

Bases de datos para la investigación sanitaria: ¿necesidad de un enfoque multidisciplinario?

Javier Macías del Campo^a, José María Gutiérrez Martínez^a y Jesús Rodríguez Álvarez^b

^aIngenieros informáticos. Profesores de la Universidad de Alcalá de Henares. Madrid.

^bMédico. Especialista en Medicina Familiar y Comunitaria. EAP Balconcillo. Guadalajara.

Muchas personas de todos los ámbitos se sienten atraídas por la informática en alguna de sus múltiples facetas: juegos, ofimática, Internet, diseño asistido por ordenador, etc. Incluso parte de ellas son capaces de programar y desarrollar aplicaciones de pequeño tamaño, actividad que en numerosos casos aprendieron de forma mayormente autodidacta. Uno de los principales atractivos del desarrollo de aplicaciones a pequeña escala es su gran componente creativo, con el valor añadido de permitir ver los resultados de lo que se va creando de una forma casi inmediata.

Por otro lado, desde hace años existen en el mundo de los ordenadores personales y disponibles para el gran público sistemas gestores de bases de datos, como por ejemplo las distintas versiones de dBase®, Clipper®, FoxBASE®, Paradox®, Lotus Approach® y Microsoft Access®. Estos gestores han ido incluyendo, con el paso del tiempo, asistentes cada vez más sencillos y potentes para el diseño de tablas, consultas, formularios e informes, además de aumentar su capacidad con respecto a la cantidad y tipo de datos almacenables, rapidez, seguridad, etc.

Todo esto hace que parezca que, de todas las aplicaciones informáticas de pequeño tamaño, las más fáciles de desarrollar sean las constituidas fundamentalmente por una base de datos. Cualquier persona con conocimientos básicos de informática puede crear, en muchos casos de una forma muy automatizada, una o más tablas donde guardar sus datos, algún formulario para facilitar la introducción de dichos datos e incluso informes o resúmenes de aspecto profesional.

Sin embargo, no es lo mismo crear una base de datos para gestionar una biblioteca personal que crear una base de datos para una investigación, ya que los requerimientos exigibles para esta última son, o deberían ser, mucho más estrictos. La razón es sencilla: las bases de datos para investigación constituyen el soporte de un estudio científico. De la calidad e integridad de los datos que en ella estén al-

macenados dependerá en gran medida la calidad, integridad y validez final de dicho estudio. Si los datos de partida contienen una tasa significativa de errores o son incompletos, todo el trabajo posterior se verá afectado. Esta situación se puede incluso agravar debido a que muchas veces es necesario un cierto preproceso de los datos, por ejemplo aplicando determinadas operaciones en las que intervienen uno o más campos de una o más tablas para obtener un resultado concreto, o/y es necesario realizar una selección de la información, por ejemplo a través de consultas complejas en las que intervienen varias tablas y múltiples criterios. Si no se realizan correctamente, ambas acciones pueden introducir errores adicionales en la información a partir de la cual deseamos obtener conocimiento.

Usualmente, las personas sin una formación académica en informática utilizan procedimientos artesanales para la creación de aplicaciones, que básicamente consisten en la práctica conocida como "codificar y corregir" o, dicho de otra forma, ensayo y error. Al no ser un método riguroso, las posibilidades de éxito (incluso de éxito parcial) disminuyen drásticamente con la complejidad de la aplicación a desarrollar, disminución que se acentúa todavía más cuanto menos experiencia posea el desarrollador. Y, lo que es todavía peor, debido a que no se realiza ninguna verificación ni validación formal de la aplicación, es posible que se dé como válida y se utilice una base de datos en la que existan errores o deficiencias potencialmente graves, errores y deficiencias que difícilmente se habrían podido detectar en las pruebas informales que se hubieran realizado.

Estas incorrecciones podrían ocasionar, entre otras, las siguientes situaciones, apareciendo solas o combinadas:

- Que se detecte cuando ya es demasiado tarde (se haya cargado la base de datos) que no se recopiló algún atributo necesario para su posterior tratamiento. En el mejor de los casos y, suponiendo que ello es posible, habría que añadir a cada elemento (registro) de la tabla implicada el dato que falte.
- Al realizar el análisis de los datos, y debido a que en la introducción de dichos datos no se realizaba ninguna

Correspondencia: J. Macías del Campo
Correo electrónico: javier.macias@uah.es

comprobación automática de los valores introducidos, se detecte un número significativo de valores atípicos, sin poder saber *a priori* si se trata o no de errores. En este caso, habría que recomprobar manualmente dichos datos si fuera posible –con el esfuerzo que ello supondría– o eliminar los elementos que contengan valores que se puedan identificar sin lugar a dudas como inválidos, lo cual podría ocasionar que el número de elementos resultante fuera inferior al necesario para poder realizar un estudio con el nivel de confianza establecido inicialmente.

