- Expediente N.º: EXP202207910

RESOLUCIÓN DE ARCHIVO DE ACTUACIONES

De las actuaciones practicadas por la Agencia Española de Protección de Datos y teniendo como base los siguientes:

HECHOS

PRIMERO: La Agencia Española de Protección de Datos ha tenido conocimiento a través de una denuncia de ciertos hechos que podrían vulnerar la legislación en materia de protección de datos.

Con fecha 19 de julio de 2022, la Directora de la Agencia Española de Protección de Datos instó a la Subdirección General de Inspección de Datos (SGID) a iniciar las actuaciones previas de investigación a las que se refiere el artículo 67 de la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (en adelante, LOPDGDD) para investigar al CONSORCIO BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACIÓN, CONSORCIO HOSPITAL CLÍNICO DE BARCELONA (HCB), FUNDACIÓN INSTITUTO DE INVESTIGACIÓN SANITARIA ILLES BALEARS, FUNDACIÓN PÚBLICA ANDALUZA PARA LA GESTIÓN DE LA INVESTIGACIÓN EN SALUD DE SEVILLA y HOSPITAL UNIVERSITARIO DOCE DE OCTUBRE, en relación con los siguientes hechos:

El Consorcio Barcelona Supercomputing Center (en lo sucesivo BSC-CNS) tiene acceso a datos personales de cuyo tratamiento serían responsables los hospitales 12 de octubre de Madrid, Clinic, Virgen del Rocío y Son Espases. Los datos accedidos serían: DNI, números de teléfono, número de colegiado, historia clínica completa estructurada y no estructurada; no existiendo, al parecer, un control de acceso adecuado a dichos datos, pudiendo cualquier trabajador (incluidos los becarios) acceder a ellos y descargárselos.

SEGUNDO: La Subdirección General de Inspección de Datos procedió a la realización de actuaciones previas de investigación para el esclarecimiento de los hechos en cuestión, en virtud de las funciones asignadas a las autoridades de control en el artículo 57.1 y de los poderes otorgados en el artículo 58.1 del Reglamento (UE) 2016/679 (Reglamento General de Protección de Datos, en adelante RGPD), y de conformidad con lo establecido en el Título VII, Capítulo I, Sección segunda, de la LOPDGDD, teniendo conocimiento de los siguientes extremos:

Durante las actuaciones de investigación se han llevado a cabo los siguientes requerimientos de información:

Requerimientos a BSC-CNS:

En las contestaciones recibidas se indica lo siguiente:

BSC-CNS es una entidad de derecho público integrada por la Administración General del Estado, por la Administración de la Generalitat de Catalunya y por la Universitat Politècnica de Catalunya y creada mediante convenio de colaboración de fecha 9 de marzo de 2015 para la gestión y promoción de la colaboración científica, técnica, económica y administrativa de las instituciones que la integran para el equipamiento y gestión del BSC-CNS, centro español de supercomputación y de I+D+I y agente de ejecución del Sistema Español de Ciencia, Tecnología e Innovación (de acuerdo con lo establecido en el artículo 3.4 de la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación) cuyas infraestructuras forman parte del mapa nacional de Infraestructuras Científicas y Técnicas Singulares (ICTS).

Así pues, esta entidad no tiene por finalidad estatutaria el procesamiento de datos o de información a niveles de público o de entidades en general, sino que se trata de procesos de supercomputación complejos con la finalidad, en la mayoría de los casos, de crear patrones o esquemas de conductas no referidos a sujetos concretos sino a colectividades. Esto lleva implícito que los datos personales no son necesarios para los análisis de supercomputación que se llevan a cabo en los proyectos científicos o investigación de esta entidad.

(…)

En cuanto a la información solicitada al respecto de tratamiento de datos de los Hospitales 12 de octubre de Madrid, Clinic de Barcelona, del Hospital Virgen del Rocío y de Son Espases, es necesario indicar que la relación que vincula a BSC-CNS con estas entidades (con la excepción del Hospital Virgen del Rocío que se trata de la Fundación Pública Andaluza Para La Gestión De La Investigación En Salud De Sevilla, ubicada en la sede de dicho hospital) se halla debidamente regulada en los convenios firmados por ambas partes y que se acompañan a esta respuesta como documentos adjuntos.

Constan efectivamente adjuntos a la respuesta al requerimiento de información los convenios de colaboración a que hace referencia la denunciada. Por otro lado, dos de los convenios aludidos se encuentran publicados en el BOE:

******URL.1***
******URL.2***

En todos los convenios suscritos por BSC-CNS con los distintos hospitales/fundaciones que constan como investigados en estas actuaciones, BSC-CNS tiene la consideración de Encargado del Tratamiento, mientras que los Hospitales/fundaciones tienen la consideración de responsable del Tratamiento.

Continúa manifestando que:

Tal y como se desprende de los convenios adjuntos a esta respuesta, los datos se recibieron anonimizados; en algunos casos el proceso de anonimización fue realizado directamente por el Responsable de Tratamiento, como el firmado con el Hospital Sos Espases, mientras que en los otros convenios referenciados (Hospital 12 octubre, Hospital Clinic y Hospital Virgen del Rocío),

en los cuales también se recibieron los datos de forma anonimizada, pero estaba habilitada la posibilidad de que en algunos casos se pudiesen recibir de forma pseudo-anonimizada y realizada ésta por el responsable del tratamiento, en el caso que hubiere sido necesario para la ejecución del proyecto.

En todos los convenios consta claramente especificado que BSC-CNS recibirá los datos anonimizados o seudonimizados:

Convenio con HCB
ANEXO 1
Contrato de tratamiento de datos de carácter personal.

Cláusula 1.3: El BSC-CNS recibirá los datos anonimizados por parte del HCB. En este contexto, no será objeto de aplicación las normas relativas a la protección de datos personales puesto que los datos personales no son identificables. Ello, no obstante, para los supuestos en los que el BSC-CNS recibiese datos seudonimizados por parte del HCB, se formaliza el presente encargo de tratamiento para regular los derechos y obligaciones de las Partes con relación a la normativa de protección de datos aplicable

Convenio con IdIsBa

Cláusula 5: Corresponde al IdIsBa poner a disposición del BSC-CNS los informes médicos de alta hospitalaria de pacientes ingresados en Neurología, anonimizados, de acuerdo a los requerimientos para su análisis.

Convenio con FISEVI
ANEXO 1
Acuerdo de tratamiento de datos de carácter personal

Cláusula 1.4: Sin perjuicio de los descrito en el apartado 1.3, el BSC-CNS recibirá los datos anonimizados por parte de FISEVI desde origen. En este contexto, no será objeto de aplicación las normas relativas a la protección de datos personales puesto que los datos personales no son identificables según lo establecido en este Acuerdo. Ello, no obstante, para los supuestos en los que el BSC-CNS recibiera los datos seudonimizados desde origen por parte de FISEVI se formaliza el presente encargo de tratamiento para regular los derechos y obligaciones de las Partes con relación a la normativa de protección de datos aplicable.

Convenio con H12O

Cláusula Primera. *Objeto de la colaboración.*
1.1 El objeto de este Convenio es el establecimiento del marco de colaboración científica y técnica entre el BSC-CNS y el H12O con el fin de dar respuesta inmediata a la crisis y emergencia sanitaria que se viviendo en España como consecuencia de la pandemia COVID-19.

