

Diseño e implementación de una aplicación informática para la gestión de un registro de pacientes diagnosticados de retraso mental de origen genético

A. Bellón-Vázquez, J. Delgado-Naranjo, C. Rodríguez-Andrés

THE DESIGN AND IMPLEMENTATION OF A COMPUTER APPLICATION FOR MANAGING
A REGISTER OF PATIENTS DIAGNOSED WITH GENETIC MENTAL RETARDATION

Summary. Introduction. In order to conduct studies of mental retardation of a genetic origin over the GIRMOGEN network, a computerised register of cases needs to be set up. Materials and methods. This register requires the implementation of a computer application consisting of a series of management programs together with the actual database. The aims of this work are: a) To analyse all the infrastructural, scientific and technical conditioning factors that allow the computer application and the databases associated to the register to be designed; b) To design all the elements that go to make up the application, namely, the application engine, consisting of all the programs responsible for providing all the operative capabilities, and the database for storing all the data structures with the variables related to the research; c) Implementing all the designs that are developed; and d) Drawing up the security protocols and work procedures involved in running the application. Results. The result of all this process is the Girmogen-Pro software application, which includes all the capabilities needed to enable it to fulfil the scientific goals of the study. It includes the security measures required to comply with the current laws regarding the handling of files containing personal information. Furthermore, the data structures that will constitute the registers to be used in the study were also been built. Conclusions. This process does not end with the implementation of the designs, since an ongoing assessment process is also performed with the aim of optimising the data structures that contain the variables that go to make up the register, so as to ensure it fits the scientific objectives of the study as well as possible. [REV NEUROL 2006; 42 (Supl 1): S125-30]

Key words. Databases. Designing computer applications. Genetic mental retardation. Registers. Research. Security.

INTRODUCCIÓN

El trabajo que se pretende desarrollar trata fundamentalmente de dar cobertura informática a un proyecto científico, consistente en la creación de un registro de pacientes diagnosticados de retraso mental de origen genético (RMOG).

La red GIRMOGEN (Grupo de Investigación en Retraso Mental de Origen Genético) es pionera en el estado español en la investigación del RMOG. Dicha investigación exige la creación de un registro o base de datos para la recogida de todas aquellas variables implicadas en las áreas de conocimiento involucradas. Este registro ha de cumplir los siguientes requisitos:

- Permitir el registro de casos en tiempo real.
- Que sea un sistema multicéntrico que posibilite el registro de casos en diferentes puntos geográficos.
- Permitir la recogida de todos los datos que se necesiten y, por tanto, disponer de la amplitud necesaria para cubrir los diferentes grupos de variables requeridos para un correcto análisis epidemiológico, con la consiguiente disminución de la probabilidad de sesgo. También debe posibilitar el control de la confusión y las interacciones entre variables.
- Posibilitar la recogida para su procesamiento y estudio de todos los casos en los que se tenga constancia de la existencia de un RMOG en cualquiera de los centros sanitarios y de investigación vinculados a la red GIRMOGEN.

Aceptado: 22.11.05.

Departamento de Medicina Preventiva y Salud Pública. Facultad de Medicina y Odontología. Universidad del País Vasco. Leioa, Vizcaya, España.

Correspondencia: Dr. Antonio Bellón Vázquez. Departamento de Medicina Preventiva y Salud Pública. Facultad de Medicina y Odontología. Universidad del País Vasco. Barrio Sarriena, s/n. E-48940 Leioa. Vizcaya. E-mail: jedelnar@terra.es

© 2006, REVISTA DE NEUROLOGÍA

MATERIALES Y MÉTODOS

En función de los supuestos de partida y condicionantes derivados de la naturaleza del registro, ya explicados en la introducción, se necesita, además del diseño de una base de datos, la adquisición o el desarrollo de un programa informático específico, dedicado a proporcionar los accesos a la base de datos y su gestión.

Entre ambas posibilidades, programa informático comercial o desarrollado, se optó por la segunda, debido a la necesidad de adaptarse al máximo a los requerimientos derivados de la investigación de GIRMOGEN, necesidad que no cubren las restricciones impuestas por un programa informático preconfigurado. El conjunto del programa más las bases de datos asociadas constituye una aplicación informática, que hemos denominado Girmogen-Pro.

Requisitos informáticos

De la interpretación, desde un punto de vista informático, de los requisitos enumerados en la introducción se derivan los siguientes aspectos para la aplicación que gestiona el registro de pacientes de la base de datos:

- Acceso a la base de datos a través de Internet.
- Estructuración fundamentada en el organigrama de grupos de la red GIRMOGEN.
- Inserción y consulta de datos en tiempo real.
- Inclusión de los diferentes listados de variables clínicas y de laboratorio por áreas de estudio.

Fases del desarrollo de la aplicación informática

Según el procedimiento indicado por Delgado et al [1], el proceso global de desarrollo e implementación de un registro informatizado, con inclusión de aquellos registros de pacientes diagnosticados de RMOG, ha de llevarse a cabo según las siguientes fases: 1. Planteamiento inicial; 2. Planificación; 3. Requerimientos; 4. Análisis y diseño; 5. Seguridad: cobertura legal y confidencialidad; 6. Implementación; 7. Tests y pruebas; 8. Evaluación; y 9. Producción e investigación.

