
Estrategias para la gestión y la promoción del uso compartido de datos científicos sobre la salud

Rafael Aleixandre Benavent
Fernanda Peset Mancebo | Antonia Ferrer Sapena
Adolfo Alonso Arroyo | Antonio Vidal Infer
Isabel de la Riva Jiménez
Juan Carlos Valderrama Zurián
Alicia García García

Ayudas a la investigación 2012

Equipo de trabajo:

Investigador Principal:
Rafael Aleixandre Benavent

Equipo investigador:
Fernanda Peset Mancebo
Antonia Ferrer Sapena
Adolfo Alonso Arroyo
Antonio Vidal Infer
Isabel de la Riva Jiménez
Juan Carlos Valderrama Zurián
Alicia García García

Índice

	Página
1. TAREAS REALIZADAS	4
2. DIFUSIÓN DEL PROYECTO	5
3. ANEXO I	8
4. ANEXO II	29
5. ANEXO III	33
6. ANEXO IV	59
7. ANEXO V	61

Durante el año transcurrido desde que se inició el proyecto se han realizado las tareas que se detallan a continuación, cumpliendo así los objetivos perseguidos en la memoria y el cronograma planificado.

El **primer objetivo** pretendía identificar sectorialmente las prácticas y necesidades actuales en gestión de datos de investigación por parte de grupos de investigación, instituciones científicas y publicaciones, así como las websites que ofrecían sus datos abiertos en España y sean susceptibles de ser preservados al tiempo que difundidos (*curation*).

El **segundo** objetivo trataba de desarrollar recursos para la gestión del ciclo de vida digital de los datos científicos (*Digital Data Curation*) de forma sostenible desde los Organismos Públicos de Investigación, mediante la redacción de guías de buenas prácticas para diferentes colectivos: investigadores, gestores de investigación, autoridades académicas.

1. TAREAS REALIZADAS

PRIMER OBJETIVO

1.1. Paquete de trabajo de encuesta

Para la realización del cuestionario se creó un panel de expertos formado por médicos de diversos sectores (investigadores, gestores y clínicos) y gestores de la información. Este panel sentó los fundamentos en los que se tenían que basar las preguntas a incluir en el cuestionario a enviar a los investigadores.

Posteriormente se diseñó una encuesta que incluía tanto aspectos relacionados con el investigador y con la institución (modelo ciclo de vida). En relación con los investigadores, se indagaba sobre los siguientes aspectos: los datos activos que se mantienen dentro de las instituciones; cómo se almacenan, se administran, comparten y reutilizan; identificar riesgos como el mal uso, la pérdida de datos o su irrecuperabilidad; conocer cómo se han creado y las posibilidades de intercambio (si usan páginas web o repositorios). Respecto a la institución, se trata de identificar: acciones de *curation* que son aplicables a los datos de investigación; diseñar los agentes y los flujos principales en la organización en España.

Una vez diseñada la encuesta, ésta se envió en junio de 2013 por correo electrónico a investigadores de las siguientes instituciones:

- Hospital Severo Ochoa de Leganés (Madrid)
- Hospital General de Castellón
- Hospital General de Alicante
- Facultad de Medicina de Valencia

Análisis de las encuestas

Los resultados del análisis de las encuestas se presentan en el **anexo I**.

1.2. Paquete de trabajo de datos ajenos

El paquete de trabajo de datos ajenos tenía los siguientes objetivos:

- Identificación de los principales centros productores de datos científicos biomédicos en España basándonos en productividad científica, e identificación del tipo de información transversal que se utiliza en la investigación (datos estadísticos, geográfico, demográficos, etc.)
- Identificación de las websites que exponen los datos a nivel internacional (inventario de repositorios, su contenido y formas de trabajo) y nacional (inventario de URL que suministran datos reutilizables).
- Identificar URL de revistas que admitan material adicional

En el segundo informe presentado se identificaron y presentaron los principales centros productores de datos científicos biomédicos en España basándonos en su productividad científica.

En el presente informe se presentan como **anexo II**, las websites que exponen los datos a nivel internacional (inventario de repositorios) y nacional (inventario de URL que suministran datos reutilizables).

En relación con la identificación de las revistas que admiten material adicional o suplementario, se han analizado las revistas de diversas áreas, entre ellas, la pediatría, cardiología, abuso de sustancias, aparato respiratorio y ciencias de la información. Algunos de estos resultados ya han sido publicados y pueden consultarse en la bibliografía de la web del proyecto, y otros se encuentran en trámite de publicación.

SEGUNDO OBJETIVO

El segundo objetivo consistía en desarrollar recursos para la gestión del ciclo de vida digital de los datos científicos (*Digital Data Curation*) de forma sostenible desde los Organismos Públicos de Investigación, mediante la redacción de guías de buenas prácticas para diferentes colectivos: Investigadores, Gestores de investigación, Autoridades académicas.

Por tal motivo, se ha elaborado una guía de buenas prácticas destinada a médicos e investigadores, profesionales de la información y gestores de la información. La guía consta de 11 capítulos y 3 anexos y trata de responder a las preguntas que se plantean estos colectivos en lo que respecta a la gestión de los datos brutos de investigación: las prácticas actuales de los investigadores en relación con la gestión y el intercambio de datos; las necesidades de apoyo para gestionar y compartir sus

datos; el desarrollo de infraestructuras necesarias para ponerlo en práctica; la orientación práctica y la formación en este campo; las normativas necesarias para la citación de los datos y sus publicaciones matrices para aumentar la visibilidad y la accesibilidad de los datos y de la propia investigación; los aspectos éticos y legales relacionados con la protección de la privacidad y del copyright. La guía contiene también un glosario de términos, un anexo con organismos internacionales y otro que incluye recursos en Internet. Esta guía se presenta como **anexo III**.

La guía, aunque es uno de los resultados finales del proyecto, se complementa con dos anexos: un vocabulario de material suplementario (**anexo IV**) y otro que incluye las publicaciones científicas sobre los aspectos técnicos y procedimentales (**anexo V**).

Todos estos materiales y anexos se encuentran depositados en la web del proyecto: www.healthdatasea.es y son de libre acceso.

Con la entrega de estos materiales, la edición impresa y en formato electrónico de la guía y su difusión, puede afirmarse que se han cumplido todos los objetivos del proyecto.

2. DIFUSIÓN DEL PROYECTO

2.1. PÁGINA WEB

Un medio importante para la difusión de los resultados del proyecto ha sido, junto a las presentaciones en

congresos y las publicaciones en revistas, la página web (<http://healthdatasea.es/>).

La web contiene varias pestañas: Inicio (figura 1); El proyecto; El equipo (figura 2); Resultados; Agenda; Área privada; Contactar, y lleva incluida la mención de que el proyecto ha sido financiado por la Fundación Mapfre, así como el logotipo de la entidad.

La web incluye en su home page un vídeo de presentación del proyecto presentado por el investigador principal en el que se menciona la financiación de la Fundación Mapfre para su desarrollo.

2.2. ARTÍCULOS PUBLICADOS A LO LARGO DE TODA LA VIGENCIA DEL PROYECTO

Se han publicado dos artículos, otro está aceptado y un cuarto artículo se encuentra pendiente de la revisión por pares. En todos ellos se indica que han sido subvencionados por la Fundación Mapfre.

Las referencias de estos trabajos son:

A) Artículos publicados

- Luis-Millán González, Tomas Saorín, Antonia Ferrer, Rafael Aleixandre, Fernanda Peset. Gestión y difusión de datos de investigación: infraestructuras para su difusión. El profesional de la información. 2013; 22 (5): 414-423.
- Aleixandre-Benavent R; Ferrer Sapena A; Alonso-Arroyo A; Peset F. Compartir datos de investigación en cardiología. Revista Española de Cardiología 2013; <http://dx.doi.org/10.1016/j.recesp.2013.08.005>



Figura 1. Home page de la web del proyecto: <http://healthdatasea.es/>

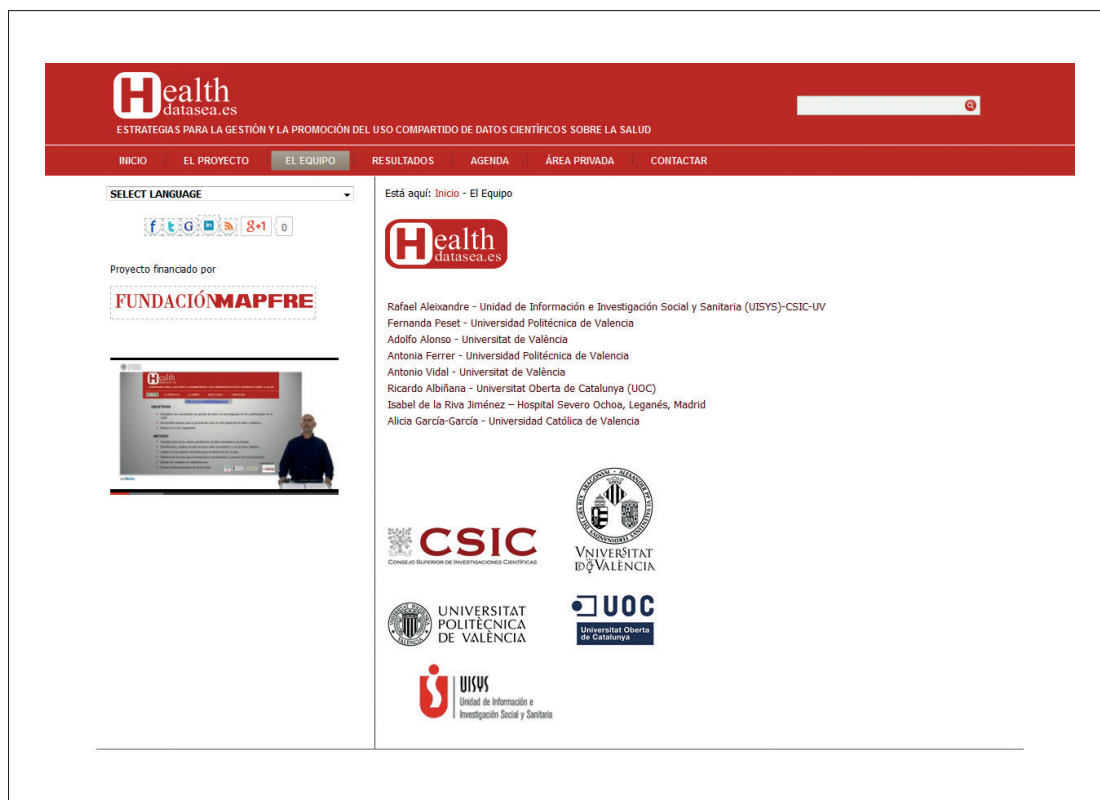


Figura 2. Página de la web del proyecto con la lista de miembros del equipo investigador

- Aleixandre-Benavent R; Ferrer-Sapena A; Alonso Arroyo A; Peset F. Sharing research data in Cardiology. *Revista Española de Cardiología* 2013; 66:1007-8.

B) Artículos en prensa

- Rafael Aleixandre Benavent (1); Antonio Vidal Infer (2); Adolfo Alonso Arroyo (2); Javier González de Dios (3); Antonia Ferrer Sapena (4); Fernanda Peset (4). Disponibilidad en abierto de los artículos y de los datos brutos de investigación en las revistas pediátricas españolas. *Anales de Pediatría*. 2013; <http://dx.doi.org/10.1016/j.anpedi.2013.11.014>.

C) Artículos en proceso de revisión

- Fernanda Peset, Luis-Millán González, Antonia Ferrer, Rafael Aleixandre, Alicia García Garrcía. Public availability of published research data in pediatrics journals. Enviado a la revista *Journal of Pediatrics*.
- Aleixandre-Benavent R; Vidal-Infer A; Alonso-Arroyo A; Valderrama Zurián JC; Bueno Cañigral F; Ferrer Sapena A. Public availability of research data in substance abuse journals. Enviado a la revista *Addictive Behaviours*. 2014;

2.3. PRESENTACIONES EN CONGRESOS

Internacionales

- Aleixandre-Benavent, R; Ribeiro-Schneider, D; Bueno-Cañigral, F.J; Ferrer Sapena, A.; Peset Mancebo, F; Garcia Garcia, A; Valderrama Zurián, J.C. Disponibilidad de los datos públicos de investigación publicados en revistas sobre drogodependencias. Congreso Internacional sobre Drogas. Brasilia. Brasil, 3-5 de mayo de 2013. <http://www.cid2013.com.br>.
- Aleixandre-Benavent, R.; Agulló Calatayud, V; Vidal Infer, A. Alonso Arroyo, A; Ferrer-Sapena, A; Peset, F. Strategies to promote scientific data sharing on substance abuse. Les Rencontres de Biarritz: Colloque Européen et International Toxicomanies Hépatites SIDA. Biarritz (France), 8-11 October 2013.
- Aleixandre-Benavent, R; Vidal Infer, A; Agulló Calatayud, V; Bueno Cañigral, FJ; Valderrama Zurián JC; Schneider DR. Public availability of research data in substance abuse journals. Les Rencontres de Biarritz: Colloque Européen et International Toxicomanies Hépatites SIDA. Biarritz (France), 8-11 October 2013.

Nacionales

- Alicia García-García; Rafael Aleixandre-Benavent; Antonia Ferrer-Sapena; Miguel Villamón; Luis-Millán González; Fernanda Peset; José Morales Aznar. ¿Cómo se gestionan los datos de investigación en ciencias de la salud? X Foro sobre Evaluación de la Educación

Superior y de la Investigación-FECIES. Granada, 25-28 de Junio, 2012.

- Aleixandre-Benavent, R.; Ferrer Sapena, A.; Peset Mancebo, F; Agulló Martínez, A; Alonso Arroyo, A; Vidal Infer, A. Estrategias para la gestión y la promoción del uso compartido de datos científicos sobre la salud. Sitges. XIII: Congreso Nacional de Documentación Médica, 12-14 de Junio de 2013. <http://www.congresodocumentacion2013.com/>
- Aleixandre-Benavent, R.; Ferrer Sapena, A.; Peset Mancebo, F; García García, A; Valderrama Zurián, J.C; Bueno-Cañigral, F.J. Disponibilidad en abierto de los datos de investigación publicados en revistas biomédicas. Sitges. XIII: Congreso Nacional de Documentación Médica, 12-14 de Junio de 2013. <http://www.congresodocumentacion2013.com/>

2.4. VÍDEO PROMOCIONAL DEL PROYECTO EN YOUTUBE

La web incluye en su home page un vídeo de presentación del proyecto y puede también visionarse en youtube en la dirección: <http://www.youtube.com/watch?v=zay3l66yYUQ>

2.5. ENTREGABLES EN LA MEMORIA FINAL

- **Anexo I.** Resultados del análisis de las encuestas
- **Anexo II.** Websites que exponen los datos a nivel internacional (inventario de repositorios) y nacional
- **Anexo III.** Guía de buenas prácticas destinada a médicos e investigadores, profesionales de la información y gestores de la información
- **Anexo IV.** Vocabulario de material suplementario
- **Anexo V.** Publicaciones científicas sobre los aspectos técnicos y procedimentales

Dada su longitud, estos anexos se entregan en ficheros aparte.