- Que se detecte un error causado por el uso de una expresión incorrecta en el cálculo de un resultado dependiente de uno o más campos o por el uso de una consulta errónea. Aunque corregir un defecto de este tipo suele ser relativamente sencillo, durante el proceso de eliminación del mismo existe la posibilidad de introducir inadvertidamente errores en otros lugares. Además, si el error se detecta cuando el estudio se encuentra en una fase avanzada, podría ocasionar la repetición de las partes ya realizadas.

- Que existiendo errores, no se detecten. Los resultados podrán estar, por tanto, falseados en mayor o menor grado. Un sólo error grave podría incluso invalidar totalmente un estudio que, sin embargo, se presente como correcto.

Una mejora sustancial en las posibilidades de éxito es aportada por la rama de la informática conocida como Ingeniería del Software, que proporciona las denominadas “metodologías de desarrollo”, cuya misión es guiar el desarrollo de aplicaciones y sistemas informáticos con el objetivo de obtener sistemas de calidad de una forma eficiente.

Antes de pasar a comentar brevemente las fases principales comunes a la mayoría de las metodologías, se quiere resaltar aquí el concepto de sistema. Para el caso particular de las bases de datos, se puede decir que la base de datos (aplicación) no es un ente independiente, sino que está contenida en un sistema de información, en el cual existen otros elementos con los cuales interacciona. Por un lado, necesitará de una determinada infraestructura tecnológica para su funcionamiento: ordenadores (PC, servidores, etc.), sistema operativo (Windows®, UNIX®, etc.), sistema gestor de bases de datos (dBase®, Microsoft Access®, etc.), redes, dispositivos de entrada-salida (escáneres, impresoras, lectores de código de barras, etc.) y dispositivos de copia de seguridad (disqueteras, unidades de cinta, grabadoras de CD, etc.), entre otros. Por otro lado, deberá interaccionar con personas o/y otras aplicaciones para obtener datos de ellas, y la información de salida deberá proporcionarse en algún formato preestablecido para poder realizar su posterior procesamiento, procesamiento que se llevará a cabo por una persona o por otra aplicación como, por ejemplo, un paquete estadístico.

Para ilustrar la importancia del concepto de sistema, nótese que de nada sirve construir una base de datos si luego no va a ser factible su uso dentro del entorno real. Esto sucede, por ejemplo, cuando es necesario que el médico utilice dicha base durante la consulta pero, por la forma en que se deben de introducir los datos en la misma, los tiempos requeridos son inaceptables. Otro caso similar

ocurriría si la base de datos no es capaz de proporcionar la información de salida en un formato válido para su posterior procesamiento, en caso de que éste fuera necesario.

De forma resumida, las fases contempladas en las metodologías para el desarrollo de sistemas de información son las siguientes:

- **Análisis:** qué es lo que se quiere y qué es lo que se tiene, estudio de viabilidad del proyecto y selección de la opción más adecuada.
- **Diseño:** cómo lo vamos a realizar.
- **Construcción:** implementar lo diseñado.
- **Implantación, aceptación del sistema y operación:** puesta en funcionamiento del sistema, validación y utilización del mismo.
- **Mantenimiento:** eliminar errores detectados durante el funcionamiento, añadir nuevas funcionalidades requeridas y adaptar a los cambios del entorno.

En las metodologías se establecen, para cada fase, las actividades a realizar, las técnicas a utilizar y los productos a obtener (una parte importante de estos últimos se utilizan para documentar todo el proceso). Las metodologías también contemplan aspectos como la seguridad, formación del usuario, manuales de uso y explotación, aseguramiento de la calidad, etc. Además, suelen ser flexibles, de forma que se pueden adaptar a cada caso particular. Esta adaptabilidad se refleja fundamentalmente en que la profundidad y el esfuerzo requeridos para la ejecución de la mayoría de las actividades es dependiente de la naturaleza particular del desarrollo a realizar. Por ejemplo, mientras que para una aplicación dada la ejecución de una actividad determinada puede requerir la confección de varias páginas de documentación, la misma actividad podría requerir simplemente una frase o párrafo (incluso podría bastar un “No procede en este caso.”) para otra aplicación de naturaleza diferente.

Sin embargo, el uso de una metodología requiere de la participación de personal adecuado, tanto para su aplicación como para cumplir las disposiciones en ella establecidas. Los perfiles del personal participante sugeridos para una base de datos de investigación sanitaria son los siguientes:

- **Cliente.** Persona o personas (investigadores sanitarios) interesadas en la realización de la base de datos para la investigación sanitaria. Es el encargado de especificar los requisitos que debe cumplir el sistema y validar finalmente el producto obtenido siendo, por tanto, el perfil más importante, ya que el objetivo de todo el proceso es satisfacer las necesidades por él expresadas.
- **Usuario.** Persona o personas que utilizarán el sistema. Puede coincidir total o parcialmente con el cliente. Su participación es esencial y no se debe olvidar ya que, en otro caso, podría ocurrir que éste lo rechazara o infrutilizara por no ser de su agrado. Colabora en la especificación del sistema y en la validación del mismo.

