



Manejo de datos de investigación

Ricardo Sánchez MD. Profesor Asociado, Facultad de Medicina, Universidad Nacional de Colombia. Fabián Jaimes MD. Profesor Auxiliar, Facultad de Medicina, Universidad de Antioquia.

SUMMARY

This article presents the general concepts and techniques related with projects that utilize computerized data, emphasizing in aspects of planning, implementing and maintaining activities required to adequately administrate research data.

RESUMEN

Este artículo presenta los conceptos generales y las técnicas que se relacionan con proyectos que utilizan datos en computadores. Se hace énfasis en aspectos de planeación, ejecución y actividades de mantenimiento requeridas para administrar adecuadamente los datos de investigación.

Los datos manejados en computador se han vuelto un elemento fundamental dentro de la investigación científica. Hace algunos años los datos se manejaban manualmente y eran excepcionales las bases de datos muy grandes. Ahora la disponibilidad de computadores impone el manejo de bases de datos voluminosas. En este escrito sólo se hará referencia al manejo de datos ya reunidos. No se profundizará sobre los procedimientos a seguir para recolectar los datos de una investigación, lo cual depende en gran medida de las características del diseño que se haya seleccionado (2).

Algunos términos utilizados en el manejo de datos de investigación son (2):

1. *Manejo de Datos de Investigación (MDI)*: Es la manipulación sistemática

de información almacenada en medios electrónicos, con el fin de proteger su contenido para investigaciones futuras o para responder preguntas de investigación. Este procedimiento permite mantener datos de buena calidad que permitan el desarrollo fácil y rápido de proyectos de investigación. Actualmente es una parte necesaria en cualquier proyecto.

2. *Sistema de MDI*: Es un conjunto sistematizado de actividades y procedimientos para adquirir, manipular, documentar y almacenar datos en un ambiente de computador. Un sistema de MDI está compuesto de procesos de MDI interrelacionados (Figura 1).
3. *Procesos de MDI*: Es un grupo de actividades directamente relacionadas con el logro de un objetivo de manejo de datos de

investigación. Por ejemplo: Recolección de datos, Entrada de datos, Limpieza de datos, Copias de respaldo, etc.)

4. *Monitorización de MDI*: Son las actividades y procedimientos para seguir y evaluar un proceso. Su objetivo es identificar oportunamente problemas para ejecutar soluciones oportunas.
5. *Perfil de recursos*: Identifica los recursos y restricciones ambientales disponibles (personal, computadores, dinero, tiempo.)

Un técnico de MDI debe estar en capacidad de diseñar un plan de MDI, poner en funcionamiento procesos de MDI, efectuar una monitorización efectiva del plan y los procesos, mantener funcionando los sistemas y corregir problemas inesperados.

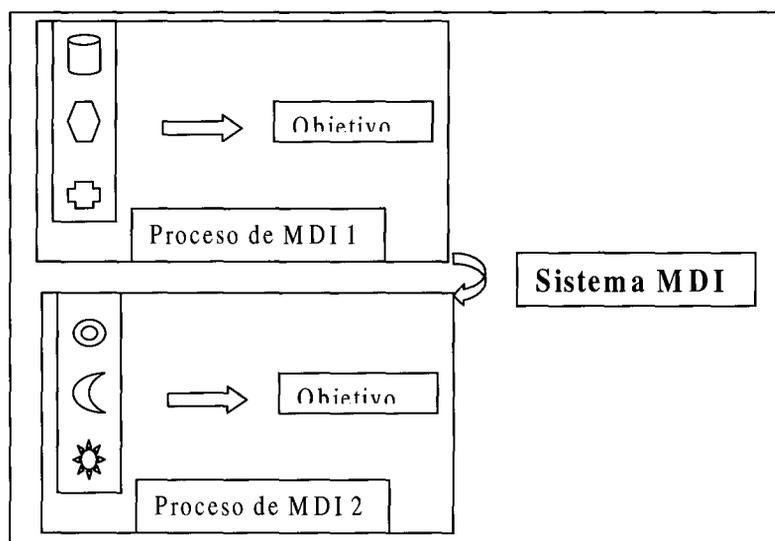


Figura 1. Estructura de un Sistema de MDI.