1.2 A tal fin, en el marco de esa colaboración que se avanza en los antecedentes administrativos, las Partes convenían de que el H12O transfiera al BSC-

CNS los datos seudonimizados que se circunscriben en el anexo 1 de este Convenio a los efectos de que el BSC-CNS pueda efectuar:

1. Los análisis automáticos del contenido de textos clínicos mediante sistemas de extracción automática y normalización de conceptos clínicos;

2. Reconocimiento de entidades clínicas de relevancia para COVID-19 y generación de datos anotados y estructurados cuyo fin consiste en el desarrollo y evaluación de calidad de sistemas basados en extracción automática de información.

1.3 En este sentido, la trasferencia de datos personales seudonimizados por parte del H12O al BSC-CNS se realizará para los fines exclusiva y estrictamente de este Convenio. Para ello, las Partes firmarán el acuerdo de encargo de tratamiento de los datos personales que se adjunta en el anexo 1.

1.4 Las actuaciones concretas de cooperación conjunta entre las Partes queda definida en el anexo 2 de este Convenio en el ámbito respectivo de sus competencias

BSC-CNS continúa señalando que:

[E]n todos estos casos, los datos tratados en los proyectos han sido anonimizados puesto que no fue necesario disponer de datos personales para los análisis y tratamiento de supercomputación encargados. Estos datos siempre se recibieron, de forma que no se permitiera identificar al interesado, en cuanto no era preciso tratar los datos, de acuerdo al principio de intervención mínima en el tratamiento de los datos.
A pesar de ello y con el objetivo de cumplir la normativa de protección de datos se recogió de forma preventiva como posible tratamiento de datos – en calidad de encargado de tratamiento por parte de BSC-CNS – el siguiente registro de actividades para esos proyectos:

Se adjunta un RAT de PROYECTOS en calidad de Encargado de Tratamiento con la siguiente descripción: Gestión de proyectos donde participa en calidad de encargado de tratamiento prestando algún servicio vía contrato o ejecutando tareas científicas y/o de investigación en virtud de convenios o acuerdos de agrupación.

El RAT contiene además los siguientes campos: Sistema de tratamiento; Origen de los datos; Finalidades; Categorías de interesados; Categorías de datos; Categorías de destinatarios y Medidas de seguridad.

Por lo que respecta a las medidas de seguridad BSC-CNS manifiesta que:

***(…)**, en el que únicamente tenían acceso los miembros de la unidad que participaban en el proyecto respectivo. Dicho sistema permite garantizar la seguridad por grupos, siguiendo los más altos estándares de seguridad. BSC-CNS contrató dicho sistema con el proveedor IBM, quien es el desarrollador de esa herramienta y garantiza su mantenimiento. Asimismo, se destaca de ese sistema, que una de sus características es la de tener espacios de ficheros separados con seguridad, que permiten que otros usuarios de la propia herramienta no puedan acceder a datos de otros grupos, si no están expresamente autorizados.*

En el segundo requerimiento enviado a BSC-CNS se solicita información relacionada con las medidas de seguridad.

Además, se solicitó diversa documentación de muestreo para verificar si se cumplían los requerimientos del RGPD y de los convenios. Por poner un ejemplo, en el Convenio firmado con el Hospital Universitario Son Espases y la Fundación Instituto de Investigación Sanitaria de Illes Balears se establece en la cláusula sexta, último párrafo, la obligación de destruir el conjunto de datos original en todo caso antes del 31 de marzo de 2021. Se solicitó a BSC-CNS remisión de la prueba de destrucción de dichos datos, adjuntándose como documento 10 de la respuesta BSC-CNS 2.

Se ha aportado como documento 4 “Informe independiente del cumplimiento de Medidas de Seguridad relativo a los ficheros o tratamiento de datos personales efectuado BSC-CNS”

Consta también aportado como documento 6, el “Análisis de riesgos”.

Tal y como manifiesta BSC-CNS, se procedió conforme a la siguiente política de actuación:

1) Descripción de los medios de trabajo que se le facilita a los trabajadores.

EL BSC-CNS proporciona los medios técnicos (ordenador, monitor, teclado, ratón) a sus empleados, acceso a los superordenadores que sea necesario y de otros medios materiales tales como dispositivo móvil, cuando así se requiera. Para ello, el responsable directo de ese trabajador deberá rellenar una petición/formulario que deberá ser aprobada por el Departamento de Operaciones y por Administración. En casos excepcionales y debidamente justificados se ha autorizado la utilización de medios personales a petición del usuario. Precisamente esta situación excepcional afecta a un trabajador que tenía acceso a los datos anonimizados en los convenios y a quién se le permitió utilizar su propio ordenador personal debido a la situación excepcional de la pandemia. Este empleado trabajaba en remoto desde País Vasco y se autorizó por parte del Responsable de la Unidad de Data Mining que usara su propio ordenador personal, a petición previa de ese trabajador, a quién se le informó de las normas de seguridad del BSC-CNS para trabajar en remoto. ***(…)***

2) Documento de política de seguridad de la Información y cualquier otro documento o procedimiento de seguridad aprobado en relación con los convenios suscritos.

Los trabajadores de la Unidad de Data Mining eran conocedores de que se trataban textos clínicos anonimizados y que estaban regulados mediante convenios. En este contexto, los trabajadores que accedían a esos datos lo hacían con las correspondientes medidas de seguridad. Los textos clínicos se trataban de manera rigurosa en cuanto a su protección; se distinguían entre “datos de salud”, que es información de salud genérica y no relativa a una persona física (artículos del ámbito, descargas de webs de dominio, patentes) e ‘historias clínicas anonimizadas’. Los primeros se usaban para anotar y entrenar modelos y se publicaban en abierto para la compartirlos con la comunidad científica en aras de investigar y obtener soluciones ante la crisis de pandemia sanitaria; mientras que los segundos (incluso siendo

anonimizados) nunca se publicaban ni se compartían con terceros. Los trabajadores que accedían a esos datos anonimizados conocían las reglas y los límites de observancia respecto a los datos. Así pues, en las reuniones semanales de grupo se repetía y se resolvían dudas sobre la política de datos. Todos estaban informados que, si se detectaba algún dato personal o hubiera sospecha de reidentificación, lo tenían que comunicar inmediatamente al Responsable de la Unidad de Data Mining. El grupo de Data Mining era lo suficientemente pequeño (7 personas) para asegurar que la información era conocida por todos y que tenían que cumplir con las pertinentes medidas de seguridad y de confidencialidad establecidas en la Política de Usuarios y de Seguridad. Se adjunta como documento núm.1.

Además, debe tenerse en cuenta que todos los usuarios conocían las obligaciones y deberes de los trabajadores en cuanto al manejo de datos y medidas de seguridad. Se adjunta como documento núm. 2 "Decálogo de los diez puntos básicos" dónde se recoge las funciones y obligaciones de los usuarios.

El personal con acceso a la información anonimizada era sabedor de que la anonimización se llevaba a cabo de origen por las entidades hospitalarias, de que el BSC-CNS no participaba ni en el proceso, ni en las técnicas anonimización y que el personal del BSC-CNS debía de respetar las medidas de control relativas al acceso a la información anonimizada. También, se les informó de que en caso de ruptura de la cadena de reidentifación tenían que avisar al Responsable de Data Mining para que éste pudiera comunicar de forma fehaciente este hecho a los Hospitales.

(…).