Por consiguiente, las fases del desarrollo de Girmogen-Pro han de seguir este esquema, y la aplicación se somete a un continuo ciclo de análisis, diseño, pruebas y evaluación.

Análisis y diseño de la aplicación

De todas estas etapas, el presente trabajo se centra en el análisis y diseño de

Girmogen-Pro. Según Delgado et al [1], la citada fase se subdivide en dos subetapas: análisis y diseño funcional, y análisis y diseño orgánico. Es precisamente esta segunda fase la que se aborda en este trabajo. Con el objeto de facilitar una mejor comprensión de las actividades desarrolladas, esta fase se ha subdividido en tres partes: a) Medios e infraestructuras utilizadas; b) Desarrollo de los programas asociados al control de accesos, gestión de usuarios y control del sistema, que llamaremos 'motor'; y c) Construcción de la base de datos.

Análisis y diseño funcional

Esta fase del desarrollo de la aplicación informática es eminentemente científica. El personal informático encargado de llevarla a cabo debe atenerse a las consideraciones técnicas derivadas de las diferentes variables epidemiológicas, criterios de inclusión dentro del registro y adecuación a estudios científicos resultantes de la utilización de este registro de pacientes. Delgado et al [1] recogen las diferentes consideraciones relativas al análisis y diseño funcional de la aplicación.

Análisis y diseño orgánico

Este paso, que debe realizarse antes de comenzar la implementación, resulta de capital importancia, ya que la bondad del diseño redonda en una optimización de los recursos y tiempo de trabajo utilizados en el desarrollo de una aplicación y en su rendimiento [2]. Además, es sumamente importante disponer de un diseño cuya flexibilidad permita el continuo desarrollo del sistema, más allá de los supuestos de partida, y que puede aportar funcionalidades surgidas a lo largo de la duración del proyecto no previstas en su enunciado original.

En el desarrollo de este proyecto no se han podido utilizar las referencias de otros trabajos semejantes, ya que en el ámbito concreto del estudio del RMOG no se han encontrado referencias documentadas sobre realizaciones de este tipo. Por tanto, el diseño de la aplicación y su implementación se han realizado partiendo de cero.

Materiales e infraestructuras

Una vez definidos los requerimientos para la construcción de la base de datos y, por ende, de todo el conjunto de la aplicación, que desde el punto de vista de la programación pueden afectar a su diseño, se necesita conocer la infraestructura para la realización del trabajo. Los medios e infraestructuras disponibles son los siguientes:

- *Un servidor de bases de datos.* Este equipo, aportado por la propia red GIRMOGEN, se destina a albergar todos los registros de la base de datos. Tanto su configuración como gestión están a cargo del equipo informático de la red.
- *Un servidor web.* Gracias a la inestimable colaboración del CIDIR (Centro de Informática para la Docencia, Investigación y Red) de la Universidad del País Vasco, se puso a disposición de la red uno de los equipos de la universidad configurado como servidor web, donde alojar todo el programa informático desarrollado. La gestión de este servidor corresponde a la propia universidad, con el compromiso de adoptar o realizar todas las acciones necesarias para posibilitar el buen desarrollo de la investigación.
- *Terminales de trabajo y control.* Una serie de equipos, aportados indistintamente por la red GIRMOGEN y la Universidad del País Vasco, para la realización y desarrollo de los trabajos y la configuración y control de los servidores.
- *Un sistema de red de transmisión de datos de alta capacidad.* El propio sistema de red de la universidad, que permite cumplir las funciones de enlace entre los servidores dedicados al proyecto y los terminales de trabajo y control, además del acceso a Internet.

Del análisis de los medios e infraestructuras mencionados, y como resultado de las distintas conversaciones mantenidas con la dirección del CIDIR de la Universidad del País Vasco, se definieron las herramientas de trabajo a utilizar, tanto por su disponibilidad como por su perfecta adecuación a los requisitos exigidos en el desarrollo de la aplicación:

- *MySQL* [3], que es uno de los sistemas gestores de bases de datos relacionales sobre SQL (lenguaje de consulta estructurado de acceso a bases de datos) más populares. Sus principales características pueden resumirse en su disponibilidad para un gran número de plataformas, diferentes opciones de almacenamiento según se desee velocidad en las operaciones o mayor cantidad de operaciones disponibles, replicación, conexión segura, así como sus altas prestaciones para niveles intermedios de transacciones

y almacenaje de datos [4,5]. Además, al desarrollarse bajo la filosofía de código abierto, permite su utilización gratuita para fines no comerciales.

- *Perl* (*practical extraction and report language*) [6], un lenguaje de programación basado en *scripts* –códigos que no necesitan ser preprocesados por un compilador para interpretarse– perfectamente optimizado para su uso en el desarrollo de aplicaciones para acceso a bases de datos a través de Internet [7]. La codificación basada en *scripts* presenta la ventaja de que no es necesaria la utilización de complejas herramientas específicas de programación y compilación de códigos; basta con disponer de sencillos editores de texto, que es importante si se tiene en cuenta que la mayor parte del trabajo de codificación se ha desarrollado de forma remota en distintos equipos.

