

Esquema de confidencialidad y tratamiento informático de los datos en el proyecto

Índice

1	Introducción.....	3
2	Esquema de usuarios y conjuntos de datos.....	3
3	Niveles de datos.....	5
3.1	Primer nivel: atención primaria	5
3.2	Segundo nivel: Transmisión	6
3.3	Tercer nivel: Procesamiento de datos	6
4	Descripción de los datos	6
4.1	Base de datos de atención primaria	6
4.2	Fichero de transferencia.....	7
4.2.1	Generación.	7
4.2.2	Estructura y contenido.....	7
4.2.3	Compresión y encriptación	8
4.2.4	Transmisión.....	9
4.3	Fichero procesado	9
4.3.1	Generación	9
4.3.2	Estructura y contenido (modelo de datos de BIFAP).....	9
4.3.3	Almacenamiento	10
5	Anonimización de los datos y mantenimiento de la confidencialidad en el proyecto BIFAP.....	10

1 Introducción

En este documento se explican los sucesivos procedimientos informáticos a que va a ser sometida la información en el Proyecto BIFAP. Estos procedimientos van a asegurar un elevado nivel de protección de datos desde que se extraen de sus registros informáticos primarios hasta que se introducen en BIFAP, y han sido concebidos para evitar vulneraciones de la confidencialidad de la información y de la privacidad de los pacientes y los médicos colaboradores.

Con el objeto de representar globalmente el proceso, se comienza describiendo un *esquema de usuarios y conjuntos de datos* del Proyecto BIFAP, y, seguidamente, se hace una *descripción de los datos* y del *tratamiento* a que van a ser sometidos éstos en el proyecto, incluida la *clasificación* de los mismos que se va a hacer en el CPD antes de trasvasarlos a BIFAP, punto en el que termina el proceso descrito en este documento.

2 Esquema de usuarios y conjuntos de datos

El presente esquema no sólo tiene en cuenta a los usuarios autorizados de los distintos conjuntos de datos relacionados con el Proyecto BIFAP, sino también a las personas que intenten utilizarlos sin tener autorización para ello (los potenciales “usuarios” malintencionados). El uso del calificativo «malintencionados» es aquí preferible, pues a efectos de protección de datos es mejor presuponer que estos intentos no autorizados de utilizar los datos no van a ser inocentes. Del mismo modo, el término usuario se emplea aquí de una manera muy laxa, pues usuarios de la información en sentido estricto sólo lo serían los investigadores del proyecto.

USUARIOS	MEDICO COLABORADOR	“USUARIOS” MALINTENCIONADOS (CANAL DE COMUNICACION NO SEGURO)	PERSONAL DEL CPD DE BIFAP	INVESTIGADORES DEL PROYECTO BIFAP
DATOS	BASE DE DATOS MEDICO AP	FICHERO DE TRANSFERENCIA (ENCRIPADO)		FICHERO PROCESADO

En la fila superior figuran los distintos tipos de usuarios que acceden o que pudieran acceder a los distintos conjuntos de datos inherentes al proyecto. Es importante recalcar que acceder no significa en este documento capacidad de comprender la información sino simplemente de ver la información, que será inteligible o no dependiendo de la etapa del proceso en que se encuentre.

En la fila inferior se representan los diferentes conjuntos de datos que tienen relación con el proyecto. Los cuatro tipos de usuarios pueden definirse de la siguiente manera:

1. **Médico colaborador.** Es el médico que aporta voluntariamente datos de las historias clínicas de los pacientes adscritos a su cupo a BIFAP.
2. **“Usuarios” malintencionados.** Serían quienes accedieran al fichero de transferencia de BIFAP sin tener autorización para hacerlo.
3. **Personal del CPD de BIFAP.** Es el conjunto de personas que gestionan, procesan y mantienen los datos de BIFAP.
4. **Investigadores del Proyecto BIFAP.** Se trata del personal investigador que tiene acceso al fichero procesado de BIFAP

En cuanto a los datos, distinguimos en el anterior esquema tres conjuntos distintos, que se definen así:

1. **Base de datos del médico de atención primaria.** Este conjunto de datos no pertenece al Proyecto BIFAP. Representa los datos de historias clínicas que corresponden al cupo de pacientes del médico colaborador.
2. **Fichero de transferencia.** Es el fichero generado por el módulo de exportación de datos. Supone un subconjunto de los datos del apartado anterior. En la mayoría de los casos el médico colaborador es el encargado de generarlo y enviarlo al CPD, pero también podrá hacerlo otro miembro autorizado de su equipo en el que aquél delegue. El esquema representa la primera de estas dos posibles situaciones, y es evidente que en la segunda, la persona autorizada ocuparía en el esquema exactamente el mismo casillero que el médico colaborador, y con las mismas implicaciones.
3. **Fichero procesado.** Se trata del fichero de datos que se genera tras procesar el fichero de transferencia en el CPD. El fichero procesado se almacenará en el CPD.

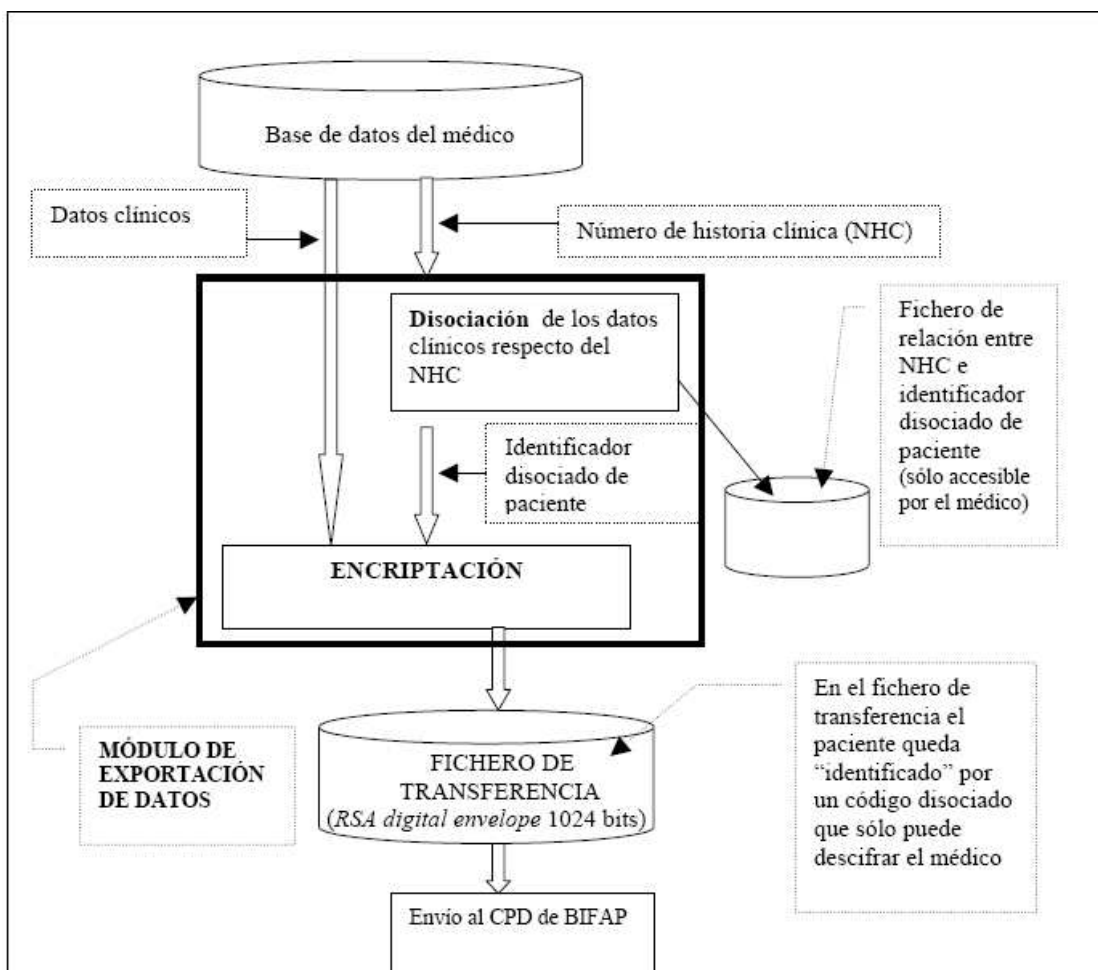
En el esquema descrito, el segmento de trazo grueso que hay en la línea central del esquema representa el blindaje de los datos que supone la encriptación del fichero de transferencia ante potenciales accesos no autorizados, e incluso para el propio personal del CPD mientras no se le aplique el algoritmo de desencriptación. Las líneas discontinuas que prolongan por sus extremos el anterior indican que los usuarios no sólo pueden acceder al soporte o fichero contiguo sino que también pueden comprender su contenido. Independientemente del tipo de trazo, la contigüidad de usuarios y ficheros a lo largo de toda la línea central solamente indica los accesos que plausiblemente se van a poder producir con mayor facilidad. Los *mecanismos de seguridad* inherentes a este esquema se explican en posteriores apartados de este documento.