En el último requerimiento de información se solicitó a BSC-CNS la siguiente información:

1.- Declaración de cumplimiento de las obligaciones que se detallan en la Política de Responsabilidades de los Usuarios del BSC-CNS, antes de acceder a la información alojada en los servidores, firmada por todos los usuarios que se detallan en la página 6 de su respuesta al requerimiento de información AEPD, de fecha 26/01/2023 y registro de entrada REGAGE23e00005485531 y en el documento núm. 3 correo electrónico del listado de usuarios que tenía acceso autorizado por el HCB.

En cumplimiento a este requerimiento, se aporta como documento 1 la documentación requerida. La Política de Responsabilidades del denunciante como usuario del BSC-CNS no consta firmada, aunque debía aceptarla para poder acceder a los recursos del BSC-CNS.

Además, dado que al menos dos de los convenios eran públicos, pues se hallaban publicados en el BOE, no es posible afirmar que el personal al servicio de BSC-CNS no conociera el deber de confidencialidad pues en ambos casos consta esa obligación:

Convenio HCB:

Cláusula Sexta. Confidencialidad.
6.1 Las Partes se comprometen a mantener confidenciales los términos y condiciones del presente Convenio, así como la información de la otra Parte a la que accedan como consecuencia del mismo, garantizando que no revelarán los datos que lleguen a conocer por la relación contractual que se establece, con la excepción de disponer de autorización expresa y por escrito de la otra Parte. Se exceptúa la obligación de confidencialidad cuando por disposición legal o a requerimiento de una autoridad judicial o administrativa, una Parte debe revelar la información confidencial de la otra.
A tal efecto, y con el fin de dar cumplimiento a la Ley 19/2014 de transparencia, acceso a la información pública y buen gobierno, se publicará la información relativa a los convenios de colaboración según lo previsto en el artículo 14 de la mencionada ley.
6.2 Cualquier información, sea cual sea su naturaleza, que pueda ser facilitada entre las Partes en relación al objeto del presente Convenio, será considerada como «Información Confidencial» será tratada confidencialmente por la Receptora, sus empleados y colaboradores, y no será revelada, total o parcialmente, sin el consentimiento previo escrito de la Emisora. (…)

ANEXO 1
Contrato de tratamiento de datos de carácter personal
CLÁUSULAS
1. Objeto
1.1 El Encargado del Tratamiento guardará secreto sobre los datos de carácter personal de los cuales tenga conocimiento con motivo de esa colaboración. Todo el personal de la entidad Encargada del Tratamiento que acceda y/o trate datos de carácter personal está sujeto al secreto profesional y deber de confidencialidad, obligación que continua una vez finalizada la relación entre las Partes.

Convenio H12O:

Octava. *Confidencialidad.*

8.1 La Información Confidencial será tratada confidencialmente por la Receptora, sus empleados y colaboradores, y no será revelada, total o parcialmente, sin el consentimiento previo escrito de la Emisora.

Además, con la tercera solicitud de información se acompañó documentación aportada por el denunciante. En concreto, las 5 hojas en las que constan datos personales que no están anonimizados y para ello se ofuscó el logueado del denunciante. Se requirió justificación de estas circunstancias y los motivos que han ocasionado que se produjera, manifestando BSC-CNS cuanto sigue:

Para poder dar respuesta a este apartado, consideramos necesario contextualizar este tratamiento de datos; en especial los relativos al Hospital 12 de octubre y Hospital Clínic de Barcelona dado que la información que se

adjunta en la denuncia hace referencia a estos hospitales. Estos convenios se firmaron durante la pandemia del COVID y tenían como objetivo dar respuesta inmediata a la emergencia sanitaria que se estaba viviendo en España durante ese momento. La colaboración científica y técnica con dichos hospitales tenía como finalidad el procesamiento automático de documentos clínicos relacionados con pacientes del COVID19, en aras de crear modelos de datos estructurados para mejorar el análisis, la gestión, el tratamiento y la predicción del desenlace y evolución durante el proceso asistencial e ingreso hospitalario de dichos pacientes. De esta manera, lo que se perseguía era disponer de información sobre las características de esta enfermedad con el fin de crear un modelo predictivo, frenar esa enfermedad y por consiguiente los contagios. Por tanto, para el BSC-CNS no era relevante, ni de su interés obtener datos identificativos de los pacientes en la medida que no había ningún tratamiento sobre pacientes.

Por este motivo y conforme a la normativa de protección de datos y al principio de intervención mínima, BSC-CNS recibía los textos clínicos anonimizados y a los que el BSC-CNS aplicaba técnicas procesamiento de lenguaje natural para permitir la extracción automática y normalización de conceptos clínicos. No se requería ningún dato personal para llevar a cabo este proceso.

En ambos convenios el BSC-CNS tenía como objeto: a) realizar actividades de minería de datos y b) desarrollo de modelos predictivos y las entidades hospitalarias tenían como obligación enviar los datos anonimizados en origen al BSC-CNS. El objetivo del BSC-CNS era recibir datos anonimizados en origen puesto que para desempeñar las tareas del convenio no se precisaba ni identificar, ni tratar datos personales. Seguidamente se describen las obligaciones y la metodología de trabajo empleada por el BSC-CNS en el marco de estos convenios:

a) Actividades de minería de datos

Las actividades de minería de datos tenían como objetivo la de facilitar la extracción de información de historias clínicas de pacientes de COVID 19 anonimizadas de relevancia para el desarrollo de modelos predictivos. La minería de textos es el proceso por el cual se extrae información relevante de textos y permite convertir la información explícita no estructurada en información implícita estructurada. En nuestro caso, la información de interés en los textos clínicos eran los antecedentes, la información sobre posibles viajes a países terceros y hábitos de salud.

Para las tareas de minería de datos se entrenan y se publican modelos del lenguaje biomédicos como base para el desarrollo de aplicaciones en el Procesamiento del Lenguaje Natural (PLN) clínico. Esto implica que generamos y publicamos modelos de lenguaje biomédico en español utilizando: a) grandes corpus biomédicos con un total de 1.1B palabras obtenidos a partir de descargas de la web, publicaciones científicas, patentes del ámbito, etc. y b) un corpus de historias clínicas electrónicas de 95 M de palabras.

Tal y como se puede apreciar, esta metodología de trabajo se publicó en mayo de 2022 en el artículo "Pretrained Biomedical Language Models for Clinical NLP in Spanish in Proceedings of the 21st Workshop on Biomedical Language Processing (pp. 193-199 https://aclanthology.org/2022.bionlp-1.19/). Según se reporta en el artículo, para cada recurso biomédico aplicamos una herramienta de limpieza con operaciones personalizadas diseñadas para leer datos en distintos formatos, dividirlos en frases, detectar el lenguaje, eliminar las frases mal formadas y repetidas, etc. Por el contrario, a las historias clínicas no les aplicamos ningún pre-proceso y se añadieron tal cual al corpus de entrenamiento porque tal pre-proceso no es necesario. Esto se explica por la diferente naturaleza de los datos: en el primer caso, los referentes al apartado (a), los datos tienen formatos diversos (HTML, PDF, RTF) y hay que extraer el texto, mientras que el segundo caso, los referentes al apartado (b) "historias clínicas"; el texto ya está en formato plano (txt) y no hay que aplicar ningún proceso de extracción o preproceso.

La no manipulación de los textos clínicos explica que no se detectara datos personales ocasionales en los textos, en cuanto no se precisaba acceder, visionar y analizar uno a uno los datos. En consonancia con lo indicado anteriormente, para las historias clínicas al venir con formato (txt) no se precisaba abrir dicho formato y no se requirió el pre-proceso.