Indicencias

En el primer trimestre del año se incorporó al proyecto la Dra. Isabel de la Riva Jiménez, del Servicio de Documentación y Archivo del Hospital Severo Ochoa de Leganés (Madrid). En el segundo trimestre se incorporaron al proyecto el Dr. Juan Carlos Valderrama Zurián (Universitat de València), Alicia García García (Universidad Católica de Valencia) y el Dr. Javier González de Dios.

3. ANEXO I. RESULTADO DEL ANÁLISIS DE LOS CUESTIONARIOS SOBRE LA GESTIÓN DE LOS DATOS DE INVESTIGACIÓN

INTRODUCCIÓN

El cuestionario constaba de 40 preguntas y fue cumplimentado por 155 investigadores en ciencias de la salud pertenecientes a diversas instituciones sanitarias y especialidades. Los resultados se dividen en tres apartados. El primero consiste en datos personales; el segundo contiene preguntas relativas a la creación y utilización de los datos; en el tercero, las preguntas indagan sobre la preservación de los datos una vez terminado el proyecto.

A. DATOS PERSONALES

De los 155 profesionales que respondieron el cuestionario, 97 eran hombres (62,6%) y 58 mujeres (37,4%).

La distribución de los 155 respondientes según la función habitual desempeñada en los equipos de trabajo en los últimos años se presenta en la tabla 1. El 44,52% eran investigadores principales, el 32,26% investigadores colaboradores, el 16,13% investigadores seniors, el 4,52% asistentes o técnicos de investigación y el 1,29%

estaban realizando estudios de doctorado e eran investigadores en formación. La distribución según el número de años que llevan investigando es la siguiente (tabla 2): el 12,9% llevaba menos de 5 años; el 19,35% de 5 a 10 años; el 30,32% de 11 a 20 años; el 37,43% más de 20 años.

B. CREACIÓN Y UTILIZACIÓN DE LOS DATOS

La pregunta relativa a si en los proyectos de investigación en los que ha participado el investigador existe alguna política o plan de gestión de los datos (implícita o por escrito) (gráfico 1) ha obtenido los siguientes resultados: la respuesta más frecuente ha sido que no se dispone de ellas (22,8%); el 15,4% de los encuestados respondió sobre la existencia de políticas de “Normas de anonimización”; el 12% indicaba “Unidades donde se almacenan para el uso del grupo de investigación”; el 11,6% marcó “Niveles de autorización para acceder y modificar datos”.

Tabla 1

Señale cuál ha sido la función habitual que ha desempeñado en los equipos de trabajo de las investigaciones desarrolladas en los últimos cinco años	¿Cuántos años lleva investigando?				Total
	Menos de 5 años	De 5 a 10 años	De 11 a 20 años	Más de 20 años	
Asistente/Técnico de investigación	3 (42,86%)	2 (28,57%)	2 (28,57%)	0	7 (4,52%)
Estudiante de doctorado	1 (50,00%)	1 (50,00%)	0	0	2 (1,29%)
Investigador colaborador	12 (24,00%)	10 (20,00%)	16 (32,00%)	12 (24,00%)	50 (32,26%)
Investigador en formación	2 (100,00%)	0	0	0	2 (1,29%)
Investigador principal	1 (1,45%)	13 (18,84%)	21 (30,43%)	34 (49,28%)	69 (44,52%)
Investigador senior	1 (4,00%)	4 (16,00%)	8 (32,00%)	12 (48,00%)	25 (16,13%)
Total	20 (12,90%)	30 (19,35%)	47 (30,32%)	58 (37,42%)	155 (100,00%)

Tabla 2

¿Cuántos años lleva investigando?	Rango de edad								Total
	Menos de 30	31-35	36-40	41-45	46-50	51-55	56-60	Más de 60	
De 11 a 20 años	0	1 (2,13%)	9 (19,15%)	9 (19,15%)	9 (19,15%)	13 (27,66%)	6 (12,77%)	0	47 (30,32%)
De 5 a 10 años	3 (10,00%)	6 (20,00%)	7 (23,33%)	6 (20,00%)	3 (10,00%)	0	4 (13,33%)	1 (3,33%)	30 (19,35%)
Más de 20 años	0	0	0	3 (5,17%)	9 (15,52%)	16 (27,59%)	19 (32,76%)	11 (18,97%)	58 (37,42%)
Menos de 5 años	9 (45,00%)	4 (20,00%)	1 (5,00%)	0	1 (5,00%)	1 (5,00%)	3 (15,00%)	1 (5,00%)	20 (12,90%)
Total	12 (7,74%)	11 (7,10%)	17 (10,97%)	18 (11,61%)	22 (14,19%)	30 (19,35%)	32 (20,65%)	13 (8,39%)	155 (100,00%)

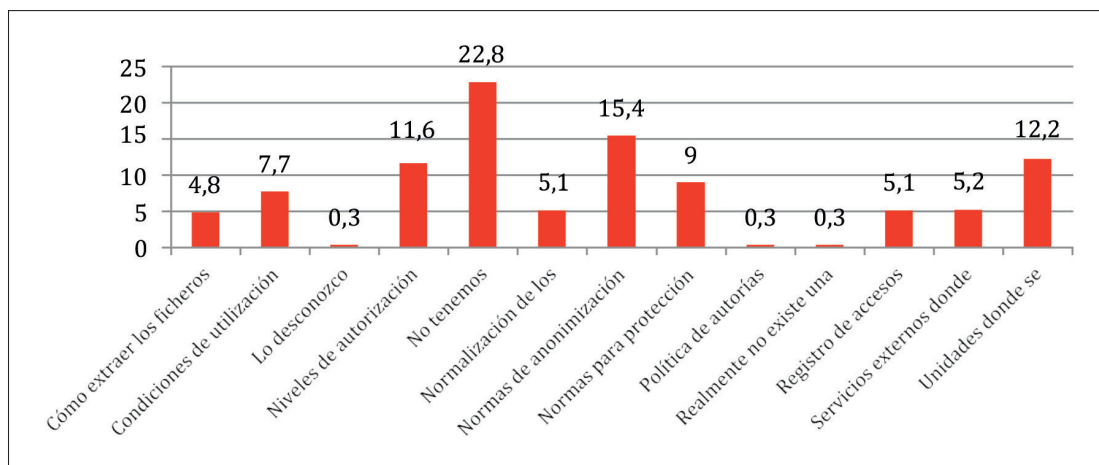


Gráfico 1. En los proyectos de investigación en los que ha participado ¿existe alguna política o plan de gestión de los datos (implícita o por escrito) que incluya alguno de estos ítem?

Los motivos para desarrollar esas políticas vienen reflejados en la gráfico 2. Casi el 30% de los entrevistados indicó que se debía a “la necesidad de la investigación para acceder, analizar o anotar los datos de otros”; el 20% indicaba “la complejidad o volumen de los datos asociados al proyecto”; el 17% indicaba que el motivo era “el tamaño del equipo de investigación”.

Respecto a los motivos por los que no existe esa política (gráfico 3), el 25,1% respondió que era debido a “la ausencia de una obligación institucional de gestión de los datos” y el 17% “por el tiempo y el esfuerzo requerido”. Otras respuestas cercanas al 13% y al 14% son: “la falta de apoyo u orientación” y “no requerido por el ente financiador del proyecto”. El 13,6% indicaba que no lo sabía o no podía contestar a la pregunta.

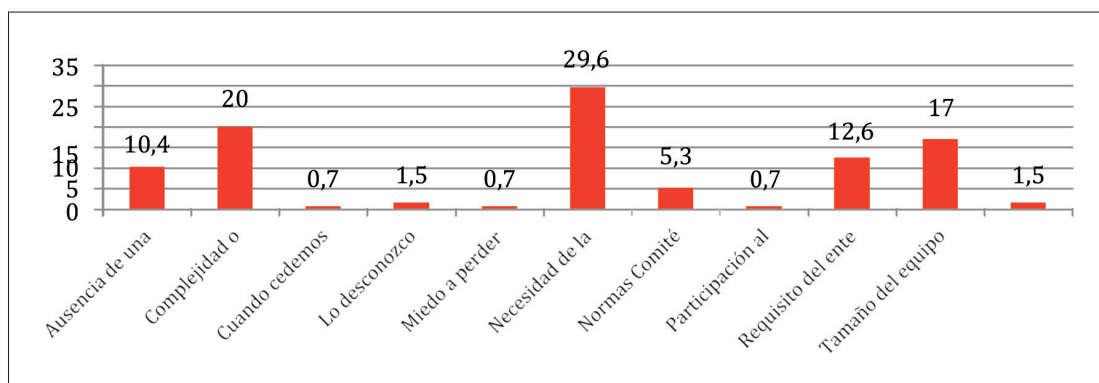


Gráfico 2. En caso de tenerla, ¿cuál fue el motivo principal para desarrollar esta política?

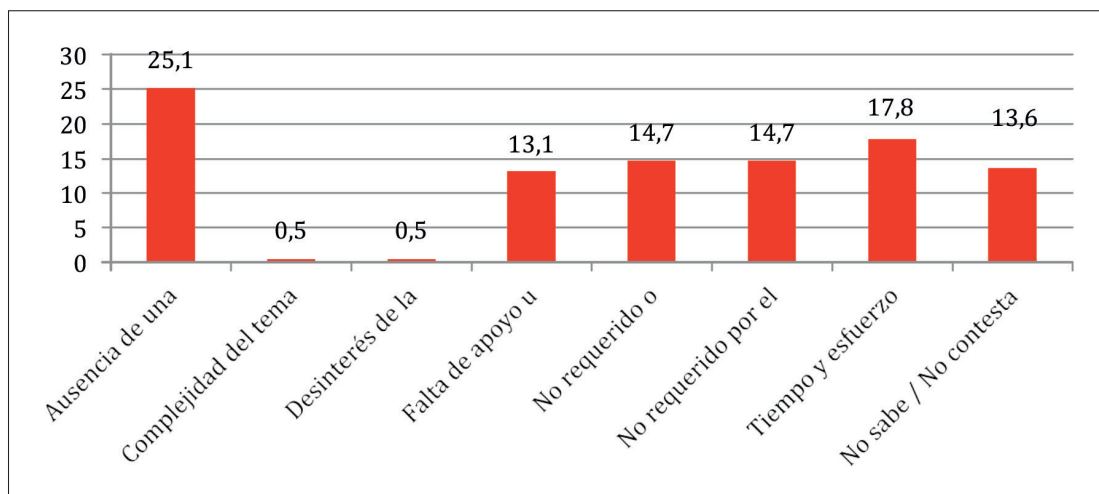


Gráfico 3. ¿Cuál cree que es el motivo principal por el que no existe ninguna política?

El relación con la pregunta ¿Quién tiene en el grupo la responsabilidad de gestionar los datos? (gráfico 4), el 29,3% indicaba que es el director del proyecto o del grupo de investigación, el 32,3% la persona que respondía a la encuesta, el 9,2% un asistente o estudiante de investigación, y el 7,4% un técnico de investigación.

Respecto al lugar de almacenamiento de los datos (gráfico 5), el 43,2% lo hace en ordenadores locales de los investigadores, el 30,8% en unidades de almacenamiento portables, el 15,6% en servidores de red institucionales y el 8,4% en un servicio de almacenamiento externo.

La siguiente pregunta hacía referencia al tipo de formato utilizado para en el tratamiento de los datos (gráfico 6). El 67,7% utiliza principalmente formatos libres, mientras que el 29,7% utiliza formatos propietarios.

Ante la pregunta de si les gustaría utilizar datos de investigación recogidos por otros investigadores, las respuestas son afirmativas mayoritariamente, con los porcentajes siguientes: 50% en los estudiantes de doctorado; 75,36% en los investigadores principales; 85,71% en los asistentes y técnicos de

investigación; 88% en los investigadores senior; 92% en los investigadores colaboradores; 100% en los investigadores en formación (tabla 3 (a)).

La contestación afirmativa a la respuesta anterior se produce en todos los rangos de tiempo de investigación y va del 81,03% en los que llevan más de 20 años al 90% en los que llevan menos de 5 años (tabla 3 (b)).

En la siguiente pregunta se indaga sobre si el investigador utiliza actualmente datos de investigación no

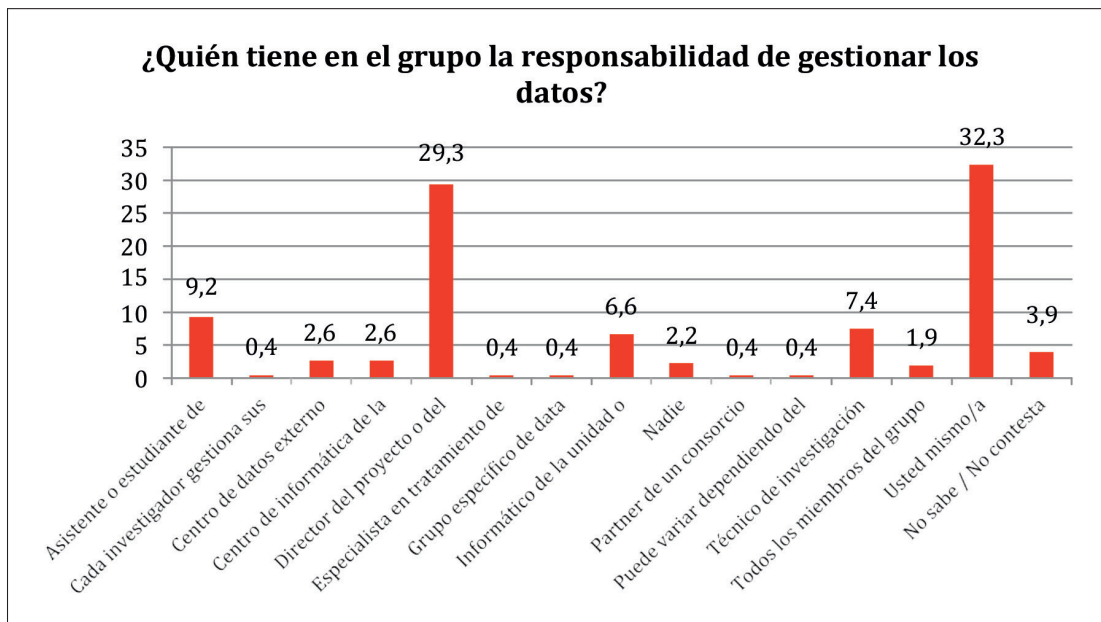


Gráfico 4. ¿Quién tiene en el grupo la responsabilidad de gestionar los datos?