3 Niveles de datos

A continuación, se describe un modelo en capas que representa las diferentes etapas de procesamiento de los datos desde que se registran en la consulta de atención primaria hasta que llegan al CPD. Para cada uno de los niveles se establecen mecanismos destinados a mantener la confidencialidad de los datos y la anonimización de los pacientes.

3.1 Primer nivel: atención primaria

En este nivel se efectúa la extracción y encriptación de datos a partir de la base de datos de atención primaria del médico colaborador. Esto se hace de forma selectiva, es decir, únicamente de aquellos datos que son relevantes para la investigación farmacoepidemiológica, y prescindiendo de los datos personales.

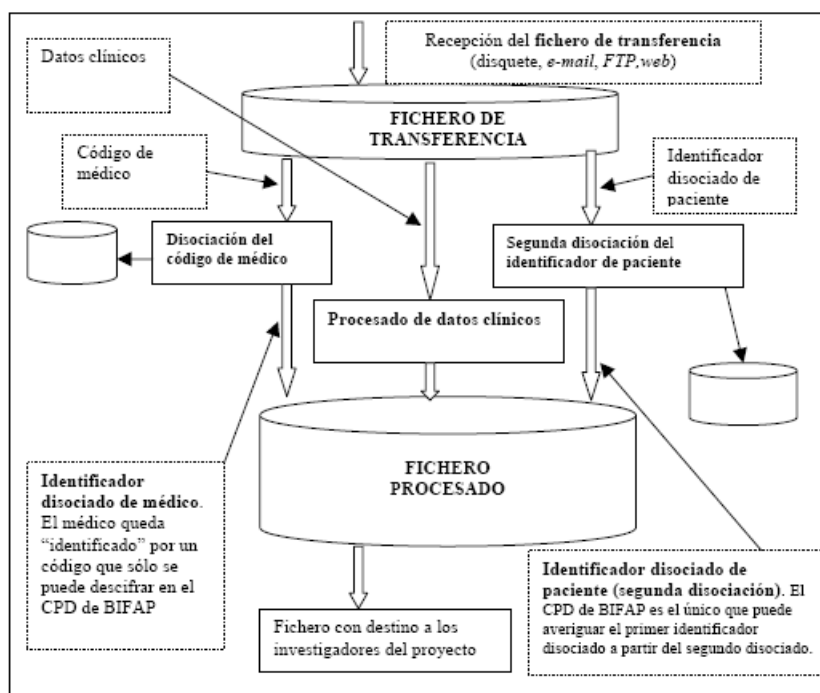


3.2 Segundo nivel: Transmisión

En este nivel se transmite el fichero de transferencia, que es un fichero en formato ASCII extendido, y encriptado con un sobre digital (*RSA digital envelope*). El fichero podrá ser enviado vía Web.

3.3 Tercer nivel: Procesamiento de datos

En este nivel se sitúa el CPD de BIFAP, a donde llega el fichero de transferencia.



4 Descripción de los datos

4.1 Base de datos de atención primaria

La base de datos de atención primaria es el registro en el que se almacena la información derivada del trabajo de los profesionales de atención primaria, que normalmente ha sido introducida por ellos mismos durante su trabajo, aunque cada vez más se están poniendo en marcha sistemas que permiten la introducción de información (p. ej., resultados de pruebas de laboratorio, informes de interconsulta, etc.) desde otros niveles asistenciales. En nuestro ámbito, el esquema predominante es que las bases de datos de atención primaria se encuentren en los propios centros de salud, y que contengan la información correspondiente a varios cupos.