Ambas herramientas operan sobre Linux, un sistema operativo de tipo UNIX de altas prestaciones tanto en seguridad como en rendimiento en plataformas multiusuario.

Diseño del 'motor' de la aplicación

Como ya se ha mencionado anteriormente, la aplicación como tal consta de dos partes: los programas de gestión y control del sistema, y la base de datos. Por tanto, su diseño se divide en dos etapas: una dedicada al diseño de todos los programas que van a aportar las funcionalidades de la aplicación, que denominaremos 'motor', y la otra dedicada más específicamente al diseño de la estructura de la base de datos en sí y todo lo relacionado con las variables contenidas en ella.

El punto de partida para comenzar el diseño del motor ha sido la propia estructura de la red GIRMOGEN, vertebrada en torno a distintos grupos de investigación, cada uno de ellos especializado en una o más áreas específicas del estudio del RMOG. La jerarquía interna de la red también se trasladó al diseño del motor mediante la creación de las siguientes categorías de usuarios: invitado, grabador, para los encargados de introducir datos; verificador, para los encargados de controlar los registros de los grabadores de sus respectivos grupos; supervisor, para los encargados de controlar los registros de todos los grupos; y administrador del sistema. Con la idea de implementar el esquema jerarquizado basado en esa organización se ideó una estructura paralela, en la que se integró toda la infraestructura disponible (Fig. 1). Además, el diseño debía permitir un proceso de control de calidad de los datos, consistente en hacer que los registros introducidos se revisaran sucesivamente por los usuarios de nivel superior, antes de considerarse aptos para su utilización en la investigación (Fig. 2). Teniendo en cuenta todas estas consideraciones, se realizó el diseño del motor de la aplicación, cuyas principales características debían ser:

- Habilitar los sistemas de acceso, tanto para la consulta como para la inserción de datos de la futura base de datos.
- Actuar de control de acceso, e identificar en cada momento la identidad del usuario, su nivel de acceso y gestionar las funcionalidades disponibles para el usuario según ese nivel de acceso. Para la realización de estas funciones se utilizaría una pequeña base de datos auxiliar con los datos de los usuarios.
- Habilitar un sistema de control de calidad de los registros.
- Proporcionar los sistemas de control de los registros existentes.
- Disponer de un sistema de gestión de altas y bajas de usuarios.
- Proporcionar un sistema de control y gestión sobre el acceso a la base de datos.
- Proporcionar un sistema de gestión, control y mantenimiento de toda la aplicación.
- Permitir la implementación de las medidas de seguridad necesarias para adecuarse a las normativas derivadas de la Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal, desarrollada por el Real Decreto 94/1999, de 11 de junio.

Además, este diseño debía permitir, a la hora de realizar su implementación y codificación, la posibilidad de añadir nuevas capacidades que se derivaran de las necesidades surgidas a lo largo del desarrollo de la investigación.

Diseño de la base de datos de la aplicación

El diseño de la base de datos debía estructurarse desde un principio a partir de las listas de variables que formaban el estudio científico. Todas las variables escogidas para su inclusión en la base de datos debían superar un proceso de selección y recopilación fundamentado sobre los criterios científicos.

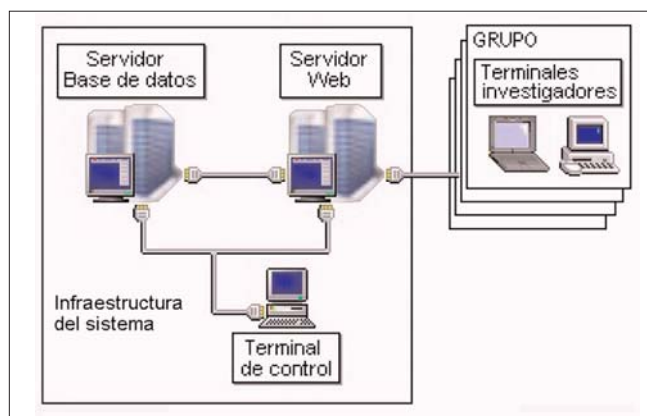


Figura 1. Esquema de la distribución de la infraestructura del sistema.

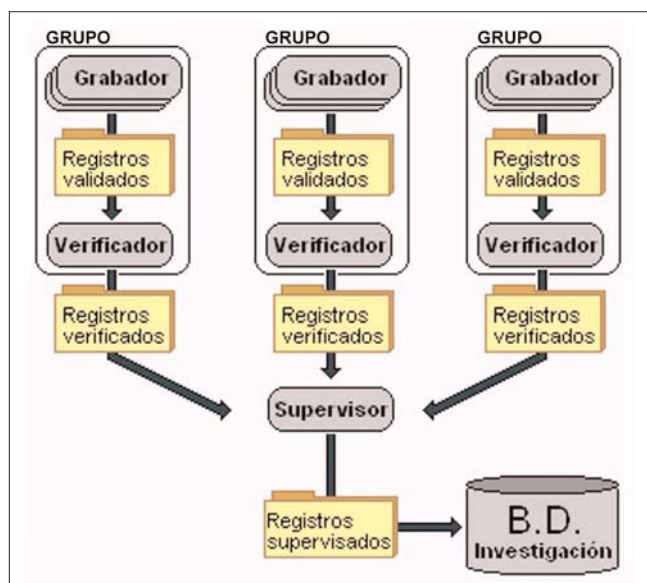


Figura 2. Esquema del proceso de aceptación de los registros.

cos derivados de la investigación. Este proceso de elaboración de los listados de variables tuvo lugar durante la etapa de diseño de la base de datos; debido a ello, no se disponía de las relaciones de variables, ni del conjunto total, ni por áreas de estudio; tan sólo una proyección de su magnitud numérica y, evidentemente, del número de áreas implicadas.