Recomendaciones generales para el MDI

1. Diseñar Procesos de MDI lo más simples que sea posible.
2. Hacer el diseño teniendo en cuenta los recursos disponibles.
3. Siempre estar buscando cómo mejorar el sistema de MDI.
4. En la planeación de un Proceso de MDI se han de considerar estos aspectos:
 - a. Adquisición de datos.
 - b. Verificación y validación de datos.
 - c. Manipulación y análisis de los datos.
 - d. Reporte de resultados.
 - e. Archivo y copias de respaldo.
 - f. Procedimientos de monitorización.
5. Tener un registro de los códigos y convenciones utilizados en el registro de los datos: Es necesario elaborar un instructivo que permita replicar el método de recolección y registro si se repite el proceso más tarde.
6. La protección de los datos es prioritaria. Se debe estar en guardia contra la infección y corrupción de los datos, la pérdida física o los accesos no autorizados.
7. Los procesos de MDI sin mecanismos de monitorización son una fuente de malestar.
8. Documentación: En todo proceso de MDI debe haber documentos escritos que registren:
 - a Elementos clave y discusiones importantes dentro de la planeación del proyecto.
 - b Descripción de procedimientos que deben efectuarse regularmente.
 - c Registro de decisiones administrativas importantes para el proyecto.
 - d Descripción de programas de computador diseñados para manejar los datos.
 - e Inventario detallado de los datos: Nombre del Archivo, tamaño, número de observaciones, número de variables, fuente de los datos, referencias cruzadas con otros documentos.
 - f Inventario de las limitaciones y errores, encontrados o posibles, dentro del proceso de MDI.

Sistemas de Transferencia de Datos (3)

Generalmente cuando se termina la recolección de los datos, estos se encuentran anotados en hojas de papel o en otro medio similar (tarjetas de cartón, acetatos, etc.). El paso siguiente consiste en transportar estos datos a un medio electrónico que permita su agrupación y manipulación. En este momento se presenta la primera transferencia. Generalmente esta transferencia se hace a una *hoja electrónica* (Excel, Quatro Pro) o a una *base de datos* (Access, Dbase). En este punto se debe disponer de un archivo que permite realizar determinadas manipulaciones de los datos. Para esta labor son preferibles las bases de datos ya que presentan las siguientes ventajas:

1. Permiten relacionar diferentes grupos de datos y unirlos con una variable común.
2. Facilitan la búsqueda de datos y las consultas teniendo en cuenta varios condicionantes.
3. Facilitan el ingreso de datos adicionales mediante el diseño de formularios amigables que además reducen el riesgo de errores de digitación.

Dadas las limitaciones de estos programas para el procesamiento estadístico, generalmente debe hacerse una segunda transferencia desde la base de datos a un programa estadístico (SAS, SPSS, STATA, MINITAB, etc).

Estructura de una base de datos (4)

Una base de datos es una estructura en la cual se almacenan, con un orden definido, un grupo de descripciones sobre determinados individuos.

Los individuos generalmente se denominan casos, observaciones o pacientes. Las descripciones que se hacen de los individuos corresponden a mediciones de algún tipo de fenómeno; dichas mediciones se denominan variables.

Es recomendable que esta estructura tenga un formato que facilite el manejo, la interpretación y el procesamiento

de los datos. Obviamente, dicha estructura dependerá de la naturaleza de los datos. En general, los datos de investigación clínica de tipo cuantitativo quedan adecuadamente organizados en una estructura que tenga las siguientes características (figura 2):

1. Presentar un arreglo ordenado de filas y columnas.
2. Las filas corresponden a los casos. Cada fila representa un conjunto de observaciones que se hacen sobre un mismo sujeto.
3. Las columnas representan las variables. Cada columna muestra los diferentes valores que toma una sola variable en cada uno de los diferentes individuos.
4. Las casillas contienen los valores. Una casilla es la intersección entre una fila y una columna. Indica el valor que tiene una variable en determinado individuo. Cuando la casilla no tiene ningún valor se habla de un dato faltante o perdido.
Para organizar una base de datos debe tenerse claridad en la definición de los casos y de las variables:

1. Casos: Deben ser sujetos que compartan uno o varios atributos, de tal manera que se puedan identificar como pertenecientes a una misma categoría. Así pues, los casos deben ser pacientes, muestras de sangre o de algún otro tejido, animales de experimentación, o radiografías, por mencionar solo algunos.

2. Variables: Son las características de los individuos estudiados y se debe unificar su nivel de medición para todos los casos. Resultaría inadecuado, por ejemplo, que la variable estatura estuviera medida en centímetros para algunos casos y en pulgadas para otros. Para cada variable debe definirse claramente:
 - a. Su nombre: Dependiendo del programa que se utilice para manejar los datos y para efectuar los procedimientos estadísticos, puede ser necesario utilizar nombres abreviados, en general no mayores de ocho caracteres. Cada nombre de variable

debe ser único (no se permiten duplicados) (5).