Para conocer exactamente los datos susceptibles de registro a través de cada programa, hay que remitirse al modelo de datos que proporcione el fabricante. En general, los programas de atención primaria van a permitir registrar datos clínicos (prescripciones, diagnósticos, resultados de pruebas complementarias, etc.) y datos administrativos (nombre, apellidos, fecha nacimiento, dirección, teléfono, etc.) de los pacientes.

4.2 Fichero de transferencia

4.2.1 Generación.

El fichero de transferencia se genera a partir de la base de datos del médico colaborador, y contiene exclusivamente los datos de relevancia para el Proyecto BIFAP. Como norma preceptiva no se extraen datos personales que permitan identificar a los pacientes. Ahora bien, es necesario distinguir a cada paciente de los demás, y esto se efectúa a través de un **procedimiento de disociación** que rompe la relación entre el código que usa el médico para identificar a su paciente (el *número de historia clínica*) y el código de paciente exportado (**número aleatorio** que se va a utilizar como **código anonimizado de paciente**, y que será el único código de paciente que se reciba en el CPD). La relación entre ambos códigos se almacenará en un fichero al que **sólo el médico** podrá acceder mediante su código de usuario y su contraseña secreta. De esta forma disociamos la identidad del paciente de sus datos clínicos.

Una vez recibido el fichero de transferencia en el CPD, se le aplicarán filtros de entrada, se procesarán sus datos (aplicación de controles de calidad, clasificación, etc.), y se efectuarán dos **nuevos procedimientos de disociación** encaminados a reforzar la protección de los datos:

1. Disociación del **código anonimizado de paciente** recibido en el fichero anterior. En el CPD se genera una tabla que relaciona el código de paciente generado por el módulo de exportación de datos BIFAP con un **nuevo código anonimizado de paciente** generado aleatoriamente. Esta segunda disociación sirve para elevar el estándar de anonimización de los pacientes.
2. Disociación del **código de médico** recibido en el fichero de transferencia. En el CPD se genera una tabla que relaciona el código de médico con un nuevo **código anonimizado de médico** generado aleatoriamente. De esta forma se va a garantizar la ausencia de datos identificativos de los médicos colaboradores en BIFAP.

Las dos tablas que se obtienen a partir del fichero de transferencia serán almacenadas en el CPD, y nunca formarán parte del *fichero procesado* (véase más adelante).

4.2.2 Estructura y contenido.

Al ser el fichero de transferencia un paso intermedio entre la base de datos del médico de atención primaria y la base de datos BIFAP, su estructura va a depender del programa de atención primaria que use el médico.

El contenido de un fichero de transferencia es el siguiente:

1. **Información sobre el médico colaborador y la exportación:** Identificación del médico, fecha de exportación, y fechas de inicio y final de exportación en exportaciones incrementales.
2. **Información administrativa del paciente:**
 - *Código anonimizado de paciente* (id_paciente). Se trata de un número aleatorio calculado a partir del número de historia clínica (NHC) del paciente. A este número se le ha aplicado una función de disociación que genera un número aleatorio. La relación entre NHC e id_paciente se almacena en el ordenador del médico en un fichero al que **sólo el médico** puede acceder mediante su nombre de usuario y contraseña.
 - *Datos administrativos* como son fecha de alta con el médico, situación administrativa del paciente, fecha de cambio de situación administrativa.
3. **Información clínica del paciente**
 - **Datos básicos** como sexo, mes y año de nacimiento, fecha de última visita, fechas de contacto con el paciente...
 - **Datos correspondientes a procesos.** Gravedad, fecha, clasificación, observaciones del médico sobre el proceso...
 - **Datos correspondientes a problemas clínicos, factores de riesgo y factores que pueden condicionar el pronóstico o la actitud clínica.**
 - **Datos complementarios.** Se trata de datos clínicos que complementan la información relativa a diagnósticos, acontecimientos clínicos y tratamientos.
 - **Antecedentes personales y familiares.**
 - **Análisis y pruebas realizadas al paciente.**
 - **Interconsultas a especialistas** y pruebas complementarias realizadas por especialistas.
 - **Vacunas.**
 - **Prescripciones.** Siempre clasificadas por el código nacional de fármaco así como información sobre el tratamiento (posología, dosis, duración, recetas emitidas...).
 - **Productos y preparados farmacéuticos** como son las vacunas hiposensibilizantes y las fórmulas magistrales, que no tienen código nacional de especialidad.