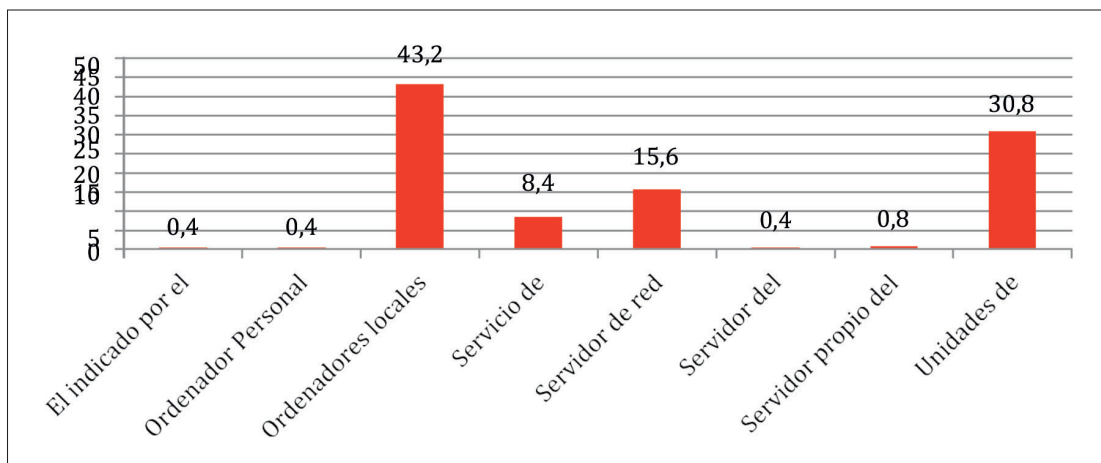


Gráfico 5. ¿Dónde suele almacenar sus datos de investigación?

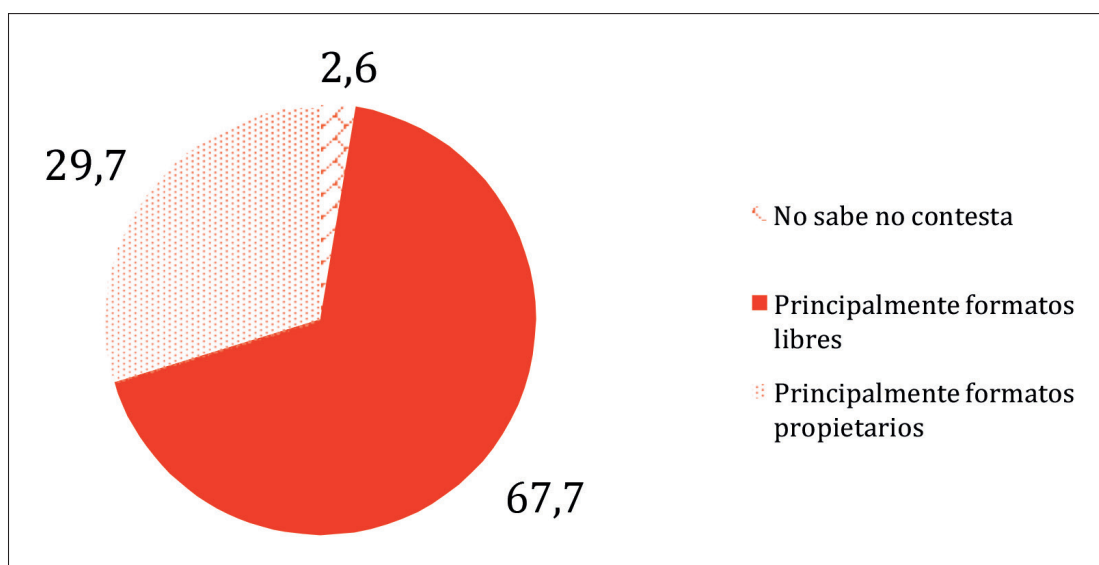


Gráfico 6. En el tratamiento de los datos de sus investigaciones utiliza

Tabla 3 (a)

¿Le gustaría utilizar datos de investigación recogidos por otros investigadores?			
Función desempeñada	No	Sí	Total
Asistente/Técnico de investigación	1 (14,29%)	6 (85,71%)	7 (4,52%)
Estudiante de doctorado	1 (50,00%)	1 (50,00%)	2 (1,29%)
Investigador colaborador	4 (8,00%)	46 (92,00%)	50 (32,26%)
Investigador en formación	0	2 (100,00%)	2 (1,29%)
Investigador principal	17 (24,64%)	52 (75,36%)	69 (44,52%)
Investigador senior	3 (12,00%)	22 (88,00%)	25 (16,13%)
Total	26 (16,77%)	129 (83,23%)	155 (100,00%)

Tabla 3 (b)

¿Le gustaría utilizar datos de investigación recogidos por otros investigadores?			
¿Cuántos años lleva investigando?	No	Sí	Total
Menos de 5 años	2 (10,00%)	18 (90,00%)	20 (12,90%)
De 5 a 10 años	5 (16,67%)	25 (83,33%)	30 (19,35%)
De 11 a 20 años	8 (17,02%)	39 (82,98%)	47 (30,32%)
Más de 20 años	11 (18,97%)	47 (81,03%)	58 (37,42%)
Total	26 (16,77%)	129 (83,23%)	155 (100,00%)

publicados de otros investigadores con los que no colabora (tabla 4 (a)). El 85,81% de todos los encuestados respondieron que no los utilizan, y esta situación se produce en todos los rangos de tiempo que llevan investigando (tabla 4 (b)).

Ante la pregunta de si utiliza actualmente datos de investigación no publicados de otros investigadores con los que no colabora se ofrecieron varias posibilidades de respuesta (tabla 5 (a)). La mayor parte de los respondientes (64,74%) manifestaron que “mis datos están disponibles para mi grupo de investigación y/o colegas colaboradores en la investigación”. Las otras respuestas obtuvieron porcentajes mucho menores que fueron del 9,83% en la

respuesta “no comparto mis datos, pero me gustaría hacerlo en un futuro”; 9,25% en la respuesta “mis datos podrían estar disponibles con los cambios apropiados (por ejemplo, datos clínicos anónimos o con la obligación de citarlos)”; 6,94% en la respuesta “mis datos están disponibles en abierto para todo el mundo”; y 6,36% en la respuesta “mis datos están disponibles en abierto para mi disciplina científica”. Sólo un 2,31% respondió que no compartía sus datos y que no quería hacerlo en el futuro y sólo un 0,58% respondió que sus datos estaban disponibles mediante el pago de una tasa. En relación con los diversos colectivos encuestados (tabla 5 (b)), el mayor porcentaje de los que respondieron que “Mis datos están

Tabla 4 (a)

¿Utiliza actualmente datos de investigación no publicados de otros investigadores con los que no colabora?			
Función desempeñada	No	Sí	Total
Asistente/Técnico de investigación	7 (100,00%)	0	7 (4,52%)
Estudiante de doctorado	2 (100,00%)	0	2 (1,29%)
Investigador colaborador	44 (88,00%)	6 (12,00%)	50 (32,26%)
Investigador en formación	1 (50,00%)	1 (50,00%)	2 (1,29%)
Investigador principal	56 (81,16%)	13 (18,84%)	69 (44,52%)
Investigador senior	23 (92,00%)	2 (8,00%)	25 (16,13%)
Total	133 (85,81%)	22 (14,19%)	155 (100,00%)

Tabla 4 (b)

¿Utiliza actualmente datos de investigación no publicados de otros investigadores con los que no colabora?			
¿Cuántos años lleva investigando?	No	Sí	Total
Menos de 5 años	17 (85,00%)	3 (15,00%)	20 (12,90%)
De 5 a 10 años	26 (86,67%)	4 (13,33%)	30 (19,35%)
De 11 a 20 años	41 (87,23%)	6 (12,77%)	47 (30,32%)
Más de 20 años	49 (84,48%)	9 (15,52%)	58 (37,42%)
Total	133 (85,81%)	22 (14,19%)	155 (100,00%)

Tabla 5 (a)

¿Utiliza actualmente datos de investigación no publicados de otros investigadores con los que no colabora?								
Función desempeñada	Mis datos están disponibles en abierto para mi disciplina científica	Mis datos están disponibles en abierto para todo el mundo	Mis datos están disponibles mediante el pago de una tasa	Mis datos están disponibles para mi grupo de investigación y/o colegas colaboradores en la investigación	Mis datos podrían estar disponibles con los cambios apropiados (por ejemplo, datos clínicos anónimos o con la obligación de citarlos)	No comparto mis datos y no quiero hacerlo en el futuro	No comparto mis datos, pero me gustaría hacerlo en un futuro	Total
Asistente/Técnico de investigación	1 (12,50%)	1 (12,50%)	0	4 (50,00%)	1 (12,50%)	0	1 (12,50%)	8 (4,62%)
Estudiante de doctorado	0	0	0	2 (100,00%)	0	0	0	2 (1,16%)
Investigador colaborador	3 (5,36%)	3 (5,36%)	1 (1,79%)	38 (67,86%)	3 (5,36%)	2 (3,57%)	6 (10,71%)	56 (32,37%)
Investigador en formación	1 (33,33%)	0	0	1 (33,33%)	0	0	1 (33,33%)	3 (1,73%)
Investigador principal	4 (5,26%)	6 (7,89%)	0	47 (61,84%)	9 (11,84%)	1 (1,32%)	9 (11,84%)	76 (43,93%)
Investigador senior	2 (7,14%)	2 (7,14%)	0	20 (71,43%)	3 (10,71%)	1 (3,57%)	0	28 (16,18%)
Total	11 (6,36%)	12 (6,94%)	1 (0,58%)	112 (64,74%)	16 (9,25%)	4 (2,31%)	17 (9,83%)	173 (100,00%)

Tabla 5 (b)

¿Utiliza actualmente datos de investigación no publicados de otros investigadores con los que no colabora?								
¿Cuántos años lleva investigando?	Mis datos están disponibles en abierto para mi disciplina científica	Mis datos están disponibles en abierto para todo el mundo	Mis datos están disponibles mediante el pago de una tasa	Mis datos están disponibles para mi grupo de investigación y/o colegas colaboradores en la investigación	Mis datos podrían estar disponibles con los cambios apropiados (por ejemplo, datos clínicos anónimos o con la obligación de citarlos)	No comparto mis datos y no quiero hacerlo en el futuro	No comparto mis datos, pero me gustaría hacerlo en un futuro	Total
Menos de 5 años	1 (4,35%)	1 (4,35%)	1 (4,35%)	14 (60,87%)	1 (4,35%)	1 (4,35%)	4 (17,39%)	23 (13,29%)
De 5 a 10 años	3 (8,82%)	4 (11,76%)	0	22 (64,71%)	4 (11,76%)	0	1 (2,94%)	34 (19,65%)
De 11 a 20 años	2 (3,77%)	4 (7,55%)	0	39 (73,58%)	5 (9,43%)	1 (1,89%)	2 (3,77%)	53 (30,64%)
Más de 20 años	5 (7,94%)	3 (4,76%)	0	37 (58,73%)	6 (9,52%)	2 (3,17%)	10 (15,87%)	63 (36,42%)
Total	11 (6,36%)	12 (6,94%)	1 (0,58%)	112 (64,74%)	16 (9,25%)	4 (2,31%)	17 (9,83%)	173 (100,00%)

disponibles para mi grupo de investigación y/o colegas colaboradores en la investigación” corresponde a los investigadores que están cursando estudios de doctorado (100%), seguido de los investigadores senior (71,43%), investigadores colaboradores (67,86%) e investigadores principales (61,84%).

La pregunta “¿Qué le preocupa a la hora de compartir los datos en un servicio externo?” (tabla 6 (a)) ha obtenido las siguientes respuestas, en orden decreciente: el 25,29% aduce “cuestiones legales: confidencialidad / derechos de propiedad intelectual”; con un porcentaje similar, 24,42%, “mal uso o interpretación de los datos”; la “pérdida de autoría” preocupa al 14,24%; el “miedo a perder la delantera en investigación” al 11,63%. Otras cuestiones como la “confidencialidad de datos del paciente”, “falta de interés para otros”, “gastar tiempo haciéndolos disponibles” y “pérdida de autoría”, son asuntos que no preocupan, a juzgar por el bajo número de encuestados que respondieron afirmativamente a estas cuestiones.

No se aprecian discrepancias en estas respuestas que llamen especialmente la atención según las categorías relativas al número de años que llevan investigando

(tabla 6 (b)). Las cuestiones legales, el mal uso o malinterpretación de los datos, el miedo a perder la delantera en la investigación, el desconocimiento de los archivos y repositorios existentes, la pérdida de tiempo haciéndolos disponibles y la pérdida de autoría son preocupaciones que afectan casi por igual a todos los grupos, independientemente del número de años que lleven investigando.

C. PRESERVACIÓN DE LOS DATOS UNA VEZ TERMINADO EL PROYECTO

Las siguientes preguntas indagan sobre las amenazas que perciben los investigadores relacionadas con los datos. En la siguiente tabla (tabla 7 (a)), se presentan las respuestas a la pregunta sobre si “los usuarios pueden ser incapaces de comprender o utilizar los datos (semántica, formatos o algoritmos implicados)”. Como se puede apreciar, no existe unanimidad en estas percepciones y los resultados son dispares, aunque por lo general no se percibe este riesgo, ya que predominan los porcentajes en las valoraciones que perciben la amenaza como menor. Tampoco los porcentajes según los años que llevan investigando permiten identificar diferencias significativas (tabla 7 (b)).