El hombre que se seleccione debe intentar reflejar lo mejor posible la característica del atributo que mide la variable.

- b. Su Tipo: de antemano debe definirse si se trata de una viable numérica, de texto o de fecha. Cuando se trate de variables categóricas debe establecer un código para denominar cada uno de los niveles de la variable.
- c. Las etiquetas: Dependiendo del programa utilizado para el procesamiento de los datos se puede asignar nombres más largos, que describen mejor cada una de las variables y categorías, y que aparecen en las salidas de los procesos para facilitar su lectura e interpretación.
- d. El código utilizado para identificar los valores faltantes: Muchos paquetes estadísticos y bases de datos asignadas automáticamente valores a los datos perdidos, generalmente un punto o el valor 999. Sin embargo el investigador puede asignar un código a este tipo de datos, cuidando que no coincida con el valor que puede tomar esa variable en algunas de las mediciones.

Cómo determinar la calidad de los datos

1. Integridad: Lo primero es verificar los datos para ver si están completos o tie-

nen defectos por errores en su procesamiento o por daños físicos luego de ser archivados.

2. Exactitud: Para verificar este punto puede usarse estrategias como medir un mismo fenómeno de diferentes maneras, comparar los datos incluidos en el archivo electrónico con los datos en el papel o usar información redundante (un mismo dato se registra varias veces, con diferente indentificado, y luego se mira si los registros coinciden).

3. Utilidad: Hace referencia al uso que se le puede dar a los datos. Hay que definir qué programas hay disponibles para manejar los datos, pues estos pueden haber sido introducidos en un formato incompatible. También es necesario determinar la capacitación del personal para entender y manejar los datos, y la disponibilidad de equipos para procesarlos

Preservar la seguridad de los datos

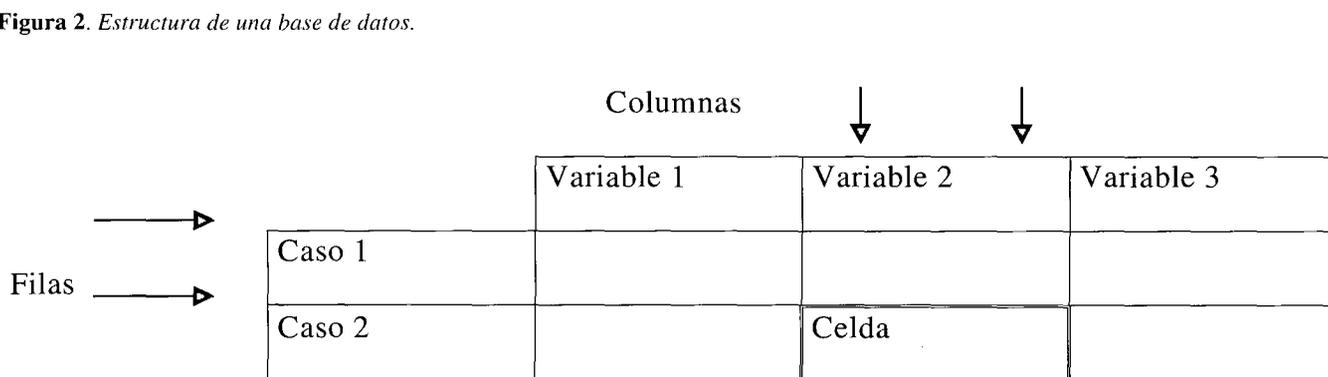
1. Hacer copias de seguridad: Si bien este es un concepto aplicado a los datos, puede extenderse a los equipos e incluso al personal. La idea básica es que no haya nada ni nadie que sea indispensable y que si falta se pueda reemplazar. En relación con los archivos electrónicos de datos se debe establecer un programa de copias de seguridad, de tal manera que existan al menos dos de estos respaldos, además del archivo original. En cuanto a los equipos, debe tenerse preparado un plan de contingencia, de manera que si un equi-

po falla, exista la posibilidad de trabajar en otra máquina que pueda manejar esos datos, asegurando que no se pierda tiempo. En relación con el personal se deben establecer claramente las responsabilidades, asignando a alguien del equipo o externo a él, funciones opcionales que deben cubrirse en un plan de emergencia. Se debe pensar que la mejor manera de lograr que algo ocurra es dejarlo al azar