4.2.3 **Compresión y encriptación**

El fichero con la estructura descrita anteriormente se genera en forma de texto plano (ASCII extendido), y posteriormente se comprime y encripta.

1. **Algoritmos de encriptación de ficheros:** Los algoritmos de encriptación elegidos para encriptar los ficheros de BIFAP son dos: RSA y DES3 (Triple DES), combinados en forma de protocolo para generar lo que se conoce como *sobre digital RSA* (*RSA digital envelope*). De este modo la desencriptación sólo podrá realizarse en el CPD de BIFAP que es donde se dispone de la clave de desencriptación.
2. **Funcionamiento de la encriptación:** Cuando desde una fuente de información BIFAP se genera un fichero con contenido susceptible de ser cifrado, el módulo de

exportación invoca a la utilidad de encriptación. Una vez encriptado el fichero, éste está listo para ser enviado al CPD de BIFAP.

4.2.4 Transmisión

Una vez generado el fichero de transferencia convenientemente encriptado, se procede al envío del mismo al CPD:

- Envío telemático del fichero mediante vía Web a www.bifap.org. En este caso, el fichero será almacenado en un archivo gestionado por el responsable de seguridad del Proyecto BIFAP.

4.3 Fichero procesado

4.3.1 Generación

El fichero procesado se obtiene a partir de los datos que se reciben del fichero de transferencia. Una vez que este fichero llega al CPD, se efectúa su desencriptación para luego aplicarle una serie de procesos. Esquemáticamente, los procesos son los siguientes:

1. Segunda disociación del código anonimizado del paciente. Se trata de generar un nuevo código anonimizado de paciente a partir del código de paciente recibido (el procedimiento se explica en el apartado de generación del fichero de transferencia).
2. Disociación del código de médico (tal y como se explica en el mismo apartado).
3. Clasificación de diagnósticos clínicos de acuerdo a los diccionarios internos de BIFAP (CIAP, CIE y otros).
4. Clasificación de prescripciones según el diccionario interno de BIFAP (que contendrá los Códigos Nacionales de Especialidades Farmacéuticas, y sus correspondientes códigos de principios activos).
5. Clasificación de información clínica relevante no encuadrable dentro de diagnósticos ni de prescripciones (para clasificar estos *datos clínicos misceláneos* se utilizará un diccionario interno de BIFAP).
6. Realización de controles de calidad de los datos.

4.3.2 Estructura y contenido (modelo de datos de BIFAP)

Los datos que contiene el fichero procesado se corresponden con los datos a incluir en BIFAP. La información recogida en el fichero procesado es la siguiente:

1. Datos referidos al **médico colaborador**.
 - Código impersonal del médico. En el CPD de BIFAP se generará un código de médico colaborador que no tiene ninguna relación con sus datos personales. Los usuarios de BIFAP en ningún momento podrán conocer la identidad del médico y de esta forma, sólo el responsable del fichero y/o los encargados del tratamiento de los datos (que, como se ha indicado, estarán sometidos por ley a secreto profesional) podrán relacionar el código del médico con su identidad.
2. Datos referidos al **paciente**:
 - **Código impersonal**. Al estar doblemente disociado el código del paciente, será imposible identificar con este código a ningún paciente.