Tabla 6 (a)

¿Qué le preocupa a la hora de compartir los datos en un servicio externo?													
Función desempeñada	Confidencialidad de datos del paciente	Cuestiones legales: confidencialidad / derechos de propiedad intelectual	Falta de interés para otros	Mal uso o interpretación de los datos	Miedo a perder la delantera en investigación	No conozco archivos digitales (repositorios o centros de documentación) a los que pueda enviar datos	No creo que mis datos de investigación estén seguros en un centro de documentación, web de una revista o repositorio	No se prevén problemas (no perder) tiempo haciendo disponibles	No tengo seguridad de como fueron recogidos	Perder tiempo haciendo disponibles	Pérdida de autoría	Pérdida de autoría, hay quien aspira a tener datos sin trabajarlos...	Total
Asistente/Técnico de investigación	0	3 (27,27%)	0	3 (27,27%)	0	0	0	2 (18,18%)	0	0	3 (27,27%)	0	11 (3,20%)
Estudiante de doctorado	0	1 (25,00%)	1 (25,00%)	1 (25,00%)	0	0	0	0	0	1 (25,00%)	0	0	4 (1,16%)
Investigador colaborador	1 (0,94%)	31 (29,25%)	0	27 (25,47%)	14 (13,21%)	4 (3,77%)	5 (4,72%)	0	0	11 (10,38%)	13 (12,26%)	0	106 (30,81%)
Investigador en formación	0	1 (20,00%)	0	1 (20,00%)	0	1 (20,00%)	0	0	0	0	2 (40,00%)	0	5 (1,45%)
Investigador principal	0	36 (23,08%)	0	38 (24,36%)	16 (10,26%)	10 (6,41%)	9 (5,77%)	5 (3,21%)	1 (0,64%)	17 (10,90%)	22 (14,10%)	1 (0,64%)	156 (45,35%)
Investigador senior	0	15 (24,19%)	0	14 (22,58%)	10 (16,13%)	1 (1,61%)	2 (3,23%)	0	0	11 (17,74%)	9 (14,52%)	0	62 (18,02%)
Total	1 (0,29%)	87 (25,29%)	1 (0,29%)	84 (24,42%)	40 (11,63%)	16 (4,65%)	16 (4,65%)	7 (2,03%)	1 (0,29%)	40 (11,63%)	49 (14,24%)	1 (0,29%)	344 (100,00%)

Tabla 6 (b)

¿Cuántos años lleva investigando?	¿Qué le preocupa a la hora de compartir los datos en un servicio externo?											Total		
	Confidencialidad de datos del paciente	Cuestiones legales: confidencialidad / derechos de propiedad intelectual	Falta de interés para otros	Mal uso o interpretación de los datos	Miedo a perder la delantera en investigación	No conozco archivos digitales (repositorios o centros de documentación) a los que pueda enviar datos	No creo que mis datos de investigación estén seguros en un centro de documentación, web de una revista o repositorio	No se prevén problemas	No se prevén problemas, GASTAR (no perder tiempo haciendo datos disponibles)	No tengo seguridad de como fueron recogidos	Perder tiempo haciendo datos disponibles		Pérdida de autoría	Pérdida de autoría, hay quien aspira a tener datos sin trabajarlos...
Menos de 5 años	1 (1,92%)	13 (25,00%)	0	10 (19,23%)	8 (15,38%)	1 (1,92%)	2 (3,85%)	1 (1,92%)	0	0	8 (15,38%)	8 (15,38%)	0	52 (15,12%)
De 5 a 10 años	0	17 (30,91%)	1 (1,82%)	13 (23,64%)	7 (12,73%)	3 (5,45%)	2 (3,64%)	0	1 (1,82%)	3 (5,45%)	3 (5,45%)	8 (14,55%)	0	55 (15,99%)
De 11 a 20 años	0	27 (25,71%)	0	28 (26,67%)	13 (12,38%)	4 (3,81%)	4 (3,81%)	3 (2,86%)	0	0	13 (12,38%)	12 (11,43%)	1 (0,95%)	105 (30,52%)
Más de 20 años	0	30 (22,73%)	0	33 (25,00%)	12 (9,09%)	8 (6,06%)	8 (6,06%)	3 (2,27%)	1 (0,76%)	0	16 (12,12%)	21 (15,91%)	0	132 (38,37%)
Total	1 (0,29%)	87 (25,29%)	1 (0,29%)	84 (24,42%)	40 (11,63%)	16 (4,65%)	16 (4,65%)	7 (2,03%)	1 (0,29%)	1 (0,29%)	40 (11,63%)	49 (14,24%)	1 (0,29%)	344 (100,00%)

Tabla 7 (a)

Califique del 1 al 7 las siguientes amenazas A SUS DATOS para los próximos 10 años								
Función desempeñada	[Los usuarios pueden ser incapaces de comprender o utilizar los datos (semántica, formatos o algoritmos implicados)]							Total
	1	2	3	4	5	6	7	
Asistente/Técnico de investigación	3 (42,86%)	0	0	3 (42,86%)	0	1 (14,29%)	0	7 (4,61%)
Estudiante de doctorado	0	0	1 (50,00%)	0	0	1 (50,00%)	0	2 (1,32%)
Investigador colaborador	4 (8,51%)	9 (19,15%)	7 (14,89%)	12 (25,53%)	9 (19,15%)	4 (8,51%)	2 (4,26%)	47 (30,92%)
Investigador en formación	0	0	0	0	1 (50,00%)	1 (50,00%)	0	2 (1,32%)
Investigador principal	9 (13,04%)	13 (18,84%)	6 (8,70%)	13 (18,84%)	13 (18,84%)	8 (11,59%)	7 (10,14%)	69 (45,39%)
Investigador senior	6 (24,00%)	3 (12,00%)	3 (12,00%)	4 (16,00%)	3 (12,00%)	5 (20,00%)	1 (4,00%)	25 (16,45%)
Total	22 (14,47%)	25 (16,45%)	17 (11,18%)	32 (21,05%)	26 (17,11%)	20 (13,16%)	10 (6,58%)	152 (100,00%)

Tabla 7 (b)

Califique del 1 al 7 las siguientes amenazas A SUS DATOS para los próximos 10 años								
¿Cuántos años lleva investigando?	[Los usuarios pueden ser incapaces de comprender o utilizar los datos (semántica, formatos o algoritmos implicados)]							Total
	1	2	3	4	5	6	7	
Menos de 5 años	4 (22,22%)	5 (27,78%)	1 (5,56%)	3 (16,67%)	3 (16,67%)	2 (11,11%)	0	18 (11,84%)
De 5 a 10 años	4 (13,33%)	4 (13,33%)	5 (16,67%)	9 (30,00%)	4 (13,33%)	3 (10,00%)	1 (3,33%)	30 (19,74%)
De 11 a 20 años	9 (19,57%)	9 (19,57%)	3 (6,52%)	6 (13,04%)	6 (13,04%)	9 (19,57%)	4 (8,70%)	46 (30,26%)
Más de 20 años	5 (8,62%)	7 (12,07%)	8 (13,79%)	14 (24,14%)	13 (22,41%)	6 (10,34%)	5 (8,62%)	58 (38,16%)
Total	22 (14,47%)	25 (16,45%)	17 (11,18%)	32 (21,05%)	26 (17,11%)	20 (13,16%)	10 (6,58%)	152 (100,00%)

La siguiente amenaza, la falta de software, hardware o soporte sostenibles de los ordenadores, también proporciona resultados heterogéneos (tabla 8 (a)). Así, en los investigadores principales y en los investigadores colaboradores, este supuesto no se percibe como amenaza en alrededor del 50% de los respondientes, mientras que en los investigadores seniors los resultados son más dispares, de manera que mientras un 36% de los respondientes no lo percibe como amenaza, un 32% sí que lo percibe. En relación con los años que llevan investigando, predominan los porcentajes altos entre quienes no perciben esta amenaza en todos los periodos.

La siguiente pregunta trata de averiguar si el hecho de que el origen y la autenticidad de los datos puede ser incierto se percibe como una amenaza y también ha proporcionado respuestas heterogéneas: el 40% de los investigadores seniors no lo perciben como amenaza, pero este porcentaje sólo es del 26% en los investigadores principales y del 33% en los investigadores colaboradores (tabla 9 (a)). En este colectivo, el 43% sí que perciben este hecho como amenaza. En relación con los años que llevan investigando, los porcentajes también son dispares entre los que no perciben amenaza y los que sí que lo perciben como amenaza (tabla 9 (b)).

Tabla 8 (a)

Califique del 1 al 7 las siguientes amenazas A SUS DATOS para los próximos 10 años								
Función desempeñada	[La falta de software, hardware o soporte sostenibles de los ordenadores]							Total
	1	2	3	4	5	6	7	
Asistente/Técnico de investigación	3 (42,86%)	0	0	2 (28,57%)	0	1 (14,29%)	1 (14,29%)	7 (4,73%)
Estudiante de doctorado	1 (50,00%)	0	0	0	0	1 (50,00%)	0	2 (1,35%)
Investigador colaborador	4 (8,70%)	17 (36,96%)	8 (17,39%)	8 (17,39%)	6 (13,04%)	2 (4,35%)	1 (2,17%)	46 (31,08%)
Investigador en formación	0	1 (100,00%)	0	0	0	0	0	1 (0,68%)
Investigador principal	21 (31,34%)	11 (16,42%)	7 (10,45%)	6 (8,96%)	9 (13,43%)	7 (10,45%)	6 (8,96%)	67 (45,27%)
Investigador senior	4 (16,00%)	5 (20,00%)	4 (16,00%)	3 (12,00%)	1 (4,00%)	5 (20,00%)	3 (12,00%)	25 (16,89%)
Total	33 (22,30%)	34 (22,97%)	19 (12,84%)	19 (12,84%)	16 (10,81%)	16 (10,81%)	11 (7,43%)	148 (100,00%)

Tabla 8 (b)

Califique del 1 al 7 las siguientes amenazas A SUS DATOS para los próximos 10 años								
¿Cuántos años lleva investigando?	[La falta de software, hardware o soporte sostenibles de los ordenadores]							Total
	1	2	3	4	5	6	7	
Menos de 5 años	6 (37,50%)	6 (37,50%)	1 (6,25%)	1 (6,25%)	2 (12,50%)	0	0	16 (10,81%)
De 5 a 10 años	6 (20,69%)	5 (17,24%)	3 (10,34%)	9 (31,03%)	1 (3,45%)	4 (13,79%)	1 (3,45%)	29 (19,59%)
De 11 a 20 años	9 (19,57%)	16 (34,78%)	6 (13,04%)	3 (6,52%)	6 (13,04%)	3 (6,52%)	3 (6,52%)	46 (31,08%)
Más de 20 años	12 (21,05%)	7 (12,28%)	9 (15,79%)	6 (10,53%)	7 (12,28%)	9 (15,79%)	7 (12,28%)	57 (38,51%)
Total	33 (22,30%)	34 (22,97%)	19 (12,84%)	19 (12,84%)	16 (10,81%)	16 (10,81%)	11 (7,43%)	148 (100,00%)

Tabla 9 (a)

Califique del 1 al 7 las siguientes amenazas A SUS DATOS para los próximos 10 años								
Función desempeñada	[El origen y autenticidad de los datos puede ser incierto]							Total
	1	2	3	4	5	6	7	
Asistente/Técnico de investigación	2 (28,57%)	1 (14,29%)	2 (28,57%)	1 (14,29%)	1 (14,29%)	0	0	7 (4,67%)
Estudiante de doctorado	1 (50,00%)	0	0	1 (50,00%)	0	0	0	2 (1,33%)
Investigador colaborador	7 (14,89%)	6 (12,77%)	5 (10,64%)	7 (14,89%)	9 (19,15%)	11 (23,40%)	2 (4,26%)	47 (31,33%)
Investigador en formación	0	0	0	1 (100,00%)	0	0	0	1 (0,67%)
Investigador principal	13 (19,12%)	5 (7,35%)	6 (8,82%)	13 (19,12%)	13 (19,12%)	11 (16,18%)	7 (10,29%)	68 (45,33%)
Investigador senior	6 (24,00%)	4 (16,00%)	0	2 (8,00%)	7 (28,00%)	2 (8,00%)	4 (16,00%)	25 (16,67%)
Total	29 (19,33%)	16 (10,67%)	13 (8,67%)	25 (16,67%)	30 (20,00%)	24 (16,00%)	13 (8,67%)	150 (100,00%)

Tabla 9 (b)

Califique del 1 al 7 las siguientes amenazas A SUS DATOS para los próximos 10 años								
¿Cuántos años lleva investigando?	[El origen y autenticidad de los datos puede ser incierto]							Total
	1	2	3	4	5	6	7	
Menos de 5 años	4 (23,53%)	3 (17,65%)	3 (17,65%)	6 (35,29%)	1 (5,88%)	0	0	17 (11,33%)
De 5 a 10 años	5 (16,67%)	2 (6,67%)	4 (13,33%)	5 (16,67%)	4 (13,33%)	5 (16,67%)	5 (16,67%)	30 (20,00%)
De 11 a 20 años	11 (23,91%)	6 (13,04%)	1 (2,17%)	2 (4,35%)	11 (23,91%)	12 (26,09%)	3 (6,52%)	46 (30,67%)
Más de 20 años	9 (15,79%)	5 (8,77%)	5 (8,77%)	12 (21,05%)	14 (24,56%)	7 (12,28%)	5 (8,77%)	57 (38,00%)
Total	29 (19,33%)	16 (10,67%)	13 (8,67%)	25 (16,67%)	30 (20,00%)	24 (16,00%)	13 (8,67%)	150 (100,00%)

Las restricciones de acceso y uso relacionadas con la propiedad intelectual se perciben como amenaza entre los diversos investigadores (tabla 10 (a)). Si consideramos conjuntamente las estimaciones 5, 6 y 7, puede decirse que el 72% de los investigadores senior, el 57%

de los investigadores colaboradores y el 55% de los investigadores principales perciben como amenaza este supuesto. Los datos son más inciertos entre los grupos según los años que llevan investigando (tabla 10 (b)).

Tabla 10 (a)

Califique del 1 al 7 las siguientes amenazas A SUS DATOS para los próximos 10 años								
Función desempeñada	[Las restricciones de acceso y uso (propiedad intelectual)]							Total
	1	2	3	4	5	6	7	
Asistente/Técnico de investigación	2 (28,57%)	0	0	3 (42,86%)	0	2 (28,57%)	0	7 (4,67%)
Estudiante de doctorado	1 (50,00%)	1 (50,00%)	0	0	0	0	0	2 (1,33%)
Investigador colaborador	3 (6,52%)	6 (13,04%)	6 (13,04%)	5 (10,87%)	9 (19,57%)	12 (26,09%)	5 (10,87%)	46 (30,67%)
Investigador en formación	0	0	0	0	0	0	1 (100,00%)	1 (0,67%)
Investigador principal	5 (7,25%)	7 (10,14%)	5 (7,25%)	14 (20,29%)	14 (20,29%)	10 (14,49%)	14 (20,29%)	69 (46,00%)
Investigador senior	0	2 (8,00%)	3 (12,00%)	2 (8,00%)	3 (12,00%)	9 (36,00%)	6 (24,00%)	25 (16,67%)
Total	11 (7,33%)	16 (10,67%)	14 (9,33%)	24 (16,00%)	26 (17,33%)	33 (22,00%)	26 (17,33%)	150 (100,00%)

Tabla 10 (b)

Califique del 1 al 7 las siguientes amenazas A SUS DATOS para los próximos 10 años								
¿Cuántos años lleva investigando?	[Las restricciones de acceso y uso (propiedad intelectual)]							Total
	1	2	3	4	5	6	7	
Menos de 5 años	2 (11,76%)	3 (17,65%)	4 (23,53%)	0	1 (5,88%)	5 (29,41%)	2 (11,76%)	17 (11,33%)
De 5 a 10 años	2 (6,67%)	3 (10,00%)	3 (10,00%)	6 (20,00%)	5 (16,67%)	5 (16,67%)	6 (20,00%)	30 (20,00%)
De 11 a 20 años	3 (6,67%)	5 (11,11%)	2 (4,44%)	4 (8,89%)	10 (22,22%)	9 (20,00%)	12 (26,67%)	45 (30,00%)
Más de 20 años	4 (6,90%)	5 (8,62%)	5 (8,62%)	14 (24,14%)	10 (17,24%)	14 (24,14%)	6 (10,34%)	58 (38,67%)
Total	11 (7,33%)	16 (10,67%)	14 (9,33%)	24 (16,00%)	26 (17,33%)	33 (22,00%)	26 (17,33%)	150 (100,00%)

La amenaza relacionada con la pérdida de la capacidad para localizar los datos no arroja resultados concluyentes (tabla 11 (a) y tabla 11 (b)).