2. Sistema de archivo: Los procesos de guardar y preservar datos deben realizarse regularmente. No es adecuado dejar el archivo de los datos para el final de una sesión de registro o de procesamiento. Cuando se deban manejar grupos de datos muy grandes puede ser recomendable crear participaciones que permitan trabajar con grupos de menor tamaño, lo cual aumenta la velocidad de procesamiento, facilita su manejo y disminuye la probabilidad de fallos en el sistema del computador. En estos casos es necesario tener una convención para nombrar los diferentes archivos, de tal manera que sea fácil identificarlos como parte de un archivo mayor e incluso conocer dentro del mismo. El sitio donde se van a archivar los datos debe contar con las medidas de protección necesarias para evitar daños físicos o corrupción de archivos electrónicos.

3. Restricción del acceso a los datos: Se debe asegurar que sólo el personal capacitado para manejar los datos ten-

Figura 2. Estructura de una base de datos.



ga acceso a los mismos. La falta de pericia para manipular algún programa de computación puede resultar desastrosa en una investigación. Debido a que uno de los requerimientos éticos tiene que ver con la confidencialidad de los datos, se recomienda restringir el acceso a los mismos mediante contraseñas o claves de entrada.

Diseño de un sistema de MDI

La primera etapa para el diseño de un sistema de Manejo de Datos de Investigación consiste en hacer un listado de los principales objetivos del proyecto, relacionados con la recolección y manejo de los datos. Como lo ilustra la figura 3, se pueden distinguir dos grupos de objetivos:

1. *Objetivos primarios:* Son los que corresponden a logros específicos y secuenciales dentro del proyecto de investigación. Los más frecuentes tienen que ver con la adquisición de los datos, con su validación, manipulación y análisis y con el reporte de resultados.

2. *Objetivos de seguimiento:* Se refieren a procesos que deben ejecutarse permanentemente, a lo largo de todo el proyecto. Fundamentalmente tienen que ver con la protección de la información, con la monitorización de los diferentes procesos, con el inventario de recursos y con la documentación.

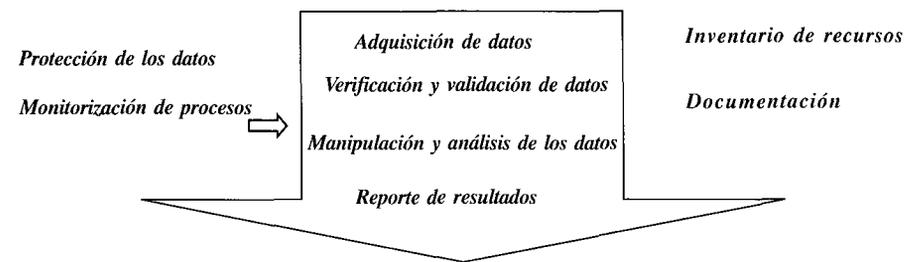
Cada uno de los anteriores objetivos implica la realización de una serie de actividades y procedimientos que, en su conjunto, configuran un Sistema de MDI. A continuación se mencionarán más detalladamente las características de cada uno de estos puntos:

Adquisición de Datos:

En general, se deben considerar aquí los siguientes procedimientos:

- Recolección de los datos escritos o disponibles en otros medios magnéticos.
- Verificación de la integridad de estos

Figura 3. Objetivos que se consideran al diseñar un Sistema de MDI.



Los Objetivos Primarios (dentro de la flecha) implican etapas, generalmente secuenciales, que se deben cumplir a lo largo del proceso de MDI. Los Objetivos de Seguimiento (en cursiva) se mantienen permanentemente a lo largo de todo el proceso, independientemente del Objetivo Primario que se esté desarrollando.

datos y nominación de los mismos mediante la asignación de códigos de identificación. Cada una de las observaciones debe tener un identificador único.

- Diseño de la base de datos. En este punto debe tenerse en cuenta la utilización de métodos que disminuyan la probabilidad de error durante la digitación, como usar listas de selección, sistemas de bloqueo de datos extraños o señales auditivas que avisen cuando hay errores en la digitación. Todas estas estrategias están fácilmente disponibles en una base de datos comercial.
- Capacitación del personal encargado de la digitación.
- Prueba piloto de la base de datos y ajuste de la misma.
- Ingreso de los datos a la base de datos (Primera transferencia).