- **Datos administrativos.** Fecha de registro, fecha de salida, estado administrativo del paciente, causa de cambio de estado...
- **Datos demográficos básicos.** Sexo, fecha de nacimiento, grupo familiar...
- **Procesos.** El diagnóstico de cada proceso vendrá codificado mediante clasificaciones aceptadas de forma generalizada como pueden ser la CIAP (Clasificación Internacional de Atención Primaria) o la CIE (Clasificación Internacional de Enfermedades). Además de este código los registros contendrán las fechas de comienzo y fin del proceso, y otros datos asociados como gravedad al inicio, evolución y observaciones relativas a los procesos de morbilidad...
- **Prescripción de medicamentos.** Para la identificación del medicamento se utilizará el Código Nacional de Especialidad Farmacéutica correspondiente al medicamento prescrito. Asociados a este código van a figurar los datos esenciales de la prescripción: fecha, posología, fecha de las recetas, y número de envases por receta.
- **Datos complementarios.** Son todos aquellos datos del paciente que no se pueden clasificar como diagnósticos ni como prescripciones, p. ej., peso, talla, presión sanguínea, resultados de laboratorio y pruebas complementarias, actividades preventivas, datos de anamnesis, etc.).

4.3.3 Almacenamiento

Los ficheros procesados se almacenarán en el CPD de BIFAP, que se encuentra en la División de Farmacoepidemiología y Farmacovigilancia de la Agencia Española de Medicamentos y Productos Sanitarios. En el CPD se adoptará un plan de seguridad que garantice absolutamente la protección de la información que allí se almacene, que es la correspondiente a los ficheros de transferencia recibidos, los ficheros procesados y la base de datos BIFAP.

Los procedimientos de seguridad se hayan descritos en un Documento de Seguridad que podrá ser auditable. Dicho documento de seguridad sigue las normas de seguridad para ficheros con nivel de protección alto.

5 Anonimización de los datos y mantenimiento de la confidencialidad en el proyecto BIFAP

En este apartado se resume la protección de los datos en el Proyecto BIFAP, con especial referencia a la protección de la identidad de los pacientes y los médicos. Este aspecto del proyecto es fundamental, y se centra en la *anonimización* de los datos, que no es otra cosa que eliminar las identidades y los datos identificativos personales de los conjuntos de datos propios del proyecto a través de los procesos de *disociación* descritos previamente.

El siguiente esquema representa la posibilidad o imposibilidad de conocer la identidad de pacientes y médicos colaboradores en los distintos niveles del Proyecto BIFAP:

En síntesis, toda la información extraída de las fuentes primarias va a ser encriptada antes de ser exportada, y además la identidad de los pacientes no forma parte de la información extraída sino que éstos van a ser individualizados por un código impersonal. En consecuencia, toda la información relativa a los pacientes va a quedar anonimizada desde el mismo momento de su extracción, y así es como se va a enviar al CPD y se va a introducir

en BIFAP. Por este motivo, **el personal del CPD no va a conocer la identidad de los pacientes**. Debido a la logística del proyecto, el personal del CPD sí que debe conocer, no obstante, a cuál de los médicos colaboradores corresponde la información remitida, y por lo tanto, ha de ser capaz de vincular sus códigos individuales con sus identidades.

La **validación de los casos** exige recorrer el camino inverso, es decir, desde la información contenida en BIFAP hasta los médicos colaboradores. Los **investigadores del proyecto no van a conocer las identidades de los médicos ni de los pacientes** a través de sus códigos disociados, y sin embargo van a requerir eventualmente que los médicos remitan información sobre algunos de sus pacientes con el objeto de validar la información de BIFAP. Para que esto sea posible, el **personal del CPD se encargará de “reasociar” los códigos de disociación del médico y de los pacientes, dirigirse al correspondiente médico, comunicarle los códigos de los pacientes** de su cupo cuya información haya que validar (es decir, los códigos disociados de paciente que llegaron al CPD en el fichero de transferencia), y solicitarles copias de los informes clínicos de esos pacientes. Por su parte, **el médico colaborador va a poder relacionar los códigos de los pacientes con sus correspondientes números de historia clínica** accediendo a través de su ordenador y mediante clave secreta personal a un fichero de relación que permite este vínculo. De este modo, **el médico va a poder identificar a los pacientes** de los que se precisa información, **fotocopiar los informes clínicos solicitados, anonimizarlos, y enviarlos al CPD**, donde se comprobará si la eliminación de datos identificativos en las fotocopias se ha hecho correctamente, y se corregirá cualquier defecto de anonimización que pudiera detectarse en las mismas, antes de pasárselas a los investigadores del proyecto.