Tabla 11 (a)

Califique del 1 al 7 las siguientes amenazas A SUS DATOS para los próximos 10 años								
Función desempeñada	[Pérdida de capacidad para localizar los datos]							Total
	1	2	3	4	5	6	7	
Asistente/Técnico de investigación	1 (14,29%)	1 (14,29%)	2 (28,57%)	1 (14,29%)	0	0	2 (28,57%)	7 (4,70%)
Estudiante de doctorado	0	0	0	0	1 (50,00%)	0	1 (50,00%)	2 (1,34%)
Investigador colaborador	5 (10,64%)	9 (19,15%)	4 (8,51%)	10 (21,28%)	6 (12,77%)	8 (17,02%)	5 (10,64%)	47 (31,54%)
Investigador en formación	0	0	1 (100,00%)	0	0	0	0	1 (0,67%)
Investigador principal	13 (19,40%)	8 (11,94%)	12 (17,91%)	11 (16,42%)	13 (19,40%)	6 (8,96%)	4 (5,97%)	67 (44,97%)
Investigador senior	2 (8,00%)	1 (4,00%)	4 (16,00%)	4 (16,00%)	6 (24,00%)	7 (28,00%)	1 (4,00%)	25 (16,78%)
Total	21 (14,09%)	19 (12,75%)	23 (15,44%)	26 (17,45%)	26 (17,45%)	21 (14,09%)	13 (8,72%)	149 (100,00%)

Tabla 11 (b)

Califique del 1 al 7 las siguientes amenazas A SUS DATOS para los próximos 10 años								
¿Cuántos años lleva investigando?	[Pérdida de capacidad para localizar los datos]							Total
	1	2	3	4	5	6	7	
Menos de 5 años	2 (11,76%)	4 (23,53%)	4 (23,53%)	3 (17,65%)	1 (5,88%)	2 (11,76%)	1 (5,88%)	17 (11,41%)
De 5 a 10 años	4 (13,79%)	6 (20,69%)	2 (6,90%)	3 (10,34%)	5 (17,24%)	3 (10,34%)	6 (20,69%)	29 (19,46%)
De 11 a 20 años	7 (15,56%)	3 (6,67%)	6 (13,33%)	10 (22,22%)	8 (17,78%)	7 (15,56%)	4 (8,89%)	45 (30,20%)
Más de 20 años	8 (13,79%)	6 (10,34%)	11 (18,97%)	10 (17,24%)	12 (20,69%)	9 (15,52%)	2 (3,45%)	58 (38,93%)
Total	21 (14,09%)	19 (12,75%)	23 (15,44%)	26 (17,45%)	26 (17,45%)	21 (14,09%)	13 (8,72%)	149 (100,00%)

El siguiente supuesto, relativo a que “la custodia actual de los datos, en una organización o proyecto, puede dejar de existir en algún punto en el futuro” se percibe como amenaza entre los diversos investigadores (tabla 12 (a)). Si de nuevo consideramos conjuntamente los grados de amenaza de 5, 6 y 7, la amenaza se percibe en el 72% de los investigadores senior, en el 63% de los investigadores colaboradores y en el 42% de los investigadores principales. Los datos son similares para todos

los colectivos según los años que llevan investigando (tabla 12 (b)).

La amenaza de que “aquellos en quienes confiamos para cuidar los datos, pueden fallarnos” no presenta resultados concluyentes debido a que las respuestas se diluyen en todos los rangos posibles, tanto entre los diversos colectivos como en el número de años que llevan investigando (tabla 13 (a) y tabla 13 (b)).

Tabla 12 (a)

Califique del 1 al 7 las siguientes amenazas A SUS DATOS para los próximos 10 años								
Función desempeñada	[La custodia actual de los datos, en una organización o proyecto, puede dejar de existir en algún punto en el futuro]							Total
	1	2	3	4	5	6	7	
Asistente/Técnico de investigación	2 (28,57%)	2 (28,57%)	1 (14,29%)	1 (14,29%)	0	1 (14,29%)	0	7 (4,73%)
Estudiante de doctorado	1 (50,00%)	0	0	0	0	0	1 (50,00%)	2 (1,35%)
Investigador colaborador	4 (8,70%)	5 (10,87%)	3 (6,52%)	5 (10,87%)	5 (10,87%)	15 (32,61%)	9 (19,57%)	46 (31,08%)
Investigador en formación	0	0	0	0	0	0	1 (100,00%)	1 (0,68%)
Investigador principal	7 (10,45%)	6 (8,96%)	9 (13,43%)	13 (19,40%)	9 (13,43%)	15 (22,39%)	8 (11,94%)	67 (45,27%)
Investigador senior	1 (4,00%)	2 (8,00%)	1 (4,00%)	3 (12,00%)	7 (28,00%)	6 (24,00%)	5 (20,00%)	25 (16,89%)
Total	15 (10,14%)	15 (10,14%)	14 (9,46%)	22 (14,86%)	21 (14,19%)	37 (25,00%)	24 (16,22%)	148 (100,00%)

Tabla 12 (b)

Califique del 1 al 7 las siguientes amenazas A SUS DATOS para los próximos 10 años								
¿Cuántos años lleva investigando?	[La custodia actual de los datos, en una organización o proyecto, puede dejar de existir en algún punto en el futuro]							Total
	1	2	3	4	5	6	7	
Menos de 5 años	4 (23,53%)	6 (35,29%)	0	2 (11,76%)	0	2 (11,76%)	3 (17,65%)	17 (11,49%)
De 5 a 10 años	4 (13,33%)	3 (10,00%)	3 (10,00%)	4 (13,33%)	3 (10,00%)	9 (30,00%)	4 (13,33%)	30 (20,27%)
De 11 a 20 años	2 (4,65%)	5 (11,63%)	4 (9,30%)	6 (13,95%)	5 (11,63%)	12 (27,91%)	9 (20,93%)	43 (29,05%)
Más de 20 años	5 (8,62%)	1 (1,72%)	7 (12,07%)	10 (17,24%)	13 (22,41%)	14 (24,14%)	8 (13,79%)	58 (39,19%)
Total	15 (10,14%)	15 (10,14%)	14 (9,46%)	22 (14,86%)	21 (14,19%)	37 (25,00%)	24 (16,22%)	148 (100,00%)

Tabla 13 (a)

Califique del 1 al 7 las siguientes amenazas A SUS DATOS para los próximos 10 años								
Función desempeñada	[Aquellos en quienes confiamos para cuidar los datos, pueden fallarnos]							Total
	1	2	3	4	5	6	7	
Asistente/Técnico de investigación	1 (14,29%)	2 (28,57%)	2 (28,57%)	0	1 (14,29%)	0	1 (14,29%)	7 (4,70%)
Estudiante de doctorado	2 (100,00%)	0	0	0	0	0	0	2 (1,34%)
Investigador colaborador	4 (8,70%)	7 (15,22%)	3 (6,52%)	7 (15,22%)	7 (15,22%)	8 (17,39%)	10 (21,74%)	46 (30,87%)
Investigador en formación	0	0	0	0	0	0	1 (100,00%)	1 (0,67%)
Investigador principal	9 (13,24%)	6 (8,82%)	10 (14,71%)	13 (19,12%)	7 (10,29%)	14 (20,59%)	9 (13,24%)	68 (45,64%)
Investigador senior	3 (12,00%)	3 (12,00%)	3 (12,00%)	2 (8,00%)	5 (20,00%)	7 (28,00%)	2 (8,00%)	25 (16,78%)
Total	19 (12,75%)	18 (12,08%)	18 (12,08%)	22 (14,77%)	20 (13,42%)	29 (19,46%)	23 (15,44%)	149 (100,00%)

Tabla 13 (b)

Califique del 1 al 7 las siguientes amenazas A SUS DATOS para los próximos 10 años								
¿Cuántos años lleva investigando?	[Aquellos en quienes confiamos para cuidar los datos, pueden fallarnos]							Total
	1	2	3	4	5	6	7	
Menos de 5 años	4 (23,53%)	4 (23,53%)	2 (11,76%)	0	3 (17,65%)	2 (11,76%)	2 (11,76%)	17 (11,41%)
De 5 a 10 años	5 (16,67%)	3 (10,00%)	3 (10,00%)	5 (16,67%)	3 (10,00%)	6 (20,00%)	5 (16,67%)	30 (20,13%)
De 11 a 20 años	3 (6,52%)	7 (15,22%)	6 (13,04%)	6 (13,04%)	5 (10,87%)	8 (17,39%)	11 (23,91%)	46 (30,87%)
Más de 20 años	7 (12,50%)	4 (7,14%)	7 (12,50%)	11 (19,64%)	9 (16,07%)	13 (23,21%)	5 (8,93%)	56 (37,58%)
Total	19 (12,75%)	18 (12,08%)	18 (12,08%)	22 (14,77%)	20 (13,42%)	29 (19,46%)	23 (15,44%)	149 (100,00%)

Dentro de este capítulo sobre la preservación de los datos, es interesante conocer cómo se almacenan los datos de investigación para preservar su acceso y utilización futura (gráfico 7). Más de la mitad de los encuestados (51,6%) respondió que los almacena en su ordenador personal; el 21,6% respondió que lo hace en el servidor

de la organización y el 9,3% en un servicio web externo como Google Drive o Dropbox. Como se aprecia en la figura, el uso de servicios externos especializados como Dryad, ArrayExpress, Figshare o DataOne, es utilizado por un porcentaje muy bajo de los encuestados (1,6%), y lo mismo ocurre con otras posibilidades.

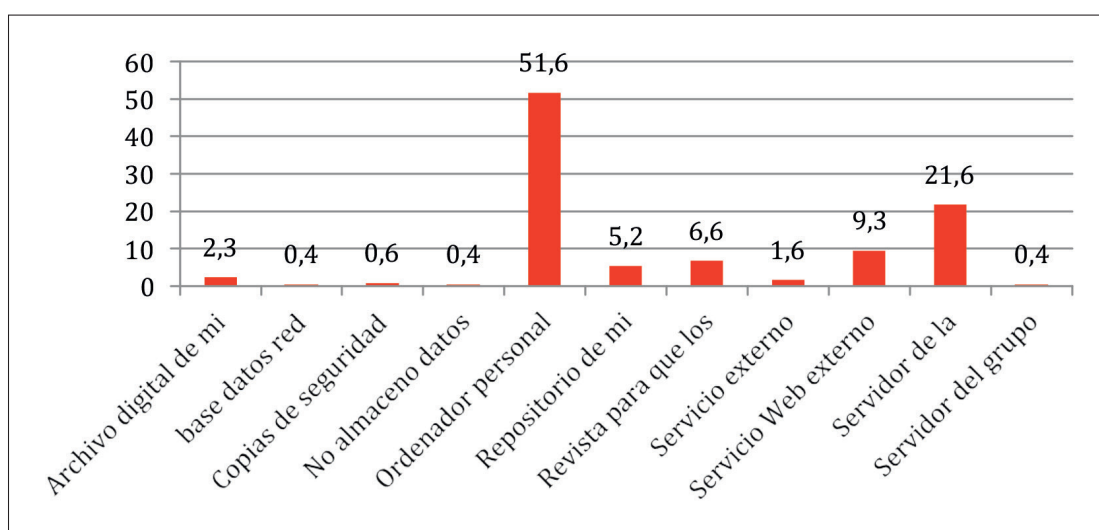


Gráfico 7. ¿Cómo almacena los datos de investigación para preservar su acceso y utilización futura?

Ante la pregunta de dónde considera más adecuado que se preserven y se compartan los datos de las investigaciones (gráfico 8), las respuestas cambian claramente respecto a la pregunta anterior, de manera que, en este caso, la consideración de preservación en el ordenador personal sólo es del 1,7%. La mayoría de los respondientes indicaron que el repositorio de la organización es el lugar apropiado (36,4%), el 23% opinó que debería ser el servidor de la organización, el 17,9% en un servicio web externo y el 16,8% en una editorial.

Un hecho que llama mucho la atención es la falta de información de los usuarios sobre la existencia de instalaciones que permitan preservar los datos digitales. Un 82,6% de los respondientes indicaron que ignoraban su existencia frente a un 10,3% que indicaron que conocían su existencia (gráfico 9).

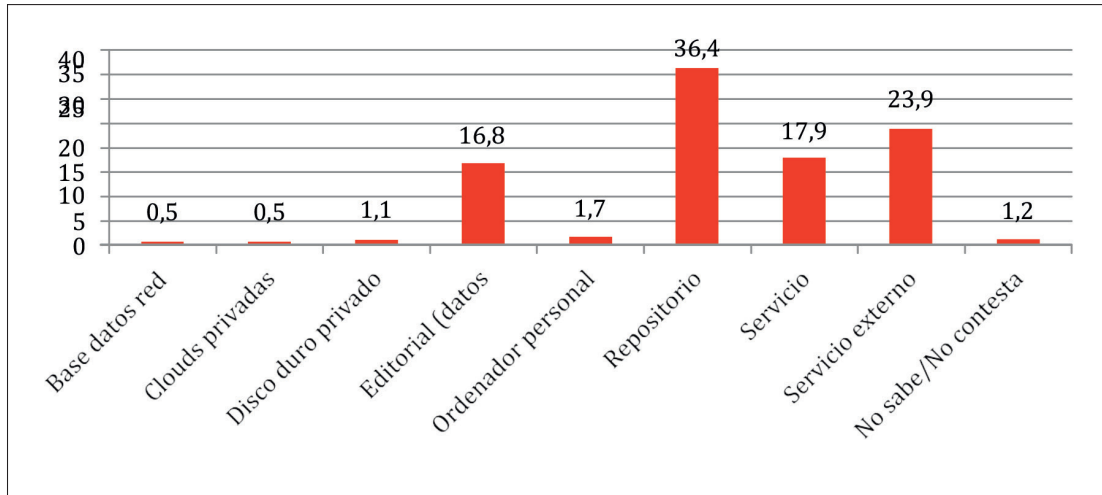


Gráfico 8. ¿Dónde considera más adecuado que se preserven y compartan los datos de sus investigaciones?