En resumen, las tareas de modelado de lenguaje no requerían el análisis exhaustivo y manipulación de los textos, ya que éstos se agregaban al conjunto de datos de entrenamiento. Así pues, los investigadores del BSC-CNS no precisaron abrir, acceder o tratar los textos para desempeñar los trabajos relativos a la minería de datos referentes a las historias clínicas; que, en cualquier caso, ya debían venir anonimizadas en origen por las entidades hospitalarias, que eran las responsables de los datos. Las historias clínicas al venir con formato (txt) se añadieron al corpus de entrenamiento, sin pre-procesamiento.

Es por ello que, de acuerdo a las evidencias aportadas por la AEPD, solo es posible que esos datos personales aparecieran al ejecutar comandos de búsqueda "ad hoc"; es decir búsquedas con patrones especialmente pensados para localizar potenciales datos personales. La finalidad de esta búsqueda la desconocemos puesto que para realizar las funciones y tareas asignadas en el BSC-CNS no se requiere acceder a estos datos personales, como se ha apuntado anteriormente. Dicho de otra manera, para realizar tareas de modelado de lenguaje no se precisa acceder a datos personales.

BSC-CNS continúa describiendo las obligaciones y la metodología de trabajo empleada por el BSC-CNS en el marco de estos convenios:

Desarrollo de modelos predictivos

Los modelos predictivos en salud se entrenan con datos correspondientes a grandes conjuntos de pacientes para ayudar a identificar a los pacientes con riesgo de padecer determinadas afecciones. En nuestro caso, el objetivo era la predicción temprana de riesgo de muerte y necesidad de ventilación asistida.

Para la generación de modelos predictivos a partir de datos estructurados, únicamente contamos con los datos del Hospital 12 de octubre. Por motivos técnicos internos al hospital, el Hospital Clinic de Barcelona no pudo hacer la exportación equivalente y nunca dispusimos de sus datos. El objetivo era aumentar los datos que la entidad HM Hospitales había puesto a disposición de la comunidad científica para la investigación sobre el COVID 19 (https://www.hmhospitales.com/coronavirus/covid-data-save-lives). Los datos de HM Hospitales recogían información sobre diagnósticos, tratamientos, ingresos, pasos por UCI, pruebas diagnósticas, resultados de laboratorio, alta o deceso, entre otros registros. Con la apertura de este dataset, HM Hospitales quería ofrecer a la comunidad médica y científica datos clínicos con los que poder obtener modelos predictivos de evolución, modelos epidemiológicos e información sobre la respuesta a los diversos tratamientos aplicados.

Según el artículo publicado en diciembre del 2022 y en el que se expone la metodología de trabajo sobre el desarrollo de modelos predictivos de estos convenios "Predicting the evolution of COVID-19 mortality risk: A Recurrent Neural Network approach. Computer Methods and Programs in Biomedicine Update, 3, 100089" (https://doi.org/10.1016/j.cmpbup.2022.100089), el objetivo fue el desarrollo de un modelo basado en Redes Neuronales Recurrentes para predecir el desenlace de pacientes ingresados con COVID-19. Se utilizaron dos conjuntos de datos: los del HM Hospitales y del H12O respectivamente. En ambos casos, los datos contenían variables estáticas que se mantienen constantes a lo largo de los ingresos, como el sexo o la edad, y variables dinámicas que se miden en distintos momentos de la estancia hospitalaria, como la medicación, los análisis de laboratorio y las constantes vitales. Los datos no contenían datos personales y en ningún caso el experimento trató con datos de campos textuales libres que pudieran contener algún tipo de dato personal no deseable de forma incidental en estos campos, ya que como se ha expuesto, estos datos no son precisos y necesarios para llevar a cabo estos proyectos. Los campos utilizados para la realización del trabajo están debidamente reportados en el artículo citado y eran los únicos a tener en cuenta por nuestros investigadores de acuerdo a las funciones asignadas en los convenios.

De nuevo, las evidencias aportadas sobre posibles observaciones de datos personales no deseables son el resultado de ejecutar comandos de búsquedas "ad hoc" sobre el conjunto de datos, tal y como se observa en el documento aportado como prueba en la denuncia presentada. Si dichos datos personales – de carácter residual e incidental – han aparecido es porque de forma intencionada se ha provocado dicha búsqueda. Así pues, la posibilidad de encontrar datos personales imprevistos no deseables es prácticamente nula debido al escaso número de registros con campo de texto libre; especialmente teniendo en cuenta que, para desarrollar modelos predictivos solo se precisan las características y la evolución de la enfermedad y en ningún caso, datos personales.

Cabe reiterar que los investigadores identificados en los proyectos de estos convenios no manifestaron a los responsables técnicos del BSC-CNS la

existencia de datos personales en la documentación. Tal y como ya se indicó en el anterior requerimiento de información, si un investigador del BSC-CNS hubiera detectado la existencia de datos personales, éste tenía la obligación de comunicar dicha circunstancia a los responsables técnicos del BSC-CNS para que el BSC-CNS pudiera adoptar medidas técnicas y jurídicas necesarias y notificar al Responsable del Tratamiento. En el caso que el investigador no hubiese practicado dicha notificación, este investigador estaría incumpliendo su contrato laboral, la política de usuarios del BSC-CNS y su obligación de obrar de buena fe y con la debida diligencia profesional que se exige a todo trabajador en el tratamiento de datos personales.

Por último, poner de relieve que hasta el envío de esta documentación que se adjunta en su denuncia, desconocíamos la existencia de estos datos personales puesto que ningún usuario reportó la existencia de los mismos. En ningún caso, los usuarios del BSC-CNS accedieron o trataron datos personales para realizar las funciones asignadas en los convenios puesto que no era preciso acceder a éstos para realizar las actividades de minería de datos o para desarrollar modelos predictivos. Y en el caso que hubiere habido algún usuario que hubiera encontrado esos datos personales es porque tuvo la mala fe y motivación de buscarlos de forma deliberada y localizada con el único propósito de dañar a esta entidad.

Se debe indicar que esta documentación con datos personales no se hallaba indexada ni se efectuó tratamiento alguno sobre ella puesto que se desconocía su existencia. Sin embargo y a pesar de ello, esta entidad aplicó todas las medidas de seguridad para garantizar la correcta utilización de esta información, según se ha acreditado ante esta Agencia con la documentación aportada en este escrito y anteriores.

En suma y para finalizar, se quiere dejar constancia de lo siguiente: a) que el BSC-CNS cumplió con las obligaciones recogidas en el convenio (realizar tareas de minería de datos y desarrollo de modelos predictivos, b) que el BSC-CNS no era el responsable de llevar a cabo la anonimización, c) que los hospitales tenían que enviar los datos debidamente anonimizados en origen con arreglo a sus obligaciones en los convenios, y d) que los proyectos COVID de los convenios contaron con la correspondiente aprobación del Comité de Ética de Investigación Clínica y en el que se evaluó la anonimización de los datos personales.

