la lista completa de las variables declaradas, sus nombres y los de sus valores. Para cada variable, la descripción incluye su nombre y código, los valores mínimo y máximo y el número de casos almacenados sin incluir valores ausentes. El portafolio pesa 1.65 kilogramos.

Se prepararon treinta copias, cinco de ellas incluían información calificada como confidencial por razones comerciales o para proteger el anonimato de los informantes. Ambos tipos de copias fueron distribuidas apropiadamente, es decir, entre oficinas del Gobierno de Malawi, instituciones académicas y el Departamento para el Desarrollo Internacional (DFID). Usuarios legítimos e investigadores autorizados están en capacidad de encontrar una copia de los datos de forma tal que pueden, por ejemplo (a) realizar análisis adicionales, (b) con acceso apropiado, usar la información del archivo como un marco de muestreo extremadamente detallado, mediante el cual pueden revisar submuestras de las parcelas de ELUS; (c) integrar los datos de ELUS con sus propios hallazgos posteriores para análisis longitudinales.

8. Información Adicional

Esperamos que este documento provea algunas ideas iniciales sobre aspectos de importancia en el archivo de datos. El Statistical Services Centre está preparando guías más detalladas sobre procedimientos de archivo para investigadores potenciales que estén iniciando proyectos, para sus

evaluadores y quizás, para agencias de gobierno en países adonde se puedan realizar los proyectos. Estaríamos complacidos de conocer de investigadores con ejemplos o experiencias de trabajo similar a la descrita aquí, de usuarios, exitosos o frustrados, de archivos de datos, o de cualquiera que desee compartir ideas sobre aspectos pertinentes.

Reconocimientos

Savitri Abeyasekera y Carlos Barahona, miembros del personal del SSC involucrados en el caso de estudio descrito aquí, se complacen en reconocer su deuda con los miembros del equipo del proyecto ELUS en Malawi, 1995-1997: Joanne Bosworth (Socióloga Rural), Jasper Steele (Economista en Sistemas de Finca) y especialmente con Steve Gossage (Líder del Proyecto ELUS) por su generosa ayuda, estímulo y hospitalidad durante la creación de éste archivo de datos. Se agradece también a Harry Potter (Asesor en Recursos Naturales del DFID en Malawi) por su apoyo a este proyecto.

El Statistical Services Centre es una organización sin fines de lucro, adscrita al Departamento de Estadística Aplicada, en The University of Reading, Reino Unido. El Centro emplea a su propio personal y realiza trabajo de entrenamiento y consultoría para clientes fuera de la Universidad. Su personal asesora al DFID en aspectos biométricos en proyectos de recursos naturales con el objetivo de apoyar su diseño e implementación efectiva.

El DFID financia la producción de la versión en inglés de ésta serie, con el propósito fundamental de ofrecer guías para la investigación y apoyar al personal involucrado en proyectos del DFID.

La versión en castellano de las guías es producida con autorización del SSC y del DFID, por la Universidad Nacional y el Centro Agronómico Tropical de Investigación y Enseñanza, Costa Rica, con el propósito de distribuir las a investigadores y proyectos en países hispanohablantes.

Comentarios sobre las guías y sugerencias para nuevos temas serán bienvenidos.

Títulos Publicados:

- Guías Estadísticas para Proyectos de Recursos Naturales
- Investigación en Fincas - Algunas Guías Biométricas
- Guías para el Manejo de Datos de Proyectos Experimentales
- Guías para Planificar Encuestas Efectivas
- Archivo de Datos de Proyectos - Lecciones de un Estudio de Caso

Próximos Títulos:

- Guías para Presentación
- Diseño de Experimentos
- Métodos Modernos de Análisis de Datos
- El Análisis de Experimentos Combinados

Las guías son producidas tanto en forma escrita como en formato electrónico accesible por computadora. Para obtener copias de las guías o información adicional, favor contactar a la UNA, el CATIE o al SSC.

Para información adicional:
Statistical Services Centre, The University of Reading
P.O. Box 240, Reading, RG6 6FN United Kingdom

e.mail: statistics@reading.ac.uk

Unidad de Métodos Cuantitativos y Análisis de Datos. Escuela de Ciencias Agrarias,
Universidad Nacional, Heredia 3000, Costa Rica.

e.mail: jcamacho@una.ac.cr