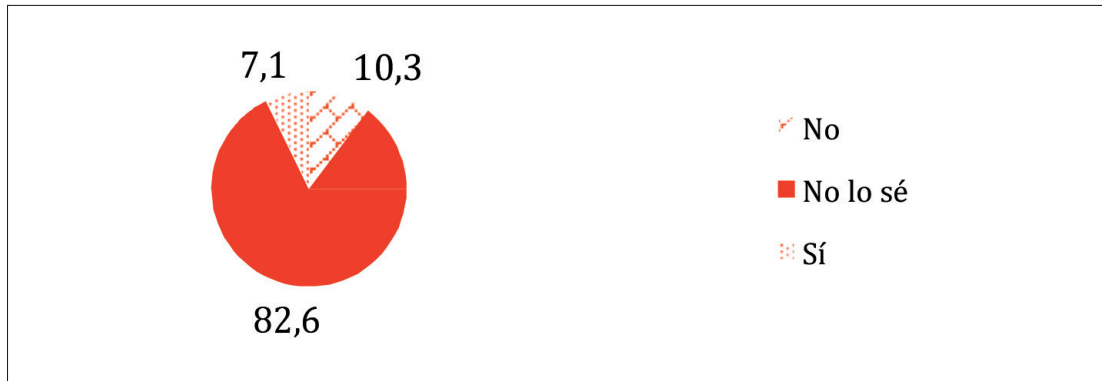


Gráfico 9. ¿Existe en su disciplina alguna instalación para la preservación de datos digitales que pueda utilizar?

La pregunta relativa a “quién proporciona guías o recomendaciones para preservar datos en la disciplina” (tabla 14 (a)), ha sido contestada con la respuesta “No lo sé” en todos los colectivos, con porcentajes que van del 33% en los investigadores principales al 45% en los colaboradores. La segunda respuesta aportada, con porcentajes en algunos casos cercanos a los anteriores, ha sido “Sociedades científicas / Asociaciones profesionales” con rangos cercanos al 29% en los Investigadores principales e Investigadores colaboradores y del 33% en quienes están realizando un doctorado. Estas dos respuestas son también las de mayor frecuencia en los rangos por años que llevan investigando (tabla 14 (b)).

Al solicitarles que priorizaran en una escala de 1 a 7 la afirmación de que “Si la investigación está financiada públicamente, los resultados deberían convertirse en propiedad pública y, por tanto, estar debidamente preservados” (tabla 15 (a)), la mayor parte de los investigadores principales de los proyectos le dieron la máxima prioridad (45%), mientras que en los investigadores

seniors y en los colaboradores, esta prioridad fue algo menor. Ningún investigador principal dio la mínima prioridad a este aspecto, pero sí un investigador senior y otro colaborador.

La prioridad que se le da a este aspecto es mayor cuanto menor es el número de años que llevan investigando, de manera que más del 80% de los investigadores que llevan hasta 10 años investigando dan una prioridad 6 o 7, mientras que este porcentaje es del 61% en los que llevan de 11 a 20 años y del 55% de los que llevan más de 20 años (tabla 15 (b)).

Las siguientes tablas presentan las razones para preservar los datos según la función desempeñada y los años que lleva investigando. Se han propuesto las siguientes preguntas: a) Estimula el avance de la ciencia basada en el conocimiento existente; b) Sirve para validar las investigaciones; c) Permite el re-análisis de datos existentes; d) Puede estimular colaboraciones interdisciplinarias; e) Tiene valor económico potencial; e) Son únicos e irremplazables.

Tabla 14 (a)

En su disciplina, ¿quién proporciona guías o recomendaciones para preservar datos?								
Función desempeñada	Editoriales (revistas)	Gobierno	Mi propia organización	Ninguno de los de arriba	No lo sé	Org. Financiadoras	Soc. cient. / Asoc. Prof.	Total
Asistente/Técnico de investigación	0	0	3 (42,86%)	1 (14,29%)	3 (42,86%)	0	0	7 (3,55%)
Estudiante de doctorado	1 (33,33%)	0	0	0	1 (33,33%)	0	1 (33,33%)	3 (1,52%)
Investigador colaborador	5 (8,33%)	1 (1,67%)	6 (10,00%)	2 (3,33%)	27 (45,00%)	1 (1,67%)	18 (30,00%)	60 (30,46%)
Investigador en formación	0	0	0	0	2 (100,00%)	0	0	2 (1,02%)
Investigador principal	9 (9,57%)	5 (5,32%)	12 (12,77%)	6 (6,38%)	31 (32,98%)	4 (4,26%)	27 (28,72%)	94 (47,72%)
Investigador senior	2 (6,45%)	1 (3,23%)	3 (9,68%)	3 (9,68%)	11 (35,48%)	2 (6,45%)	9 (29,03%)	31 (15,74%)
Total	17 (8,63%)	7 (3,55%)	24 (12,18%)	12 (6,09%)	75 (38,07%)	7 (3,55%)	55 (27,92%)	197 (100,00%)

Tabla 14 (b)

En su disciplina, ¿quién proporciona guías o recomendaciones para preservar datos?								
¿Cuántos años lleva investigando?	Editoriales (revistas)	Gobierno	Mi propia organización	Ninguno de los de arriba	No lo sé	Organizaciones financiadoras	Sociedades científicas / Asociaciones profesionales	Total
Menos de 5 años	3 (12,50%)	0	5 (20,83%)	3 (12,50%)	7 (29,17%)	0	6 (25,00%)	24 (12,18%)
De 5 a 10 años	4 (10,26%)	4 (10,26%)	3 (7,69%)	1 (2,56%)	14 (35,90%)	1 (2,56%)	12 (30,77%)	39 (19,80%)
De 11 a 20 años	4 (7,27%)	1 (1,82%)	7 (12,73%)	2 (3,64%)	26 (47,27%)	0	15 (27,27%)	55 (27,92%)
Más de 20 años	6 (7,59%)	2 (2,53%)	9 (11,39%)	6 (7,59%)	28 (35,44%)	6 (7,59%)	22 (27,85%)	79 (40,10%)
Total	17 (8,63%)	7 (3,55%)	24 (12,18%)	12 (6,09%)	75 (38,07%)	7 (3,55%)	55 (27,92%)	197 (100,00%)

Tabla 15 (a)

Priorice del 1 al 7 la importancia de las siguientes razones para preservar datos								
Función desempeñada	[Si la investigación está financiada públicamente, los resultados deberían convertirse en propiedad pública y, por tanto, estar debidamente preservados]							Total
	1	2	3	4	5	6	7	
Asistente/Técnico de investigación	0	0	0	0	2 (33,33%)	1 (16,67%)	3 (50,00%)	6 (3,97%)
Estudiante de doctorado	0	0	0	0	0	2 (100,00%)	0	2 (1,32%)
Investigador colaborador	1 (2,08%)	1 (2,08%)	1 (2,08%)	2 (4,17%)	10 (20,83%)	20 (41,67%)	13 (27,08%)	48 (31,79%)
Investigador en formación	0	0	0	0	0	1 (100,00%)	0	1 (0,66%)
Investigador principal	0	4 (5,80%)	6 (8,70%)	2 (2,90%)	12 (17,39%)	14 (20,29%)	31 (44,93%)	69 (45,70%)
Investigador senior	1 (4,00%)	1 (4,00%)	1 (4,00%)	5 (20,00%)	3 (12,00%)	9 (36,00%)	5 (20,00%)	25 (16,56%)
Total	2 (1,32%)	6 (3,97%)	8 (5,30%)	9 (5,96%)	27 (17,88%)	47 (31,13%)	52 (34,44%)	151 (100,00%)

Tabla 15 (b)

Priorice del 1 al 7 la importancia de las siguientes razones para preservar datos								
¿Cuántos años lleva investigando?	[Si la investigación está financiada públicamente, los resultados deberían convertirse en propiedad pública y, por tanto, estar debidamente preservados]							Total
	1	2	3	4	5	6	7	
Menos de 5 años	1 (5,88%)	1 (5,88%)	0	1 (5,88%)	0	9 (52,94%)	5 (29,41%)	17 (11,26%)
De 5 a 10 años	1 (3,33%)	0	0	0	4 (13,33%)	8 (26,67%)	17 (56,67%)	30 (19,87%)
De 11 a 20 años	0	2 (4,35%)	2 (4,35%)	2 (4,35%)	12 (26,09%)	17 (36,96%)	11 (23,91%)	46 (30,46%)
Más de 20 años	0	3 (5,17%)	6 (10,34%)	6 (10,34%)	11 (18,97%)	13 (22,41%)	19 (32,76%)	58 (38,41%)
Total	2 (1,32%)	6 (3,97%)	8 (5,30%)	9 (5,96%)	27 (17,88%)	47 (31,13%)	52 (34,44%)	151 (100,00%)

Como se aprecia en las tablas siguientes (tablas 16 a 21), la mayor parte de los investigadores han respondido positivamente ante estas afirmaciones. La valoración ha sido más positiva en las cuatro primeras aseveraciones, mientras que ha obtenido una menor valoración la

que se refiere al valor económico potencial de los datos. La afirmación que indica que los datos son únicos e irremplazables ha obtenido valoraciones dispares. Estas aseveraciones se mantienen en todos los rangos por años que llevan investigando.

Tabla 16

Priorice del 1 al 7 la importancia de las siguientes razones para preservar datos								
Función desempeñada	[Estimula el avance de la ciencia basada en el conocimiento existente]							Total
	1	2	3	4	5	6	7	
Asistente/Técnico de investigación	0	0	0	1 (14,29%)	2 (28,57%)	1 (14,29%)	3 (42,86%)	7 (4,67%)
Estudiante de doctorado	0	0	0	0	1 (50,00%)	1 (50,00%)	0	2 (1,33%)
Investigador colaborador	1 (2,08%)	0	5 (10,42%)	7 (14,58%)	6 (12,50%)	18 (37,50%)	11 (22,92%)	48 (32,00%)
Investigador en formación	0	0	0	0	1 (100,00%)	0	0	1 (0,67%)
Investigador principal	1 (1,49%)	0	4 (5,97%)	3 (4,48%)	16 (23,88%)	15 (22,39%)	28 (41,79%)	67 (44,67%)
Investigador senior	1 (4,00%)	1 (4,00%)	2 (8,00%)	1 (4,00%)	8 (32,00%)	7 (28,00%)	5 (20,00%)	25 (16,67%)
Total	3 (2,00%)	1 (0,67%)	11 (7,33%)	12 (8,00%)	34 (22,67%)	42 (28,00%)	47 (31,33%)	150 (100,00%)

Tabla 17

Priorice del 1 al 7 la importancia de las siguientes razones para preservar datos								
Función desempeñada	[Sirve para validar las investigaciones]							Total
	1	2	3	4	5	6	7	
Asistente/Técnico de investigación	0	0	0	0	1 (16,67%)	2 (33,33%)	3 (50,00%)	6 (3,97%)
Estudiante de doctorado	0	0	0	0	1 (50,00%)	1 (50,00%)	0	2 (1,32%)
Investigador colaborador	1 (2,08%)	0	2 (4,17%)	4 (8,33%)	11 (22,92%)	19 (39,58%)	11 (22,92%)	48 (31,79%)
Investigador en formación	0	0	0	0	0	0	1 (100,00%)	1 (0,66%)
Investigador principal	0	1 (1,45%)	1 (1,45%)	7 (10,14%)	16 (23,19%)	23 (33,33%)	21 (30,43%)	69 (45,70%)
Investigador senior	0	1 (4,00%)	1 (4,00%)	2 (8,00%)	7 (28,00%)	6 (24,00%)	8 (32,00%)	25 (16,56%)
Total	1 (0,66%)	2 (1,32%)	4 (2,65%)	13 (8,61%)	36 (23,84%)	51 (33,77%)	44 (29,14%)	151 (100,00%)

Tabla 18

Priorice del 1 al 7 la importancia de las siguientes razones para preservar datos								
Función desempeñada	[Permite el re-análisis de datos existentes]							Total
	1	2	3	4	5	6	7	
Asistente/Técnico de investigación	0	0	0	0	2 (33,33%)	1 (16,67%)	3 (50,00%)	6 (4,05%)
Estudiante de doctorado	0	0	0	0	1 (50,00%)	1 (50,00%)	0	2 (1,35%)
Investigador colaborador	1 (2,17%)	1 (2,17%)	1 (2,17%)	5 (10,87%)	8 (17,39%)	17 (36,96%)	13 (28,26%)	46 (31,08%)
Investigador en formación	0	0	0	0	0	0	1 (100,00%)	1 (0,68%)
Investigador principal	0	1 (1,47%)	1 (1,47%)	2 (2,94%)	15 (22,06%)	22 (32,35%)	27 (39,71%)	68 (45,95%)
Investigador senior	1 (4,00%)	1 (4,00%)	0	1 (4,00%)	8 (32,00%)	7 (28,00%)	7 (28,00%)	25 (16,89%)
Total	2 (1,35%)	3 (2,03%)	2 (1,35%)	8 (5,41%)	34 (22,97%)	48 (32,43%)	51 (34,46%)	148 (100,00%)

Tabla 19

Priorice del 1 al 7 la importancia de las siguientes razones para preservar datos								
Función desempeñada	[Puede estimular colaboraciones interdisciplinarias]							Total
	1	2	3	4	5	6	7	
Asistente/Técnico de investigación	0	0	0	2 (33,33%)	1 (16,67%)	1 (16,67%)	2 (33,33%)	6 (4,00%)
Estudiante de doctorado	0	0	0	0	1 (50,00%)	1 (50,00%)	0	2 (1,33%)
Investigador colaborador	0	0	1 (2,13%)	5 (10,64%)	15 (31,91%)	15 (31,91%)	11 (23,40%)	47 (31,33%)
Investigador en formación	0	0	0	1 (100,00%)	0	0	0	1 (0,67%)
Investigador principal	0	0	2 (2,90%)	8 (11,59%)	17 (24,64%)	27 (39,13%)	15 (21,74%)	69 (46,00%)
Investigador senior	0	1 (4,00%)	1 (4,00%)	3 (12,00%)	9 (36,00%)	6 (24,00%)	5 (20,00%)	25 (16,67%)
Total	0	1 (0,67%)	4 (2,67%)	19 (12,67%)	43 (28,67%)	50 (33,33%)	33 (22,00%)	150 (100,00%)