En el tercer requerimiento efectuado por esta inspección se solicitó a BSC-CNS que se aportara el contrato de encargo de tratamiento de datos con IBM en caso de que esta entidad tuviera que tratar datos personales en el seno de este contrato. En la respuesta BSC-CNS 3, se aporta cláusula de tratamiento de datos personales del Pliego de referencia y se manifiesta que:

En este sentido, no se aporta el contrato de encargado de tratamiento porque no hubo tratamiento de datos personales. El personal de IBM no dispone de acceso a los datos almacenados en el sistema GPFS, ya que no lo necesita para las tareas de mantenimiento contratadas y de ahí que no hiciera falta añadirlo en el contrato.

Continuando con las medidas de seguridad BSC-CNS expresa que:

La política de seguridad de BSC-CNS pretende garantizar la seguridad en su máximo nivel incorporando aquellos elementos que permitan desarrollar los encargos recibidos sin menoscabo de la seguridad y confidencialidad. Por ello, se sustenta en que tan sólo determinados perfiles del departamento de operaciones pueden ostentar el cargo de administradores para otorgar o modificar permisos de los usuarios. ***(…).***

Las medidas de seguridad que se implementan en los proyectos se analizan específicamente caso por caso, desde la visión de seguridad informática y confidencialidad de la información que se trate. En este caso, por las características de las investigaciones y tareas a realizar en estos proyectos, los usuarios tienen que estar autorizados y se les dota de permisos de acceso, lectura y modificación, que permiten la descarga y tratamiento de esa información para poder trabajar en los Proyectos. No se accede a la información de los proyectos sino se está debidamente y expresamente autorizado. También, los usuarios que acceden a la información que se aloja en esos servidores deben de tratarla, custodiarla y almacenarla de acuerdo a las obligaciones que se detallan en la Política de Responsabilidades de los Usuarios del BSC-CNS y firman asimismo una declaración de cumplimiento de esas obligaciones antes de acceder a la información alojada en los servidores. Se acompaña como documento cinco dicha política de responsabilidades.

En este punto es necesario recalcar en materia de protección de datos personales, que los usuarios de datos del BSC-CNS no tratan ni visionan datos personales en estos proyectos. En el supuesto caso que hubiera una posible re-identificación deben de comunicarlo a la mayor brevedad posible al responsable del Proyecto, quién éste dará aviso inmediato a las personas encargadas en materia de protección de datos de la organización. El objetivo de esa comunicación sería la de analizar los términos de los convenios suscritos y detener de forma preventiva ese tratamiento de datos hasta la comprobación del error detectado que ha implicado que el responsable de tratamiento de esos datos incorporara esos datos no anonimizados y la adecuación, en su caso, de las medidas legales y técnicas necesarias para poder continuar con la ejecución del proyecto.

Por otro lado, los empleados del BSC-CNS son informados de las normas de cumplimiento que deben atender en el tratamiento de datos, de la normativa de protección de datos y de los aspectos que les afecta como usuarios de datos por cuenta de BSC-CNS a la firma del contrato laboral. Se adjunta, como documento número seis la cláusula de cumplimiento en el contrato laboral con relación a dichas materias.

Por lo que respecta a la adecuación y adopción del Esquema Nacional de Seguridad (ENS), hay que indicar que en la actualidad el BSC-CNS se halla inmerso en este procedimiento. A tal fin, el BSC-CNS ha contratado a través de Rediris a las consultorías especializadas "Procesia" e "Inycom" para la prestación de este servicio, cuyo objetivo es garantizar una adecuada

adaptación a las exigidas del ENS. Por ello se ha adjuntado como DOC CINCO, el documento actual que establece las normas de uso de todo sistema de información, teniendo en cuenta que la fecha prevista para la finalización e implementación de la ENS será a finales de 2023.

Asimismo, dejar constancia que tanto el Delegado de protección de datos, así como una consultoría externa especializada en normativa de protección de datos, SEGURDADES, S.L., ha participado en la revisión y análisis de estos convenios a fin de comprobar el correcto cumplimiento de esta normativa de protección de datos y velar por los principios reglamentarios exigidos.

Requerimiento a Hospital Clinic de Barcelona (HCB):

En la respuesta HCB informa que:

La colaboración llevada a cabo entre el Hospital y el BSC-CNS se contextualiza por el desarrollo del proyecto de investigación "Desarrollo de modelos predictivos para el soporte a la toma de decisiones clínicas en el ámbito de la COVID", (en adelante, el "Proyecto") el cual dispone de dictamen favorable de aprobación por parte del Comité de Ética y de Investigación Clínica del HCB, con código HCB/2020/0824 adjuntándose el mismo como Documento nº1.

Dicha colaboración fue además regulada entre ambas partes mediante el Convenio de Colaboración entre el Barcelona Supercomputing Center – Centro Nacional de Supercomputación y el HCB, de fecha de 29 de mayo de 2020, en el que se incluye como Anexo I al mismo el Contrato de Tratamiento de Datos de carácter personal. Se adjunta al presente escrito dicho Convenio como Documento nº2.

El protocolo del Proyecto, añadido al presente escrito como Documento nº3, describe de manera exhaustiva y detallada el objetivo y diseño del mismo, así como de los procesos de preparación y evaluación de los datos. Se realiza un profundo análisis del método de extracción de estos datos así como del traspaso de los mismos para conseguir el objetivo del Proyecto. Se detalla entonces los diferentes equipos de investigación que tendrán acceso a los datos, los permisos y funciones asignadas, así como incluyéndose el proceso de anonimización especialmente robusto diseñado en base a:

- Plan de Impulso de las Tecnologías del Lenguaje, de Mayo de 2019, del Gobierno de España, adjunto como Documento nº4
- Guías de anotación de información de salud protegida de Octubre de 2018 del Gobierno de España, adjunto como Documento nº5
- Protocolo de anonimización de Informes para el Corpus COVID-19 anonimizado HCB-BSC de Julio de 2022, siendo un documento de trabajo conjunto para la fase de revisión y validación clínica de la anonimización, adjunto como Documento nº6
- Adendas a las Guías Meddocan para anonimización HCB, con conceptos identificados por los clínicos del HCB y añadidos a las guías consultadas, adjuntándose como Documento nº7

En el protocolo del Proyecto referenciado se describe, así mismo, en su punto 9.3, el análisis de la detección de los riesgos en relación con la posible identificación de los pacientes en las diferentes etapas del procesamiento de los datos, identificándose las medidas de contención de los mismos, análisis realizado por el equipo investigador del Proyecto.

De las diferentes reuniones mantenidas con el Director de Informática Médica, se ha elaborado por el Investigador Principal del Proyecto el informe adjunto como Documento nº8 titulado "Gestión de la información documental entre el Hospital Clínic de Barcelona ("HCB") y el Barcelona Supercomputing Center ("BSC") para el desarrollo del proyecto de Investigación sobre COVID19 articulado por el convenio de colaboración entre el BSC y el Hospital Clínic firmado el 29 de mayo de 2020, titulado "Desarrollo de modelos predictivos para el soporte a la toma de decisiones clínicas en el ámbito de la COVID19" ("Proyecto COVID19") (en adelante, el "Informe"). En dicho informe se describe detalladamente el tipo de análisis de datos que implica el Proyecto, el método de transmisión de los mismos y los métodos de anonimización y seudonimización seguidos, incluyéndose pantallazos de los procesos, así como determinaciones de los circuitos y funciones de cada parte.