Considerando este punto, resultó imposible realizar un diseño totalmente estructurado compartimentado por áreas de estudio o por categorías de variables, o mediante una estructura jerárquica cerrada. Por ello, se decidió idear una estructura basada en módulos independientes, articulados sobre registros base. Las principales características de este diseño eran las siguientes:

- El registro base o raíz se orientaba a servir de índice de todas las variables correspondientes a los módulos asociados a dicho registro. La composición de este registro base determinaba el número máximo de módulos que podían incluirse en la estructura de la base de datos. Ésta iba a ser la única organización de datos dentro de la base de datos con una estructura fija y cerrada.
- Cada módulo tenía que ser independiente de los demás, intentando agruparse en él un número de variables indeterminado, relacionadas por una temática común, pero siempre pertenecientes a una única área del estudio científico. Su organización, en cada caso, iba a depender de las variables incluidas sin limitar su cantidad o tipo, y se determinaría a partir del análisis de los listados de variables.
- Todos los módulos, con independencia del tipo o cantidad de variables incluidas o del área de la investigación a los que se asociaran, debían estructurarse al mismo nivel en la base de datos, por debajo del registro

base, con inclusión de los posibles submódulos que se necesitasen habilitar dentro de cada módulo.

Un inconveniente de este diseño es que no permite realizar un proceso completo de normalización de toda la base de datos [8] y, por ello, no se puede conseguir una eliminación total de redundancias en la base de datos, con la consiguiente merma en su rendimiento. Sin embargo, el hecho de tener distintos módulos totalmente independientes representaba una gran ventaja, pues permitía tanto su desarrollo como su funcionamiento operativo, y una mayor libertad en la construcción de las distintas estructuras de datos asociados a las variables de cada área específica del estudio científico.

Otro aspecto que deb tenerse en cuenta es que se trata de una base de datos actualizable en tiempo real, lo que implica que en pleno proceso de investigación se van a realizar continuos cambios en los valores de las variables del estudio científico, con las consiguientes interferencias en los trabajos de investigación. Para solucionar este problema se decidió disponer de dos bases de datos gemelas en paralelo, una de producción, donde crear, modificar y supervisar los registros introducidos, y otra de investigación, donde alojar los registros considerados definitivos, con la finalidad de usarse en los estudios.

RESULTADOS

Como resultado de la implementación, construcción y puesta en funcionamiento de todos los diseños y desarrollos comentados en los puntos anteriores, se ha obtenido la aplicación Girmogen-Pro en su versión 1.0, cuya materialización se desarrolla mediante los siguientes puntos:

- Configuración y puesta en marcha de los medios e infraestructuras sobre los que funciona la aplicación.
- Codificación del motor de la aplicación.
- Construcción de la estructura de datos de la base de datos de la aplicación.
- Desarrollo de los protocolos de seguridad y procedimientos de trabajo en el uso de la aplicación.

Configuración y puesta en marcha de los medios e infraestructura

Los dos servidores, entre los medios e infraestructura que conforman el soporte tecnológico sobre el que se ejecuta la aplicación, son los equipos que han requerido una actuación directa para adecuarse a los objetivos de la investigación. Ambos se configuran con una dedicación específica:

- El servidor web se encarga de proporcionar la interconexión de los usuarios con el servidor de bases de datos y el alojamiento para todos los programas de la aplicación. Este servidor es el que proporciona las medidas de seguridad más importantes de comunicaciones telemáticas. Entre ellas: a) Sistema de conexión a la aplicación, con control de acceso a la red y cifrado de claves; b) Un canal seguro de comunicación entre el servidor y el usuario, mediante encriptación por protocolo https (*hypertext transfer protocol secure*); y c) Sistema de control (sistema de logs) de conexiones.
- El servidor de bases de datos se ha dedicado a almacenar todas las estructuras de datos que va a usar la aplicación. Además, se han implementado medidas de control y seguridad adicionales, como: a) Canal seguro de comunicación en la transferencia de datos con el servidor web; y b) Sistema de control (sistemas de logs) de conexiones. La configuración de este equipo se prepara para limitar el acceso al equipo únicamente a los terminales de control, y se aísla del resto de equipos conectados a la red de datos.
- El resto de equipos, como los terminales de control y técnicos, se han integrado en la estructura ya explicada anteriormente (Fig. 1), con la utilización de la red de datos de la Universidad del País Vasco.