Tabla 20

Priorice del 1 al 7 la importancia de las siguientes razones para preservar datos								
Función desempeñada	[Tiene valor económico potencial]							Total
	1	2	3	4	5	6	7	
Asistente/Técnico de investigación	0	2 (33,33%)	1 (16,67%)	2 (33,33%)	0	0	1 (16,67%)	6 (4,00%)
Estudiante de doctorado	0	1 (50,00%)	1 (50,00%)	0	0	0	0	2 (1,33%)
Investigador colaborador	3 (6,38%)	5 (10,64%)	6 (12,77%)	15 (31,91%)	10 (21,28%)	7 (14,89%)	1 (2,13%)	47 (31,33%)
Investigador en formación	0	1 (100,00%)	0	0	0	0	0	1 (0,67%)
Investigador principal	3 (4,35%)	7 (10,14%)	8 (11,59%)	14 (20,29%)	17 (24,64%)	10 (14,49%)	10 (14,49%)	69 (46,00%)
Investigador senior	3 (12,00%)	2 (8,00%)	2 (8,00%)	6 (24,00%)	6 (24,00%)	3 (12,00%)	3 (12,00%)	25 (16,67%)
Total	9 (6,00%)	18 (12,00%)	18 (12,00%)	37 (24,67%)	33 (22,00%)	20 (13,33%)	15 (10,00%)	150 (100,00%)


Tabla 21

Función desempeñada	Priorice del 1 al 7 la importancia de las siguientes razones para preservar datos							Total
	[Son únicos e irremplazables]							
	1	2	3	4	5	6	7	
Asistente/Técnico de investigación	0	0	2 (33,33%)	1 (16,67%)	0	0	3 (50,00%)	6 (4,11%)
Estudiante de doctorado	0	0	1 (50,00%)	0	0	0	1 (50,00%)	2 (1,37%)
Investigador colaborador	0	3 (6,67%)	7 (15,56%)	3 (6,67%)	10 (22,22%)	12 (26,67%)	10 (22,22%)	45 (30,82%)
Investigador en formación	0	0	0	0	1 (100,00%)	0	0	1 (0,68%)
Investigador principal	1 (1,47%)	3 (4,41%)	4 (5,88%)	10 (14,71%)	12 (17,65%)	10 (14,71%)	28 (41,18%)	68 (46,58%)
Investigador senior	1 (4,17%)	3 (12,50%)	5 (20,83%)	4 (16,67%)	5 (20,83%)	3 (12,50%)	3 (12,50%)	24 (16,44%)
Total	2 (1,37%)	9 (6,16%)	19 (13,01%)	18 (12,33%)	28 (19,18%)	25 (17,12%)	45 (30,82%)	146 (100,00%)

ESTRATEGIAS PARA LA GESTIÓN Y LA PROMOCIÓN DEL USO COMPARTIDO DE DATOS CIENTÍFICOS SOBRE LA SALUD

Alexandre-Benavent, R. (1,2); Ferrer Sapena, A. (3); Paset Mancebo, F (3); Agulló Martínez, A (4); Alonso Arroyo, A (5,2); Vidal Infer, A (5,2).

1. Instituto de Historia de la Medicina y de la Ciencia López Piñero (CSIC-Universitat de València). 2. Unidad de Información e Investigación Social y Sanitaria-UISYS, Universitat de València. 3. Universidad Politécnica de Valencia. 4. Unidad de Documentación Clínica y Admisión, Hospital Clínico Universitario, Valencia. 5. Departamento de Historia de la Ciencia y Documentación, Universitat de València.



Stiges, 3-5 Mayo 2013
<http://www.congresdocumentacion2013.com>

OBJETIVOS

Proyectos OpendataScience y Health-datasea:

- Identificar las necesidades en gestión de datos de investigación de los profesionales de la salud
- Desarrollar recursos para la gestión del ciclo de vida digital de los datos científicos
- Promover su uso compartido

METODO


- a) Identificación de los centros productores de datos biomédicos en España
- b) Distribución y análisis de una encuesta sobre necesidades y uso de datos digitales
- c) Análisis de las políticas de publicación en abierto de las revistas
- d) Elaboración de guías para investigadores, profesionales y gestores de la investigación
- e) Diseño de campañas de sensibilización
- f) Puesta en funcionamiento de un buscador

RESULTADOS INICIALES

Distribución de la encuesta:

- Datos personales
- Información sobre la creación y reutilización de los datos
- Información sobre la Preservación de los datos una vez terminado el trabajo de investigación

✓ Estudiar las mejores formas de promover el acceso abierto y preservar e intercambiar los datos científicos es un propósito de interés internacional necesario para el avance de la ciencia.



Gestión de datos de investigación*

Este cuestionario pretende ser un primer paso de investigación (CSIC) sobre la creación y el desarrollo de los datos científicos en formato digital.

El objetivo es conocer las necesidades que existen para los investigadores de la medicina, que se reflejen en un cuestionario que se usará para el desarrollo de la investigación de los datos científicos, desde el ciclo de vida digital de los datos científicos, desde su creación hasta su preservación y reutilización.

El cuestionario está dividido en tres secciones: Datos personales, Datos de investigación y Datos de preservación de los datos.

El cuestionario está dividido en tres secciones: Datos personales, Datos de investigación y Datos de preservación de los datos.

El cuestionario está dividido en tres secciones: Datos personales, Datos de investigación y Datos de preservación de los datos.

El cuestionario está dividido en tres secciones: Datos personales, Datos de investigación y Datos de preservación de los datos.

B. Creación y reutilización de los datos

El cuestionario pretende ser un primer paso de investigación (CSIC) sobre la creación y el desarrollo de los datos científicos en formato digital.

El objetivo es conocer las necesidades que existen para los investigadores de la medicina, que se reflejen en un cuestionario que se usará para el desarrollo de la investigación de los datos científicos, desde el ciclo de vida digital de los datos científicos, desde su creación hasta su preservación y reutilización.

El cuestionario está dividido en tres secciones: Datos personales, Datos de investigación y Datos de preservación de los datos.

El cuestionario está dividido en tres secciones: Datos personales, Datos de investigación y Datos de preservación de los datos.

El cuestionario está dividido en tres secciones: Datos personales, Datos de investigación y Datos de preservación de los datos.

A. Datos personales

1. Edad:

A. Menor

B. Medio

C. Mayor

2. Fecha de la última publicación en el *Journal of Investigation* desde su creación hasta su publicación en el *Journal of Investigation* en los últimos 7 años:

A. Menos de 1 año

B. Entre 1 y 3 años

C. Más de 3 años


C. Preservación de los datos una vez terminado el proyecto

1. ¿Cualquier parte de los datos científicos de su investigación se ha preservado?

A. Sí

B. No

Financiado por:



Gestión de datos de investigación*

Este cuestionario se inscribe en un proyecto de investigación I+D+i cuyo objetivo es fomentar el intercambio de los datos científicos en formato digital.

Entendemos por "datos de investigación todo aquel material que ha sido registrado durante la investigación, reconocido por la comunidad científica y que sirve para certificar los resultados de la investigación que se realiza. Además de estas características también se especifica que debe provenir de una fuente única y deben ser difíciles o imposibles de obtener de nuevo" (Torres-Salinas; Robinson-García; Cabezas-Clavijo, 2012). Ejemplos: vídeos, gráficos, tablas, rawdata, datos numéricos...

Por reutilización de datos entendemos la posibilidad de crear nuevo conocimiento a partir de datos existentes, combinándolos de nuevas maneras, no imaginadas para su forma original. En un sentido estricto son más reutilizables los datos en bruto que un vídeo, por ejemplo.

Solicitamos su colaboración rellenando la encuesta que figura a continuación, que consiste en 40 preguntas divididas en 3 apartados y debería durar entre 10 y 15 minutos.

Siendo conscientes de la multitud de encuestas que llegan por correo, agradecemos que dedique unos minutos a la nuestra

* Los resultados de esta encuesta serán compartidos con proyectos similares

*Obligatorio



A. Datos personales

1. Edad *

2. Sexo *

- Hombre
 Mujer

3. Escoja de la lista la Universidad o Centro de Investigación donde haya realizado la mayor parte de su trabajo científico en los últimos 5 años *

- Universitat Politècnica de València
 Universitat de València

4. ANEXO II. INVENTARIO DE CENTROS QUE EXPONEN DATOS

La metodología de consulta en cada base de datos se ha hecho en función del tesoro y las materias de clasificación que emplean. Se interrogan los siguientes registros de bancos de datos: Databib; Odisea y Re3data.

Databib clasifica los datos por las materias Health and Medical Sciences. Odisea clasifica por las materias Clinical Medicine; Molecular Biology and Genetics; Neuroscience and Behaviour; Pharmacology and Toxicology. Re3data clasifica por las materias Medicine

1. DATABIB <http://databib.org/>

Materia: Health and Medical Sciences

Centros:

- ACEpepDB: Peptide Database
- Adult Blood Lead Epidemiology and Surveillance (ABLES) Interactive Database
- Agency for Healthcare Research and Quality (AHRQ) BioGrid Australia Limited
- Biomedical Informatics Research Network (BIRN) Brain Biodiversity Bank, The
- Breast Cancer Tissue Bank (BCTB)
- Canada's Michael Smith Genome Sciences Centre Cancer GEnome Mine (CanGEM)
- Centers for Disease Control and Prevention Data & Statistics [CDC] Clinical Trials Registry - India (CTRI)
- Comprehensive Epidemiologic Data Resource (CEDR)
- Data Bank, The
- Database of Protein Disorder (DisProt) DrugBank
- Global Health Data Exchange (GHDx) HIV Spatial Data Repository
- Histome: The Histone Infobase Health and Medical Care Archive Health and Retirement Study (HRS) Human Mortality Database (HMD) Infevers
- International Classification of Diseases
- Limited Access Datasets From NIMH Clinical Trials
- MedEffect Canada - Canada Vigilance Adverse Reaction Online Database Melanoma Molecular Map Project (MMMP)
- NIDDK Central Repository
- National Addiction and HIV Data Archive Program National Center for Health Statistics
- National Database for Autism Research (NDAR) Neuroscience Information Framework
- Open fMRI

- Orthopedic Foundation for Animals Records PsychData
- Reservoirs of Antibiotic Resistance Network (ROAR)
- Scientific Registry of Transplant Recipients
- Substance Abuse and Mental Health Services Administration (SAMHSA) SumsDB (Surface Management System Database) and WebCaret Online
- Visualization
- Surveillance Epidemiology and End Results
- Swedish Initiative for Research on Microdata in the Social And Medical Sciences (SIMSAM)
- Swedish National Data Service (SND)
- TBNET India: A National Portal for Tuberculosis Initiative The Immuno Polymorphism Database
- The fMRI Data Center
- VectorBase
- Veterinary Medical Database Visible Human Project, The
- WHO Global Health Observatory Data Repository Whole Brain Atlas
- XNAT
- caArray

2. ODISEA <http://odisea.ciepi.org/>

Materias: Clinical Medicine (11) Molecular Biology and Genetics (35) Neuroscience and Behaviour (2) Pharmacology and Toxicology (1)

Clinical Medicine

Centros:

- Africa Centre for Health and Population Studies
- <http://www.africacentre.ac.za/Default.aspx?tabid=69>
- Hon Media <http://www.hon.ch/HONmedia/>
- Hardin.MD. Hardin Mate Directory.
- <http://www.lib.uiowa.edu/hARDIN/MD/ped.html>
- Wellcome Images. <http://images.wellcome.ac.uk/>
- Pediatric Surgery @ Brown. <http://med.brown.edu/pedisurg/Brown/IBCategories.html>
- **fMRIDC Data Center.** <http://www.fmridc.org/fmridc>
- Okayama University Scientific Achievement Repository. <http://ousar.lib.okayama-u.ac.jp/Index.e>
- Biblioteca Virtual em Saúde. <http://bvsmis.saude.gov.br/php/index.php>

- Clinical Trial. <http://clinicaltrials.gov/>
- Clinicals Trial Registry India. <http://ctri.nic.in/Clinicaltrials/login.php>
- Cancer GenomeAtlas. <https://tcga-data.nci.nih.gov/tcga/>

Molecular Biology and Genetics

- Biological General Repository for Interaction Datasets (BioGRID) <http://thebiogrid.org/>
- Ensembl Bacteria <http://bacteria.ensembl.org/index.html>
- Ensembl Plants <http://plants.ensembl.org/index.html>
- Ensembl Fungi <http://fungi.ensembl.org/index.html>
- Ensembl Protists <http://protists.ensembl.org/index.html>
- Ensembl Metazoa <http://metazoa.ensembl.org/index.html>
- TARDIS (The Australian Repositories for Diffraction ImageS) <http://tardis.edu.au/>
- RSCB PDB <http://www.rcsb.org/pdb/home/home.do>
- Biological Magnetic Resonance Data Bank (BMRB) <http://www.bmrwisc.edu/>
- ProteinDataBank Europe <http://www.ebi.ac.uk/pdbe/>
- World Wide Protein Data Bank <http://www.wwpdb.org/>
- GenBank <http://www.ncbi.nlm.nih.gov/>
- Gene Expression Omnibus (GEO), <http://www.ncbi.nlm.nih.gov/projects/geo/>
- UniSave <http://www.ebi.ac.uk/uniprot/unisave/>
- Genome Reviews database <http://www.ebi.ac.uk/GenomeReviews/>
- human proteomics initiative (HPI) <http://www.uniprot.org/program/Chordata>
- IntEnz <http://www.ebi.ac.uk/intenz/index.jsp>
- PANDIT: Protein and Associated NucleotideDomains with Inferred Trees <http://www.ebi.ac.uk/goldman-srv/pandit/>
- CSA <http://www.ebi.ac.uk/thornton-srv/databases/CSA/>
- EBI Metagenomics <https://www.ebi.ac.uk/metagenomics/>
- CluSTr <http://www.ebi.ac.uk/clustr/>
- IMGT/HLA Databas <http://www.ebi.ac.uk/imgt/hla/>
- Immuno Polymorphism Database (IPD) <http://www.ebi.ac.uk/ipd/>
- HUGO Gene Nomenclature Committee (HGNC) <http://www.genenames.org/>
- Interpro <http://www.ebi.ac.uk/interpro/>

- PRoteomics IDentifications database (PRIDE) <http://www.ebi.ac.uk/pride/>
- Database of Genomic Variants archive (DGVA) <http://www.ebi.ac.uk/dgva/>
- Ensembl <http://www.ensembl.org/index.html>
- Ensembl Genomes <http://www.ensemblgenomes.org/>
- The European Genome-phenome Archive (EGA) <https://www.ebi.ac.uk/ega/>
- Uniprot <http://www.uniprot.org/>
- European Nucleotide Archive ENA <http://www.ebi.ac.uk/ena/>
- EM Data Bank <http://emdatbank.org>
- DNA Data Bank of Japan (DDBJ) <http://www.ddbj.nig.ac.jp/>
- ArrayExpress <http://www.ebi.ac.uk/arrayexpress/>