En el contexto del Proyecto, el BSC-CNS presta un servicio al HCB de análisis de riesgos de re-identificación y anonimización completa de los datos a tratar en el Proyecto, en base al Contrato de Encargado del Tratamiento formalizado el 29 de mayo de 2020. En el Informe se expone de manera exhaustiva el método de transmisión de los datos, el equipo autorizado a acceder a éstos, así como el tipo de información que se requiere, el procesamiento mediante tecnología PLN que proporciona el BSC-CNS, así como el resultado de los datos una vez finalizado el servicio en cuestión. La información completa de los datos se encuentra almacenada en el gestor asistencial del HCB, al que solamente tienen acceso los profesionales del HCB mediante un sistema de usuario nominal, clave de acceso secreta personal intransferible y autogestionada y los permisos de acceso correspondientes en el marco de las actuaciones necesarias para alcanzar los objetivos del Proyecto. Por lo tanto, ningún personal del BSC-CNS dispone de usuario para acceder al gestor asistencial indicado, siendo que la gestión de autorización de usuarios al gestor asistencial del Hospital la realiza Informática Médica de acuerdo con un procedimiento normalizado de registro, autorización y gestión de la identidad digital del usuario, de acuerdo con el documento Procediment SAP Assistencial Externs v14, adjuntado como Documento nº9.

Por lo que respecta a los Informes del Delegado de Protección de Datos en relación con estos tratamientos, HCB informa:

[E]l año 2020, cuando fue aprobado el Proyecto por el Comité de Ética y de Investigación Clínica del HCB, éste estaba integrado, entre otros, por un miembro experto en conocimientos suficientes del Reglamento (UE) 2016/679, delegado del Delegado de Protección de Datos, de acuerdo con la Disposición adicional decimoséptima 2h) de la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. Así entonces, se participó y tomó parte de todas las valoraciones realizadas por el Comité en cuanto al análisis de proyectos de investigación que comportasen el tratamiento de datos personales sobre

los que el Hospital es responsable, teniéndose en cuenta en el análisis del proyecto su intervención y valoración en la deliberación colegiada.

Requerimiento al INSTITUTO DE INVESTIGACIÓN SANITARIA ILLES BALEARS (IdIsBa)

Se recibe contestación en el sentido siguiente:

- Que en primer lugar, es necesario poner en conocimiento que objeto de esta colaboración en ningún momento se han remitido datos de carácter personal al Consorcio Barcelona Supercomputing Center – Centro Nacional de Supercomputación, en lo sucesivo BSC-CNS, habiéndose producido el envío con datos anonimizados.

- Que la necesidad de enviar datos anónimos se recoge con especial detalle en los Convenios firmados entre BSC-CNS, IDISBA y el Hospital universitario. (…)

- Que con motivo de los citados convenios, exclusivamente se ha proporcionado información de aproximadamente 1.000 informes de alta todos ellos debidamente anonimizados.

- Que debido a que se procesarán datos anonimizados, las Actividades de tratamiento a realizar no requieren regirse por el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, (en adelante, "RGPD") y la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

- Que es por la circunstancia de no comunicar datos personales, y unido al hecho de que en ningún caso el tratamiento de datos a anonimizar era masivo, es que no se consideró necesaria la elaboración de una evaluación de impacto de privacidad en virtud de lo expuesto por el artículo 35.4 y siempre desde la posición del cesionario de información anonimizada.

- En documento anexo, se entrega el documento descriptivo con los procesos de anonimización de informes y registros extraídos de la historia clínica electrónica, así como del proceso de remisión de los datos a la entidad Consorcio Barcelona Supercomputing Center.

El proceso de anonimización se describe como sigue:

1.- Proceso de selección de informes y anonimización: Tras seleccionar mediante consulta a base de datos los documentos con las características requeridas se procede a su extracción en formato XML. En la base de datos se archiva exclusivamente el contenido del informe y no su cabecera que es la que contiene datos sensibles que permiten la identificación del paciente. Los informes generados para impresión o para publicación en PDF fusionan el cuerpo del informe (archivo XML) con una cabecera que se construye extrayendo los datos de filiación del paciente y del episodio. Por tanto es posible hacer una extracción de informes en formato XML que no contienen ningún dato de filiación ni número de identificación que permita relacionarlo con

el paciente al que pertenece. Este conjunto de documentos anonimizados fue revisado manualmente con dos finalidades: verificar que correspondían con el tipo de información requerida para el proyecto de investigación y asegurar que no incluía datos de identificativos del paciente de acuerdo con lo requerido por Protección de Datos en aquel momento. Los ficheros tenían numeración aleatoria ya que no se requería para la investigación la identificación del caso.

2.- Proceso de envío de información: el conjunto de informes se incluyó en una carpeta que se compactó y encriptó y se subió a la plataforma que para tal uso habilitó el grupo de trabajo ictusnet para su posterior uso.

Requerimiento a la FUNDACIÓN PÚBLICA ANDALUZA PARA LA GESTIÓN DE LA INVESTIGACIÓN EN SALUD DE SEVILLA (FISEVI):

En fecha 9 de enero de 2023 se realizó requerimiento de información mediante notificación electrónica entregada al día siguiente. Dicho requerimiento se reiteró en fecha 28 de febrero de 2023, mediante notificación electrónica entregada el día 6 de marzo de 2023. No se ha obtenido contestación a los requerimientos, aunque, tal y como ha aclarado el BSC-CNS, a pesar de la firma inicial del convenio, FISEVI no llegó a enviar datos al BSC-CNS. HCB mantenía informado a FISEVI de sus progresos con el modelo común de datos, pero en ningún caso el BSC-CNS recibió datos del Hospital FISEVI.

Requerimiento al HOSPITAL UNIVERSITARIO DOCE DE OCTUBRE (H12O):

Se recibe la información siguiente:

[E]ste Comité Delegado de Protección de Datos (en adelante CDPD), en su calidad de interlocutor de la Consejería de Sanidad con esta Agencia, acorde a lo establecido en el artículo 37.2 de la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y Garantía de los Derechos Digitales (LOPDGDD), comparece y formula en tiempo y forma las siguientes consideraciones en relación a lo solicitado:

1. Descripción detallada con documentación acreditativa (pantallazos, documentos, protocolos…) del proceso de remisión de los datos al BSC-CNS.

"El envío de los datos se llevó a cabo a través de un acceso SFTP suministrado al equipo técnico del BSC para la recepción de los datos." Así se señala en el documento que nos ha sido remitido por el H120, según consta en el documento "Proceso de seudonimización de datos para investigación en COVID-19", y que se adjunta como Anexo 1.

(…)

Adicionalmente adjuntamos como Anexo 2 el documento de la Gerencia del responsable del tratamiento, H12O, en virtud del cual (i) nos informan de los hechos en relación con el requerimiento que les fuera remitido por Uds. y que ahora contestamos, y (ii) facilitan la documentación que desde esta Delegación les aportamos.

2. Descripción y evidencias del proceso de anonimización o, en su caso seudonimización de los datos sanitarios facilitados a BSC-CNS.

*En el referido Anexo 1**,** adjunto, se establece el proceso de seudonimización de los datos facilitados a BSC-CNS para investigación en COVID- 19. Damos por reproducido lo que en el mismo se señala e ilustra.*

Adicionalmente se adjunta el documento denominado "Modelo de datos-COVID19" en el que se describe el modelo de información establecido para facilitar los datos seudonimizados, con la descripción de los campos que contenían las distintas tablas que fueron diseñadas para el proyecto de referencia como Anexo 3.

3. Análisis de riesgos y evaluaciones de impacto llevadas a cabo con ocasión de estos tratamientos y descripción de la metodología utilizada para ello.