Codificación del motor

Como ya se ha mencionado, el motor de la aplicación proporciona todas las herramientas necesarias para su funcionamiento. Todos estos programas se han implementado mediante *scripts* codificados en Perl. Cada uno de estos *scripts* tiene una utilidad específica muy determinada; por ello, cada funcionalidad de la aplicación es la suma de uno o más de estos *scripts*. Debido a esta especialización, cada grupo de *scripts* funciona independientemente del resto, lo que ha permitido un continuo proceso de depuración de las funciones ya implementadas y la adición de otras nuevas. Todos los programas desarrollados proporcionan los siguientes servicios:

Programas de gestión de accesos

- *Control de acceso*: limita el acceso para usuarios autorizados; proporcio-

na sistemas de control horario de entrada y salida y medidas de seguridad adicionales ante accesos no autorizados.

- *Sistema de identificación*: permite la identificación y autenticación del usuario que accede a la aplicación y le asigna los servicios disponibles según su nivel de acceso: invitado, grabador, verificador, supervisor o administrador.

Programas de utilidades para usuario

- Panel de controles, que proporciona el acceso a todos los servicios disponibles según el nivel de acceso del usuario.
- Herramientas para la actualización de datos del usuario y visualización de usuarios de la aplicación.
- Servicio interno de mensajería.
- Servicio de consulta de manuales de uso de la aplicación y ayudas.
- Sistema de acceso al área de trabajo con los registros de variables.

Programas de utilidades para administrador

- Herramientas de gestión de usuarios.
- Guías de localización de errores y ayudas de soporte.

Programas de control del sistema

- Sistema de monitorización, estadísticas y control de accesos a la aplicación.
- Sistema de monitorización, estadísticas y control de la actividad con los registros de datos.
- Sistema de monitorización y control de acceso de la base de datos.

Otro aspecto destacado de la implementación del motor es la elaboración del entorno visual sobre el que actúa el usuario, es decir, la presentación en pantalla. Dado que cada plataforma (sistema operativo) puede presentar en pantalla un aspecto distinto de una misma realización, se ha elegido una codificación bajo los estándares HTML4 (*hypertext markup language*). De esta forma obtenemos una visualización en pantalla por parte del usuario independiente del sistema operativo que utilice, que no requiere la instalación de ningún programa informático adicional y que sólo necesita un navegador web con una antigüedad menor de tres años. El lenguaje HTML puede codificarse incrustado en los *scripts* de Perl; de esta forma, los propios *scripts*, además de gestionar las funciones para los que se han creado, también gestionan la visualización en pantalla. Todos los *scripts* que necesitan interactuar con el usuario se asocian a una serie de formularios CGI (*common gateway interface*) –que es la especificación estándar para la interconexión de aplicaciones externas con servidores web– codificados sobre HTML, que recogen y procesan los datos introducidos por el usuario [7,9,10]. Por tanto, todos los programas que forman el motor de la aplicación se encargan de proporcionar al usuario su entorno visual, además de los servicios y utilidades diseñadas para llevar a cabo la explotación del registro.

Paralelamente a la codificación del motor, se ha elaborado una detallada documentación de todo el proceso, con objeto de servir de base para la edición de los manuales de uso destinados a administradores y usuarios de la aplicación, y la realización de nuevas funcionalidades o desarrollos posteriores.

Construcción de la base de datos

Tras desarrollar el motor de la aplicación, se inició el proceso de construcción de las estructuras de datos. Esta etapa, la más importante desde el punto de vista científico, consiste en integrar los distintos listados de variables en una única estructura informática. En primer lugar, se analizaron los listados de variables en su conjunto para determinar las compartimentaciones que debían realizarse. La primera partición consistió en dividir las variables en tres categorías: variables generales del registro, variables independientes del tiempo y variables dependientes del tiempo. Las variables generales del registro agrupan todas aquellas que sirven de referencia espaciotemporal y de identificación del registro. Las variables independientes del tiempo, a su vez, se dividieron en cinco subgrupos: datos del paciente, del padre, de la madre, del embarazo y del parto. Las variables dependientes del tiempo también se dividieron en cinco grupos: área clínica, pruebas diagnósticas, exploración neurológica, fenotipos cognitivo y conductual, y laboratorio genético.

A continuación, se realizó un análisis detallado de cada uno de los listados para obtener toda la información de cada variable. En este estudio se determinaban todos los factores necesarios para su codificación informática:

- *Denominación*. Todas las variables se etiquetaron de forma clara e inequívoca para facilitar su identificación y reconocimiento.

– *Posibles intervalos o valores de la variable*. Éste es uno de los trabajos más complejos y comprometidos del análisis de las variables, ya que, al llevar a cabo un estudio mediante el uso de herramientas estadísticas o de análisis de variables, la aparición de campos con valores no definidos o indeterminados suponen un problema, por el hecho de necesitar su interpretación semántica. Se intentó, en la medida de lo posible, limitar estos valores a selecciones de valores prefijados o parámetros cuantificables.

– *Posible interacción con otras variables según el intervalo o valor de cada variable*. En muchas ocasiones hay grupos de variables que se asocian y sus valores afectan al resto de variables del grupo. Todas estas interacciones tenían que definirse claramente para evitar la obtención de resultados contradictorios, evitar valores con información carente de utilidad, o servir de variables raíz, para la activación o desactivación de otros conjuntos de variables.