- Estadístico o/y otros perfiles especializados. Dependiendo del tratamiento final que se vaya a dar a la información obtenida, puede ser necesaria la participación de un estadístico (si lo que se pretende realizar con dicha información es un análisis estadístico complejo) u otros perfiles.

- Informático. Persona o personas expertas encargadas de la realización del desarrollo.

- Personal del Dpto. de Sistemas de Información. Si el sistema se va a usar dentro de una determinada institución o va a utilizar de alguna forma recursos de ésta, es necesario contar con el Dpto. de Sistemas de Información de dicha institución. Ellos podrán proporcionar infraestructura para la realización del sistema, detallarán las posibles normativas que regulen este tipo de desarrollos o sistemas en dicha organización, y podrán prestar ayuda o información para la conexión del nuevo sistema de información con otros existentes para que, por ejemplo, se pueda nutrir de forma automática de datos ya presentes sin necesidad de tener que introducirlos manualmente.

- Personal de aseguramiento de la calidad. Controla todo el proceso, comprobando que lo obtenido se adecua a los requisitos y normativas establecidas. En ningún caso debe coincidir el personal encargado del aseguramiento de la calidad de un desarrollo con cualquier otro personal de dicho desarrollo porque podría afectar a su imparcialidad.

En los casos más sencillos, el cliente coincide con el usuario y el informático encargado de desarrollar el proyecto perteneciente al Dpto. de Sistemas de Información donde se utilizará el sistema. Aunque en principio bastaría únicamente con estas dos personas, sería muy conveniente, por no decir imprescindible, la participación de una tercera que tuviera la experiencia suficiente para encargarse del aseguramiento de la calidad.

Un caso que merece atención especial ocurre cuando el cliente (el investigador sanitario) posee la experiencia y formación suficiente para poder ser el propio desarrollador del sistema. Nótese que la formación y experiencia requerida dependerá de la complejidad del sistema a desarrollar. En este caso, suponiendo que el investigador dispusiera del tiempo necesario y así lo quisiera, podría asumir la labor de desarrollo, lo cual aporta ventajas evidentes.

Sin embargo, sigue siendo igual de necesaria la participación de una persona encargada del aseguramiento de la calidad y, cuando proceda, de personal del Dpto. de Sistemas de Información u otro personal con perfil especializado. Y, en ningún caso, se podrán relajar los procedimientos requeridos, como sucede a veces en estos casos aduciendo razonamientos baladíes del tipo “no necesito documentar el proceso de desarrollo, ya que lo he hecho yo mismo y sé perfectamente lo que he hecho” o “no he hecho ningún manual de usuario, ya que el manejo de la aplicación es evidente y si el usuario tiene algún problema me lo puede preguntar directamente a mí.”

Aunque a primera vista pueda parecer más engorroso, el uso de una metodología y de todo lo que ello conlleva, como la participación de personal con el perfil adecuado, es la única forma de asegurarse una probabilidad de éxito alta en el desarrollo de cualquier sistema informático, tanto en lo que respecta a la calidad final del mismo como a la eficiencia con la que se realiza todo el proceso. Por desgracia, muchas veces se recurre al informático con aplicaciones problemáticas desarrolladas de manera informal, mal diseñadas y con poca o ninguna documentación de soporte. En estos casos, no es raro que sea necesario desechar lo realizado y comenzar desde el principio, esta vez sí, siguiendo una metodología de trabajo. Y, en caso de que el informático decida “apuntalar” la aplicación problemática, raramente podrá asegurar con un mínimo de confianza que dicha aplicación es suficientemente correcta y que la información obtenible de ella tendrá la calidad requerida, exigiendo además, por lo general, un esfuerzo conjunto mayor que el que habría sido necesario si se hubiera realizado el desarrollo correctamente desde el principio.

BIBLIOGRAFÍA GENERAL

- Duff J. Databases – Back to the Future. PC Update Online, 1994. [<http://www.melbpc.org.au/pcupdate/9407/9407article10.htm>].
- McConnell S. Desarrollo y Gestión de proyectos informáticos. McGraw-Hill 1997; 152-153.
- Métrica versión 3, del Ministerio de Administraciones Públicas, es un ejemplo de metodología. Se puede descargar de <http://www.map.es/csi/metrica3/index.html#herramientas>. Téngase en cuenta que ésta es una metodología muy amplia y sería necesario adaptarla al ámbito particular tratado aquí, si se deseara hacer uso de la misma.