El proyecto fue trasladado para su evaluación al Comité de Ética del Centro, siguiendo con los procedimientos internos, y de conformidad con lo establecido en la DA 17ª letra h) de la LOPDGDD, el cual analizó el mismo el 14/01/2021, obtenido un resultado como el que se transcribe: "Conclusión: Por tanto, considera que no procede evaluación". Se adjunta dicho dictamen como Anexo 4.

Adicionalmente indicar en cuanto a metodologías que se emplean en la Consejería de Sanidad de la Comunidad de Madrid y por parte de todos los responsables que la integran:

- *para gestión y análisis de riesgos de los sistemas de información, MAGERIT empleando la herramienta PILAR.*

- *para las evaluaciones de impacto de protección de datos se sigue la metodología de la propia AEPD, empleando al efecto distintas herramientas, algunas elaboradas específicamente para la CSCM.*

4. Informes del Delegado de Protección de Datos en relación con estos tratamientos.

El H12O, como responsable del tratamiento, no recabó un informe específico y/ o concreto a este CDPD en relación con la participación en el proyecto, si bien queda constatado que actuó con diligencia en el sentido de haber suscrito el correspondiente convenio en el que se establecían las medidas que se debían adoptar para llevar a cabo el mismo, y solicitó el preceptivo informe del CEIM.

Conforme ha recibido el requerimiento de la AEPD que ahora contestamos, ha acudido a esta Delegación, informándonos de todo ello, así el Anexo 2 adjunto.

Proceso de seudonimización de datos para investigación en COVID-19 se describe como sigue:

Introducción

Este documento describe el proceso de seudonimización de los datos de COVID-19 para ser empleados en propósitos adicionales a la asistencia individual del paciente.

Datos de pacientes

La Historia Clínica Electrónica (HCE) del Hospital Universitario 12 de Octubre cuenta con diferentes identificadores de carácter personal a nivel de paciente (NHC, CIPA, DNI, nombre y apellidos, domicilio etc.). Es por ello que en los diferentes procesos de extracción de datos para uso secundario (investigación, entrenamiento de modelos, evaluación, etc.) se lleva a cabo un proceso de seudonimización de la información, eliminando todos los datos sensibles y vinculando cada registro de un mismo paciente a un código artificial y único, no vinculado al individuo, "ID_SEUDO".
Esta correspondencia entre código seudonimizados-códigos sensibles se lleva a cabo en la infraestructura del Hospital en la propia base de datos de la HCE, con todos los mecanismos de seguridad que esto implica, no siendo accesible esta información por los consumidores de los datos del caso de uso que los demanda. A continuación, se muestra una captura de pantalla de esta tabla de correspondencia en la HCE, omitiendo datos sensibles:


Result | Messages

ID_SEUDO	CODIGO	NUMEROHC	DEFINITIVO	LOCALIZACION	ESTADO	NOMBRE	APELLIDO1	APELLIDO2	FNOMBRE	FAPELLIDO1	FAPELLIDO2	SEXO	POBLACION	PROVINCIA	AUTONOMIA	PAIS	DOMICILIO
301785																	
301672																	
302514																	
302448																	
283388																	
283810																	
302798																	
303224																	
284569																	
304798																	
305140																	
304321																	
305478																	
300122																	
282385																	
284046																	
285203																	

Tabla de correspondencia para la seudonimización de la HCE

Así mismo, en el documento "ModeloDatos-COVID-19" se puede ver la estructura y contenido de los ficheros exportados, todos ellos con el único identificador de paciente "ID_SEUDO", y sin ningún otro dato identificativo o personal del paciente.

Para el envío de datos no estructurados se implantó un proceso automatizado que extrae el cuerpo del informe (en este caso de uso aplica a Radiología) sin ningún metadato identificativo del paciente o el profesional. La identificación se realiza asignando a cada fichero de texto como nombre un código interno del sistema de información donde han sido generados.

Datos de profesionales

No se extraen datos personales de profesionales en los procesos de obtención de datos para uso secundario. En proyectos específicos que así lo requieran (no es el caso del proyecto con BSC), se identifica cada registro con un código interno que indica el usuario que lo creó en la base de datos del sistema.

Envío de datos

*El envío de los datos se llevó a cabo a través de un **(...)** suministrado al equipo técnico del BSC para la recepción de los datos.*

Requerimiento al denunciante

En fecha 16 de febrero de 2023, se requirió al denunciante documentación acreditativa de los hechos denunciados, dado que en la denuncia se facilitaban unos enlaces (******ENLACE.1*** y ******ENLACE.2***) a los que no era posible acceder.

Con fecha 21 de febrero de 2023, se dio cumplimiento al requerimiento por parte del denunciante, que aporta cinco páginas donde aparecen algunos datos personales que no parecen de los propios pacientes, sino de sus familiares. Constan además datos de los médicos que los atienden.

Aporta además un documento de algo más de 70MB de información donde se constata que la información está anonimizada. Se han hecho numerosas búsquedas en este documento y se ha localizado, dentro de un grandísimo volumen de información anonimizada, datos de personal médico y algún dato personal aislado no detectado por los procesos de anonimización ejecutados de forma automática para los campos de texto e informes médicos.

Dado que la información facilitada se consideró insuficiente, se reiteró la solicitud de documentación en fecha 14 de marzo de 2023.

El día19 de marzo de 2023, el denunciante solicita una ampliación de plazo que le fue concedida mediante notificación electrónica entregada el mismo día 21 de marzo de 2023, sin que se haya dado cumplimiento al requerimiento.

FUNDAMENTOS DE DERECHO

I
Competencia

De acuerdo con las funciones que el artículo 57.1 a), f) y h) del Reglamento (UE) 2016/679 (Reglamento general de protección de datos, en adelante RGPD) confiere a cada autoridad de control y según lo dispuesto en los artículos 47 y 48.1 de la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (en lo sucesivo LOPDGDD), es competente para resolver estas actuaciones de investigación la Directora de la Agencia Española de Protección de Datos.

Asimismo, el artículo 63.2 de la LOPDGDD determina que: "Los procedimientos tramitados por la Agencia Española de Protección de Datos se regirán por lo dispuesto en el Reglamento (UE) 2016/679, en la presente ley orgánica, por las disposiciones reglamentarias dictadas en su desarrollo y, en cuanto no las contradigan, con carácter subsidiario, por las normas generales sobre los procedimientos administrativos."

II
Seguridad del tratamiento

El artículo 32 del RGPD estipula lo siguiente:

"1. Teniendo en cuenta el estado de la técnica, los costes de aplicación, y la naturaleza, el alcance, el contexto y los fines del tratamiento, así como riesgos de probabilidad y gravedad variables para los derechos y libertades de las personas físicas, el responsable y el encargado del tratamiento aplicarán medidas técnicas y organizativas apropiadas para garantizar un nivel de seguridad adecuado al riesgo, que en su caso incluya, entre otros:

a) la seudonimización y el cifrado de datos personales;

b) la capacidad de garantizar la confidencialidad, integridad, disponibilidad y resiliencia permanentes de los sistemas y servicios de tratamiento;

c) la capacidad de restaurar la disponibilidad y el acceso a los datos personales de forma rápida en caso de incidente físico o técnico;

d) un proceso de verificación, evaluación y valoración regulares de la eficacia de las medidas técnicas y organizativas para garantizar la seguridad del tratamiento.