De todas las variables inicialmente analizadas, se excluyeron aquellas que no podían codificarse adecuadamente y no resultaban imprescindibles para el estudio. Las restantes se han utilizado para construir las estructuras de datos de la base de datos gestionada por la aplicación. Dentro de cada subgrupo, las variables se distribuyeron en distintos módulos según su número y por temáticas específicas. Cada módulo se construyó con el empleo de un número de subestructuras dentro de la base de datos –tablas– basado en el análisis cualitativo de todas las variables del módulo, su magnitud numérica y la distribución operativamente más óptima (Tabla).

La actual estructura de cada módulo es el resultado de un continuo y largo proceso de inclusión y actualización en la fase de diseño y desarrollo de la aplicación informática Girmogen-Pro.

Además de construir la estructura de datos de todas las áreas y realizar la codificación de todas las variables, se ha necesitado implementar su presentación en pantalla y la recogida y grabación de los valores que puedan tomar. Por ello, para cada módulo se codificaron una serie de *scripts* encargados de realizar la visualización en pantalla, recogida de valores y su grabación en la base de datos. Es en esta parte donde se llevó a cabo la codificación de todos los filtros destinados a restringir la recogida de los valores de cada variable a los intervalos o parámetros establecidos al realizar su análisis.

Protocolos de seguridad y procedimientos de trabajo en el uso de la aplicación

El trabajo con datos de carácter personal está sometido a la legislación vigente, tal y como indican Delgado et al [1]. Esta normativa afecta a todas aquellas personas con acceso a este tipo de datos, que en el caso del registro que se pretende implementar corresponden al nivel más alto, y se ha necesitado la elaboración de una serie de protocolos de seguridad:

- Protocolo de inscripción como usuario de la aplicación y acceso a la base de datos.
- Sistema de asignación de niveles de acceso y claves.
- Registro de incidencias.
- Protocolo de petición de datos para la realización de investigaciones.
- Elaboración y difusión de las normas y reglamentos de seguridad.

Todos ellos son de obligado cumplimiento para todos los usuarios con acceso a la aplicación. Además, el propio trabajo con Girmogen-PRO está sometido a una serie de procedimientos, encaminados, por un lado, a complementar los protocolos de seguridad y, por otro, a cumplir las consideraciones técnicas derivadas de la investigación científica. Todo ello ha significado la elaboración de una serie de reglamentos sobre procedimientos de trabajo en el uso de la aplicación, entre ellos: a) Conexiones y desconexiones de la aplicación; b) Accesos a los datos; c) Creación y consulta de datos; d) Validación, verificación y supervisión de los registros de datos; e) Utilización de los registros de datos para empleo en la investigación; y f) Gestión, mantenimiento y control de los procesos de la aplicación. Estos reglamentos han de ponerse en conocimiento de todos los usuarios de la aplicación.

DISCUSIÓN

En el proceso de desarrollo e implementación de la aplicación Girmogen-Pro, la construcción de las estructuras de datos asociadas a la base de datos es la que se relaciona más directamente

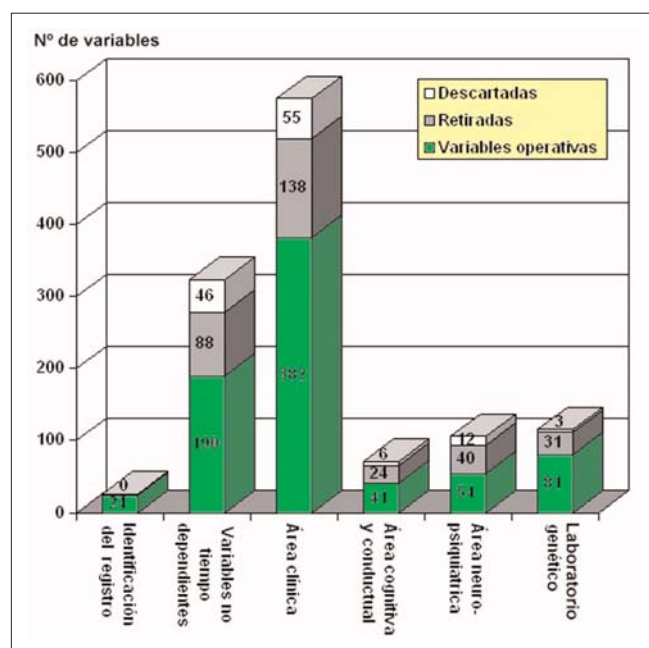


Figura 3. Relación de variables operativas, retiradas y descartadas.

Tabla. Organización de las variables de los registros pertenecientes a la base de datos de la aplicación.