2. Verificación y validación:

Además de la estructura de seguridad que se da a la base de datos para reducir el riesgo de ingreso de datos falsos o equivocados, deben ejecutarse pruebas de confrontación de los datos originales contra los datos equivalentes en el archivo electrónico. Otro sistema, ya mencionado, consiste en incluir varias veces un dato marcado, para posteriormente confrontar si todas las entradas registran la misma información. Las características de los errores encontrados pueden hacer necesario el reajuste

de la base de datos, la puesta en marcha de procesos de capacitación o el cambio de personal. Debe tenerse en cuenta que el éxito de un programa de MDI radica más de un 80% en las personas. Los equipos electrónicos y de cómputo no son tan importantes como se cree. Por esta razón es prioritario que las personas que manejen los datos tengan un adecuado entrenamiento.

3. Manipulación y análisis de datos:

Los procedimientos más frecuentes son:

- Generación de un sistema de convenciones que permita identificar los archivos correspondientes a la investigación.
- Identificar cada una de las variables junto con sus niveles de medición
- Definir los procedimientos utilizados en cada una de las transferencias. Cuando se vaya a realizar fragmentación de la base de datos original para manejar grupos de datos más pequeños, se debe iniciar el procedimiento conociendo perfectamente qué grupos de datos o de variables se incluirán en la nueva división. El paso de los datos hacia un programa estadístico generalmente corresponde a la segunda transferencia y se debe efectuar de tal manera que garantice la integridad de los datos, principalmente en lo referente a los nombres y etiquetas de las variables, así como a las ca-

- racterísticas de las mismas. Un error que frecuentemente se comete durante la segunda transferencia es convertir datos numéricos en datos de texto que no permiten una adecuada manipulación estadística.
- d. Seleccionar adecuadamente el paquete estadístico: Si se requieren múltiples procedimientos estadísticos procurar conseguir un programa que sea capaz de realizarlos todos. Entre menos programas estadísticos se utilicen y menos transferencias se realicen, menor será la posibilidad de errores en el proceso de transferencia de datos.
4. Reporte de Resultados:
Los procedimientos que se tienen en cuenta son:
- Selección de la publicación en la que se presentarán los resultados: Si bien existen normas estandarizadas de publicación de artículos científicos, pueden presentarse algunas especificidades que deben tenerse en cuenta cuando se envíe el material a consideración de los editores.
 - Evaluación de las características del auditorio y selección de medios audiovisuales. El reporte de los resultados de una investigación puede hacerse ante auditorios científicos, para lo cual

deben tenerse en cuenta las normas básicas de este tipo de presentaciones. La incorporación de los anteriores elementos dentro de un proyecto de investigación debería hacerse de manera sistemática, teniendo en cuenta que los datos recogidos son un elemento central dentro de cualquier estudio. La aparente complejidad del diseño de un sistema de manejo de datos de investigación no debe tomarse como pretexto para no invertir los recursos que sean necesarios con el fin de velar por su calidad, seguridad y adecuada disponibilidad. Este aspecto debe tenerse en cuenta al efectuar los cálculos de presupuesto de cualquier investigación.

REFERENCIAS BIBLIOGRÁFICAS

- Feigal D, Black D, Grady D, Hearst N, Fox C, Newman TB, Hulley SB.** Planificación de la gestión y el análisis de datos. En: Hulley SB, Cummings SR: Diseño de la Investigación Clínica. Un Enfoque Epidemiológico. Barcelona: Ediciones Doyma S.A;1993.p.175-186.
- Calvert WS, Meimei Ma J.** Concepts and Case Studies in Data Management. North Carolina :SAS Institute Inc; 1996.
- Pocock SJ. Clinical Trials.** A Practical Approach. Chichester: John Wiley and Sons; 1983.
- Lebart L, Morineau A, Piron M.** Tratamiento estadístico de datos. Métodos y programas. Barcelona: Marcombo; 1985.
- SPSS 9.0 Manual del usuario.** Chicago: SPSS Inc.;1999.
- Spilker B.** Guide to clinical trials. Philadelphia: Lippincott - Raven Publishers; 1996.p.473-478.
- Comité Internacional de Editores de Revistas Médicas. Requisitos uniformes para los manuscritos enviados a revistas biomédicas. Acta Médica Colombiana 1997; 22:199-211.

Nota: Tomado del libro "Estrategias de Investigación Médica Clínica "E Ardila , R Sánchez, J Echeverry Eds. (en prensa).