2. Al evaluar la adecuación del nivel de seguridad se tendrán particularmente en cuenta los riesgos que presente el tratamiento de datos, en particular como consecuencia de la destrucción, pérdida o alteración accidental o ilícita de datos personales transmitidos, conservados o tratados de otra forma, o la comunicación o acceso no autorizados a dichos datos.

3. La adhesión a un código de conducta aprobado a tenor del artículo 40 o a un mecanismo de certificación aprobado a tenor del artículo 42 podrá servir de elemento para demostrar el cumplimiento de los requisitos establecidos en el apartado 1 del presente artículo.

4. El responsable y el encargado del tratamiento tomarán medidas para garantizar que cualquier persona que actúe bajo la autoridad del responsable o del encargado y tenga acceso a datos personales solo pueda tratar dichos datos siguiendo instrucciones del responsable, salvo que esté obligada a ello en virtud del Derecho de la Unión o de los Estados miembros."

III
Principios relativos al tratamiento

La letra f) del artículo 5.1 del RGPD propugna:

"1. Los datos personales serán:

(…)

f) tratados de tal manera que se garantice una seguridad adecuada de los datos personales, incluida la protección contra el tratamiento no autorizado o ilícito y contra su pérdida, destrucción o daño accidental, mediante la aplicación de medidas técnicas u organizativas apropiadas («integridad y confidencialidad»)."

IV
Hechos denunciados e investigados

Los hechos denunciados se concretan en que el BSC-CNS tiene acceso a datos personales de cuyo tratamiento serían responsables distintos hospitales. Los datos accedidos serían: DNI, números de teléfono, número de colegiado, historia clínica completa estructurada y no estructurada; no existiendo un control de acceso adecuado a dichos datos, pudiendo cualquier trabajador (incluidos los becarios) acceder a ellos y descargárselos.

Las actuaciones de investigación se han centrado en analizar el tratamiento de datos que BSC-CSN realiza como encargado de tratamiento. No se ha investigado el tratamiento de datos que se efectúa por parte de los Hospitales/Fundaciones que son responsables de tratamiento. A estos últimos se les ha solicitado la descripción y las evidencias de los procesos de anonimización/seudonimización, así como las transferencias de los datos anonimizados o seudonimizados al BSC-CNS en el marco de los convenios que han sido firmados entre los responsables (hospitales/fundaciones) y el encargado (BSC-CNS).

Las actuaciones de investigación se han orientado fundamentalmente a analizar si el tratamiento que BSC-CNS realiza permite una identificación completa de los titulares de los datos tratados, así como a examinar las medidas de seguridad implementadas por BSC-CNS para controlar los accesos del personal investigador a tales datos.

En las actuaciones de investigación no ha podido acreditarse que resulte posible, con la actividad llevada a cabo por BSC-CNS como encargado de tratamiento, identificar a las personas afectadas por el tratamiento.

Analizada la documentación aportada con respecto a los procesos de anonimización/seudonimización, con las capturas aportadas se constata una primera capa de seudonimización de los datos demográficos e identificativos, en esta primera capa se eliminan los datos de los campos de la cabecera del documento pero permanecen los campos de texto en los que se ha introducido comentarios por parte de los médicos, en estos campos también pueden existir datos que permitan identificar al paciente, pero para su ofuscación se utilizan los algoritmos de Procesamiento de Lenguaje Natural (PLN) que existen en el BSC, de esto último se aporta captura de pantalla que evidencia el proceso de aplicación de esta tecnología.

Respecto de la remisión de datos de los responsables a BCN-CNS, de las evidencias remitidas por los hospitales/fundaciones (responsables del tratamiento) se puede concluir que tenían instaurados e implementados procesos de anonimización/seudonimización y, en el presente caso, tanto en las alegaciones aportadas por los responsables como por el encargado de tratamiento (BSC-CNS) se afirma que se siguieron esos procedimientos.

Respecto del tratamiento de datos que hace BSC-CNS, hay indicios sólidos que nos permiten concluir que no se ha hecho un tratamiento que permita identificar a afectados titulares de datos personales. Las evidencias aportadas por el denunciante sobre posibles observaciones de datos personales no deseables son el resultado de ejecutar comandos de búsquedas “ad hoc” sobre el conjunto de datos, pero no se acredita que estos datos fueran procesados de forma manual, sino que se introduce la información directamente en los algoritmos y sistemas de aprendizaje automático. La

no manipulación de los textos clínicos explica que no se detectaran datos personales ocasionales en los textos, en cuanto no se precisaba acceder, visionar y analizar uno a uno los datos. El objetivo del BSC-CNS, tal y como ha afirmado, era recibir datos anonimizados en origen puesto que para desempeñar las tareas del convenio no se precisaba ni identificar, ni tratar datos personales. BSC-CNS recibía los textos clínicos anonimizados y a los que el BSC-CNS aplicaba técnicas de procesamiento de lenguaje natural para permitir la extracción automática y normalización de conceptos clínicos.

Pese a que la información tratada por BSC-CNS (y remitida por parte de los hospitales) estaba anonimizada, existían las siguientes medidas de seguridad acreditadas:

- Acceso a la información por trabajadores a través de sistema de gestión de permisos y autorizaciones basado en grupos y listas de acceso.

- No se permitía a los trabajadores almacenar la información en sus ordenadores profesionales o personales, lo cual quedaba recogido por la política de uso y seguridad que debían aceptar. Ha quedado acreditado que de los siete trabadores del grupo de investigación del BSC-CNS, tres de ellos firmaron manualmente la política de uso y declaración responsable y el resto la aceptaron a través de una casilla de verificación que debían marcar de forma expresa aceptando haber leído la política de uso y declaración responsable.

- El uso de ordenadores personales era autorizado en casos excepcionales y debidamente justificados debido a la situación de pandemia del momento.

- ***(…)***.

V
Conclusión

Por lo tanto, en base a lo indicado en los párrafos anteriores, no se han encontrado evidencias que acrediten la existencia de infracción en el ámbito competencial de la Agencia Española de Protección de Datos.

De conformiddad con lo señalado, por la Directora de la Agencia Española de Protección de Datos, SE ACUERDA:

PRIMERO: PROCEDER AL ARCHIVO de las presentes actuaciones.

SEGUNDO: NOTIFICAR la presente resolución al CONSORCIO BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACIÓN.

De conformidad con lo establecido en el artículo 50 de la LOPDGDD, la presente Resolución se hará pública una vez haya sido notificada a los interesados.

Contra esta resolución, que pone fin a la vía administrativa según lo preceptuado por el art. 114.1.c) de la Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas, y de conformidad con lo

establecido en los arts. 112 y 123 de la citada Ley 39/2015, de 1 de octubre, los interesados podrán interponer, potestativamente, recurso de reposición ante la Directora de la Agencia Española de Protección de Datos en el plazo de un mes a contar desde el día siguiente a la notificación de esta resolución o directamente recurso contencioso administrativo ante la Sala de lo Contencioso-administrativo de la Audiencia Nacional, con arreglo a lo dispuesto en el artículo 25 y en el apartado 5 de la disposición adicional cuarta de la Ley 29/1998, de 13 de julio, reguladora de la Jurisdicción Contencioso-Administrativa, en el plazo de dos meses a contar desde el día siguiente a la notificación de este acto, según lo previsto en el artículo 46.1 de la referida Ley.

940-020323

Mar España Martí
Directora de la Agencia Española de Protección de Datos