	N.º de variables	N.º de módulos	N.º de tablas en la base de datos
Datos generales	24	1	1
Datos del paciente	48	2	4
Datos del padre	40	2	5
Datos de la madre	58	2	6
Datos del embarazo	34	1	3
Datos del parto	14	1	1
Área clínica	382	6	15
Pruebas diagnósticas	31	1	2
Exploración neurológica	54	1	2
Fenotipos cognitivo y conductual	41	1	1
Laboratorio genético	81	1	4

te con la investigación científica. Este proceso de construcción sigue abierto, ya que son muchas las consideraciones que obligan a realizar continuos cambios y actualizaciones en las variables de los registros, entre ellas:

- El número de variables recopiladas para su inclusión en la investigación superó ampliamente las proyecciones iniciales, de 500 a más de 1.000. Incluir un número de variables tan elevado representa un obstáculo para la realización de su codificación, la construcción de la estructura de los registros y su operatividad; aunque las menos relevantes, desde el punto de vista científico, se retiraron antes de incluirse en la base de datos, y su número era demasiado elevado todavía. Por ello, se han realizado recortes en las variables, tanto cien-

tífica como informáticamente, en todas las áreas de investigación (Fig. 3). Desde un punto de vista científico, los recortes se han realizado retirando aquellas variables consideradas no indispensables o redundantes. Los recortes efectuados en la implementación se han realizado respetando el carácter científico de las variables. Se han llevado a cabo mediante una agrupación por codificación, consistente en la creación de subestructuras de datos más complejas que concentran grupos de variables en un número menor de nuevas variables, dentro de cada módulo. Sin embargo, este aumento en el número de estructuras dentro de la base de datos implica un notable incremento de las llamadas variables técnicas, encargadas de gestionar y enlazar todas las estructuras de datos con sus respectivos módulos, y que requieren una gran cantidad de código para su implementación. Además, un aumento en el número de estructuras de datos conlleva un aumento en la complejidad de los sistemas de gestión y control de la base de datos. Dicha reducción, por otro lado, únicamente se aplica a un pequeño porcentaje de las variables presentes. Por tanto, existe todavía un proceso de reducción de variables con el fin de conseguir una estructura lo menos pesada posible.

- Desde el punto de vista informático, la interpretación de los valores o intervalos de cada variable es más sencilla cuanto más deterministas y restringidos sean éstos. Muchas variables se delimitan a valores de tipo ‘sí’/‘no’, intervalos prefijados, selección de opciones, etc., y no presentan ningún problema de interpretación; pero otras muchas no. Para conseguir que los valores de estas últimas variables se analicen o no dependan de interpretaciones semánticas, es necesario buscar formas de codificación que permitan acotarlos. Por otro lado, para conseguir esta acotación, hay que programar una serie de complejos filtros [8] en los CGI de recogida de datos. Estos filtros limitan la entrada de datos, y obligan a que éstos permanezcan dentro de unos intervalos limitados o tengan valores determinados, cualitativamente hablando. El problema de la implementación de estos filtros reside en la cantidad de código adicional necesario, y fundamentalmente en la necesidad de utilizar complejos algoritmos de filtrado, cuya correcta ejecución no es siempre posible en todas las diferentes plataformas existentes (navegadores web). En otros casos, ni siquiera se aplican por la complejidad que requiere el algoritmo del filtro. Por tanto, es necesario seguir buscando mejores codificaciones, que permitan una interpretación de los valores de cada variable lo más ajustada posible a los objetivos científicos y no resulte imposible de implementar.
- El conjunto de variables que forman el registro no está cerrado. A medida que avanza la tecnología y aparecen nuevos estudios, hay algunas que quedan desfasadas y tienen que sustituirse, o es necesario realizar la inclusión de otras nuevas para actualizar los datos de la investigación que se va a realizar. Todos estos cambios implican, a su vez, modificaciones en las estructuras de datos y sus contenidos. Al actualizar unas variables, hay que estudiar el efecto que estos cambios puedan tener en otras y, al añadir nuevas, debe tenerse en cuenta cómo incluirlas en los registros ya existentes sin mermar la de la aplicación, basada en la homogeneidad de todos los registros.
- Otro condicionante que debe tenerse en cuenta es la imposibilidad, derivada de la normativa de la legislación vigente, de

incluir en la misma estructura de datos una identificación inequívoca de un paciente junto con otros datos de carácter médico, sociológico, etc. Por ello, se deben disociar estos datos, mediante un sistema de codificación en la identificación de cada caso. Este sistema, basado en datos comunes, como nombre, apellidos, fecha y lugar de nacimiento, etc., tiene el inconveniente de imponer una serie de variables de obligada cumplimentación para la inclusión de un registro en la base de datos, ya que si alguno de éstos no existe, entonces no es posible obtener el código identificativo. Cuando el número de variables utilizadas en esta identificación es más amplio, tenemos un sistema de identificación muy fiable,

pero aumenta la posibilidad de tener registros excluidos del estudio, por no disponer de suficientes datos para su identificación. Por ello, es necesario buscar la mejor selección de campos para obtener el código identificativo, con el fin de conseguir un equilibrio entre la necesidad de incluir el máximo número de registros en la base de datos y la identificación inequívoca de los pacientes por parte del sistema.

La evaluación continua de todos los cambios efectuados, encaminados a ajustarse a los objetivos científicos del estudio, será lo que finalmente determine el conjunto y la codificación de todas las variables incluidas en la base de datos.

BIBLIOGRAFÍA

1. Delgado-Naranjo J, Rodríguez-Andrés C, Bellón-Vázquez A. Fundamentos teóricos para la creación de un registro de pacientes diagnosticados de retraso mental de origen genético: perspectiva desde la epidemiología y la salud pública. *Rev Neurol* 2006; 42 (Supl 1): S117-23.
2. Lucas-Gómez A. Diseño y gestión de sistemas de bases de datos. Madrid: Paraninfo; 1995.
3. MySQL AB. URL: <http://www.mysql.com>. Fecha última consulta: 02.11.2005.
4. Pérez-López C. MySQL para Windows y Linux. Madrid: Ra-Ma; 2003.
5. Maslakowski M, Butcher T. Teach yourself MySQL in 21 days. Indianapolis, IN: SAMS Publishing; 2000.
6. Perl. URL: <http://www.perl.org>. Fecha última consulta: 02.11.2005.
7. Castro E. Guía de aprendizaje Perl y CGI. Madrid: Prentice Hall; 2001.
8. Castro-Fernández MA. Normalización de bases de datos (I). *Sólo Prog Mundo-Linux* 2003; 52: 54-9.
9. Cohen I. CGI/Perl y JavaScript. Barcelona: Gestión; 2000.
10. Wyke RA, Thomas DB. Fundamentos de programación en Perl. Bogotá: McGraw-Hill; 2002.

DISEÑO E IMPLEMENTACIÓN DE UNA APLICACIÓN INFORMÁTICA PARA LA GESTIÓN DE UN REGISTRO DE PACIENTES DIAGNOSTICADOS DE RETRASO MENTAL DE ORIGEN GENÉTICO

Resumen. Introducción. Para la realización del estudio en retraso mental de origen genético por la red GIRMOGEN se necesita la creación de un registro informatizado de casos. Materiales y métodos. Este registro requiere la implementación de una aplicación informática, consistente en una serie de programas de gestión y la propia base de datos. Los objetivos de este trabajo son: a) El análisis de todos los condicionantes infraestructurales, científicos y técnicos que permiten llevar a cabo el diseño de la aplicación informática y las bases de datos asociadas al registro; b) El diseño de todos los elementos de la aplicación: el motor de la aplicación, consistente en todos los programas encargados de proporcionar todas sus funcionalidades operativas, y la base de datos donde alojar todas las estructuras de datos con las variables relacionadas con la investigación; c) La implementación de todos los diseños realizados; y d) La elaboración de los protocolos de seguridad y procedimientos de trabajo en el uso de la aplicación. Resultados. Como resultado de todo este proceso, se ha obtenido la aplicación Girmogen-Pro, en la que se han implementado todas las funcionalidades necesarias para su adecuación a los propósitos científicos del estudio. Se han incluido las medidas de seguridad derivadas del cumplimiento de la normativa legal vigente sobre tratamiento de ficheros de datos de carácter personal, y se han construido las estructuras de datos que compondrán los registros a utilizar en el estudio. Conclusión. Este proceso no termina con la implementación de los diseños, ya que existe un proceso continuo de evaluación encaminado a optimizar las estructuras de datos que contienen las variables que componen el registro, con objeto de adaptarse lo mejor posible a los objetivos científicos del estudio. [REV NEUROL 2006; 42 (Supl 1): S125-30] **Palabras clave.** Bases de datos. Diseño de aplicaciones informáticas. Investigación. Registros. Retraso mental de origen genético. Seguridad.

DESIGN E IMPLEMENTAÇÃO DE UMA APLICAÇÃO INFORMÁTICA PARA A GESTÃO DE UM REGISTO DE DOENTES DIAGNOSTICADOS COM ATRASO MENTAL DE ORIGEM GENÉTICA

Resumo. Introdução. Para a realização do estudo sobre atraso mental de origem genética pela rede GIRMOGEN é necessário o design de um registo informatizado de casos. Materiais e métodos. Este registo requer a implementação de uma aplicação informática, consistente numa série de programas de gestão e a própria base de dados. Os objectivos deste trabalho são: a) A análise de todos os condicionantes infraestruturais, científicos e técnicos que permitem levar a cabo a criação da aplicação informática e as bases de dados associadas ao registo; b) O design de todos os elementos da aplicação: o motor da aplicação, consistente em todos os programas encarregues de proporcionar todas as suas funcionalidades operativas, e a base de dados onde alojar todas as estruturas de dados com as variáveis relacionadas com a investigação; c) A implementação de todos os designs realizados; e d) A elaboração dos protocolos de segurança e procedimentos de trabalho no uso da aplicação. Resultados. Como resultado de todo este processo, obteve-se a aplicação Girmogen-Pro, na qual foram implementados todas as funcionalidades necessárias para a sua adequação aos propósitos científicos do estudo. Foram incluídas as medidas de segurança derivadas do cumprimento da normativa legal vigente sobre tratamento de ficheiros de dados de carácter pessoal, e foram construídas as estruturas de dados que comporão os registos a utilizar no estudo. Conclusão. Este processo não termina com a implementação dos designs, já que existe um processo contínuo de avaliação encaminado para otimizar as estruturas de dados que contêm as variáveis que compõem o registo, com o propósito de adaptar-se o melhor possível aos objectivos científicos do estudo. [REV NEUROL 2006; 42 (Supl 1): S125-30] **Palavras chave.** Atraso mental de origem genética. Bases de dados. Design de aplicações informáticas. Investigação. Registos. Segurança.