Neuroscience and Behaviour

- Flytrap <http://www.fly-trap.org/>
- NeuroGrid <http://www.neurogrid.ac.uk/>

Pharmacology and Toxicology

- Australian New Zelan Clinical Trials Registry <http://www.anzctr.org.au/>

3. RE3DATA <http://www.re3data.org/>

Materia: Medicine

Resultados: 115 repositories

[http://service.re3data.org/search/results/filter?term=&d=25&filter_subjects_active\[medicine\]=Medicine](http://service.re3data.org/search/results/filter?term=&d=25&filter_subjects_active[medicine]=Medicine)

- Banco de Información para la Investigación Aplicada en Ciencias Sociales, Repositorio Institucional
- ClinicalTrials.gov
- **Danish Data Archive**
- DSMZ Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures Ecological Society of America esa Ecological Archives
- NeuroMorpho
- Neuroscience Information Framework
- World Data Center of Microorganisms
- ArrayExpress
- Centers for Disease Control and Prevention, Data & Statistics
- CSIRO Data Access Portal
- Database of Genomic Variants

- Data Sharing for Demographic Research
- DRYAD
- Eumetsat
- eyeMoviePedia
- GenBank
- Gene Expression Atlas
- GigaDB
- GWAS Central
- Health and Medical Care Archive
- IFPRI Datasets
- International Mouse Strain Resource
- Mouse Genome Informatics
- National Archive of Computerized Data on Aging
- National Center for Ecological Analysis and Synthesis Data Repository
- NCBI
- Nucleic Acid Database
- Online Offering of the Federal Centre for Health Education
- PHI-base
- RCSB Protein Data Bank
- Research Data Centre of the DZA
- Research Data Centres of the Federal Statistical Office and the statistical offices of the Länder
- Robert Koch Institut, Health Monitoring
- South African Data Archive
- Substance Abuse and Mental Health Data Archive
- Survey of Health, Ageing and Retirement in Europe
- Swedish National Data Service
- The Dataweb Dataverse
- UCD Digital Library
- UPF Digital Repository - Recursos i dades primàries
- VegBank
- 1000 Genomes
- Addgene
- Agency for Healthcare Research and Quality, Data & Surveys
- ARS
- Beta Cell Biology Consortium Reagent and Data Resources
- Brain Biodiversity Bank
- Bugwood Image Database System
- California Vectorborne Disease Surveillance System
- Canada's Michael Smith Genome Sciences Centre
- caNanoLab
- CanGEM
- Cell Centered Database
- ChemSpider
- Comprehensive Epidemiological Data Resource
- CyberCell Database
- DARIS
- DNA Data Bank of Japan
- DrugBank 4.0 Beta
- EcoGene
- Encyclopedia of Life
- Ensembl
- Environmental data explorer
- European Nucleotide Archive
- European Social Survey
- GermOnline
- Global Biodiversity Information Facility, Data Portal
- Global Catalogue of Microorganisms
- Global Health Data Exchange
- Global Initiative on Sharing Avian Influenza Data
- ICRISAT Dataverse Network
- InGeoCloudS
- International Long Term Ecological Research
- <odesi>
- Mansfeld's World Database of Agriculture and Horticultural Crops
- Melanoma Molecular Map Project
- MIRAGE
- MolTable
- Mouse Phenome Database
- MyTardis
- National Addiction & HIV Data Archive Program
- National Ecological Observatory Network
- National Forestry Database
- NCBI Reference Sequences
- NDAR
- NIDDK Central Repository

- Online Mendelian Inheritance in Man
- openLandscapes
- Open Research Data
- Peptidome
- Protein Data Bank in Europe
- RunMyCode
- SumsDB
- The Antibody Registry
- The Knowledge Network for Biocomplexity
- The Long Term Ecological Research Network
- UCSC Genome Bioinformatics
- UK-AIR
- United States Census Bureau
- Universal PBM Resource for Oligonucleotide Binding Evaluation
- US Transplant - Scientific Registry of Transplant Recipients
- Washington State University College of Pharmacy United States Transuranium & Uranium Registries
- West African Vegetation
- WestNile.ca.gov
- caArray
- ChEMBL
- German Centre for Cancer Registry Data
- Infevers
- Open Rotterdam Glaucoma Imaging Data Sets
- Pseudobase
- QTL Archive
- The Electron Microscopy Data Bank at PDBe
- The whole brain Atlas
- Karlsruhe Astrophysical Database of Nucleosynthesis in Stars

5. ANEXO III. GUÍA DE BUENAS PRÁCTICAS PARA GESTIONAR Y COMPARTIR DATOS. GUÍA PARA INVESTIGADORES Y CENTROS DE INVESTIGACIÓN

Índice

Prólogo

CAPÍTULO 1. ¿Por qué compartir los datos de investigación?

CAPÍTULO 2. Los organismos financiadores de la investigación frente a los datos abiertos de investigación

CAPÍTULO 3. ¿Cómo compartir los datos?

- 3.1. Algunas consideraciones previas
- 3.2. Posibilidades para el depósito de datos brutos de investigación
- 3.3. ¿En qué repositorios puedo depositar mis datos?
- 3.4. Características de los depósitos de datos
- 3.5. Consejos para elegir un depósito para los datos
- 3.6. Un ejemplo práctico en Dryad

CAPÍTULO 4. El acceso abierto a los datos en las revistas científicas

CAPÍTULO 5. La necesidad de proteger a las partes implicadas

CAPÍTULO 6. Aspectos éticos y legales

CAPÍTULO 7. Encuesta sobre la gestión de los datos de investigación de los investigadores españoles en Ciencias de la Salud.

CAPÍTULO 8. Directrices para el futuro y conclusiones

CAPÍTULO 9. Glosario

CAPÍTULO 10. Recursos en Internet

CAPÍTULO 11. Bibliografía

ANEXO 1. Organismos internacionales

ANEXO 2. Recomendaciones para los centros de investigación sanitarios para fomentar el uso compartido de datos (Piwowar, 2008)

ANEXO 3. Argumentos comúnmente aceptados en nombre de Open Data (Murray-Rust)

PRÓLOGO

La investigación científica puede mejorarse mediante la gestión y el intercambio de los datos de la investigación. Una buena gestión de los datos permite generar nuevos conocimientos, formular nuevas hipótesis y verificar con certeza los resultados de la investigación.

Los principios del uso compartido de datos (*data sharing* en terminología sajona) están teniendo cada vez un mayor respaldo por parte de las agencias financiadoras de la investigación, dada la creciente preocupación por obtener el mayor retorno posible de la inversión, tanto en términos de la calidad de los productos derivados de la investigación como en la reutilización de datos de la investigación.

Así, en los EE.UU., la National Science Foundation requiere que las solicitudes de subvención incluyan un plan de gestión de datos. Los National Institutes of Health han invertido en la infraestructura necesaria para compartir datos a gran escala y a largo plazo y han desarrollado normativas y directrices al respecto y lo mismo ha hecho Wellcome Trust y el Medical Research Council en el Reino Unido.

En España, el Ministerio de Economía y Competitividad subvenciona el proyecto “Opendatascience, centro de recursos para la preservación y gestión de datos abiertos de investigación”, que tiene como objetivo fomentar el intercambio y reutilización de los datos primarios de investigación de forma generalizada y normalizada. También las fundaciones privadas están interesadas en investigar en este campo y, por ejemplo, la Fundación Mapfre subvenciona el proyecto “Estrategias para la gestión y la promoción del uso compartido de datos científicos sobre la salud”, con propósitos similares al anterior pero centrados en el campo de las ciencias de la salud.

Dado este entorno en rápida evolución, el grupo OpenDataScience del VLC Valencia International Campus of Excellence, compuesto por investigadores del CSIC, la Universitat de València y la Universitat Politècnica de Valencia, en colaboración con la Universidad Católica de Valencia y otras instituciones, considera que es una prioridad dar respuesta a estos requisitos y apoyar a los investigadores y gestores de la investigación con el fin último de promover la buena gestión y el uso compartido de los datos de investigación en España.

El grupo OpenDataScience ha emprendido un estudio piloto sobre el estado del intercambio de datos en España, centrado en analizar los siguientes aspectos: las prácticas actuales de los investigadores en relación con la gestión y el intercambio de datos; las necesidades de apoyo para gestionar y compartir sus datos; el desarrollo

de infraestructuras necesarias para poner en práctica; la orientación práctica y la formación en este campo; las normativas necesarias para la citación de los datos y sus publicaciones matrices para aumentar la visibilidad y la accesibilidad de los datos y de la propia investigación; los aspectos éticos y legales relacionados con la protección de la privacidad y del copyright.

Si los datos de investigación están bien organizados, documentados, conservados y accesibles, y su precisión y validez están controlados en todo momento, el resultado será disponer de datos de alta calidad, una investigación eficiente, resultados basados en pruebas sólidas y ahorro de tiempo y de recursos. Este procedimiento debe planearse antes de que comience la investigación y no necesariamente va a suponer mucha dedicación si se integra como práctica habitual de los investigadores y de los grupos de investigación.

La responsabilidad de la gestión de datos recae principalmente en los investigadores, pero las instituciones pueden proporcionar la orientación, herramientas, infraestructura y personal de apoyo necesario para ayudar en muchas facetas relacionadas con la gestión de los datos.

Las directrices que proporciona esta guía tiene como objetivo ayudar a los investigadores y administradores de datos a desarrollar estrategias eficaces para maximizar el valor de sus datos de investigación. Pretendemos dar respuesta, a partir de nuestra experiencia, a los problemas que pueden plantearse o que pueden surgir durante el proceso de gestión de sus datos.

Esta guía se complementa con la información contenida en la web del grupo (www.dataea.com; www.healthdata-sea.com), así como con la que aportan los enlaces incluidos en la sección de recursos y en la bibliografía.

CAPÍTULO 1. ¿POR QUÉ COMPARTIR LOS DATOS DE INVESTIGACIÓN?

Los datos brutos de investigación son recursos muy valiosos que pueden ser utilizados y reutilizados en el futuro con fines científicos y educativos.

Compartir datos facilita la formulación de nuevas investigaciones, evita la duplicación de experimentos y proporciona recursos para la educación y la formación.

¿Por qué compartir los datos de investigación?

Los datos de investigación son un recurso valioso con un gran valor que por lo general requieren mucho tiempo y dinero para ser producidos. Muchos datos tienen un valor significativo fuera del uso de la investigación inicial. Las virtudes de los datos abiertos y las ventajas de compartirlos están hoy en día firmemente establecidas

en la literatura, a pesar de que su implementación sea todavía limitada.

La facilidad con la que hoy en día se pueden almacenar y difundir online los datos digitales, posibilita que muchas instituciones estén dispuestas a compartir los datos de investigación para aumentar el impacto y la visibilidad de sus investigaciones.

Ventajas de compartir los datos de investigación

- Fomenta la investigación y el debate científico.
- Promueve la innovación, pues los investigadores tienen mayores oportunidades para descubrir nuevos conocimientos.
- Aumenta el número de análisis estadísticos ciegos.
- Estimula los descubrimientos adicionales.
- Posibilitan establecer a nuevas colaboraciones entre los usuarios de los datos y sus creadores.
- Maximiza la transparencia.
- Permite la comprobación de los resultados de la investigación, previniendo los fraudes.
- Fomenta la mejora y validación de los métodos de investigación.
- Reduce el coste de nuevas investigaciones, pues evita la repetición de costosos proyectos que utilizan los mismos medicamentos o aparatos que se utilizaron previamente.
- Aumenta la visibilidad y el impacto y de la investigación.
- Promueve la investigación que generó los datos y sus resultados.
- Puede proporcionar mayor crédito al investigador que ha creado los datos.
- Proporciona recursos para la educación y la formación.

CAPÍTULO 2. LOS ORGANISMOS FINANCIADORES DE LA INVESTIGACIÓN FRENTE A LOS DATOS ABIERTOS DE INVESTIGACIÓN

Los organismos financiadores de la investigación cada vez más siguen la orientación de la Organización para el Desarrollo Económico y la Cooperación (OCDE) sobre que los datos en los que se basan las investigaciones financiadas con fondos públicos deben ser en lo posible abiertos y puestos a disposición de la comunidad. En junio de 2013, la carta pública de los líderes del G8 indicaba: “los datos abiertos son un recurso sin explotar con un enorme potencial para fomentar la construcción de sociedades más fuertes, más interconectadas que satisfagan mejor las necesidades de nuestros ciudadanos y permitir que florezca la innovación y la prosperidad”.

Por su parte, Horizon 2020, el VII Programa Marco de la Unión Europea, prevé que el acuerdo de subvención de los proyectos financiados por Consejo Europeo de Investigación, se podrán establecer las condiciones en las que se deberá ofrecer el acceso abierto a los resultados de los proyectos.

Existen numerosas iniciativas mundiales en este campo y también en España se están llevando a cabo proyectos destinados a fomentar el intercambio y reutilización de los datos primarios de investigación. En España, el

Grupo de Trabajo de “Depósito y Gestión de datos en Acceso Abierto” del proyecto RECOLECTA, ha elaborado un informe que refleja los aspectos relevantes que intervienen en la gestión de los datos de investigación en los repositorios científicos. Digital CSIC, el repositorio de publicaciones y de datos del Consejo Superior de Investigaciones Científicas, también pone a disposición de los investigadores una guía de buenas prácticas y políticas de datos de investigación (<http://digital.csic.es/politicas/politicaDatos.jsp>).



www.digital.csic.es

Por otra parte, el Ministerio de Economía y Competitividad financia el proyecto “Opendatascience, centro de recursos para la preservación y gestión de datos abiertos de investigación” (<http://www.datasea.es/dt/>), y la Fundación Mapfre el proyecto “Estrategias para la gestión y la promoción del uso compartido de datos científicos sobre la salud”, (<http://www.healthdatasea.es/>). Entre las acciones de estos proyectos destaca las campañas de sensibilización mediante la edición de guías de buenas prácticas

destinadas a convencer a autores y editores de la conveniencia de liberar y compartir los datos brutos de investigación para un aprovechamiento más óptimo. También es necesario poner al servicio de los investigadores el asesoramiento técnico y las infraestructuras necesarias para que puedan llevar a cabo estas prácticas, así como la elaboración de una normativa que regule todo este proceso para evitar el uso indebido o fraudulento.

Publicar y compartir los datos científicos



G. BOULTON, IAC, CONSERVATORIA

WORLD VIEW

A personal take on events

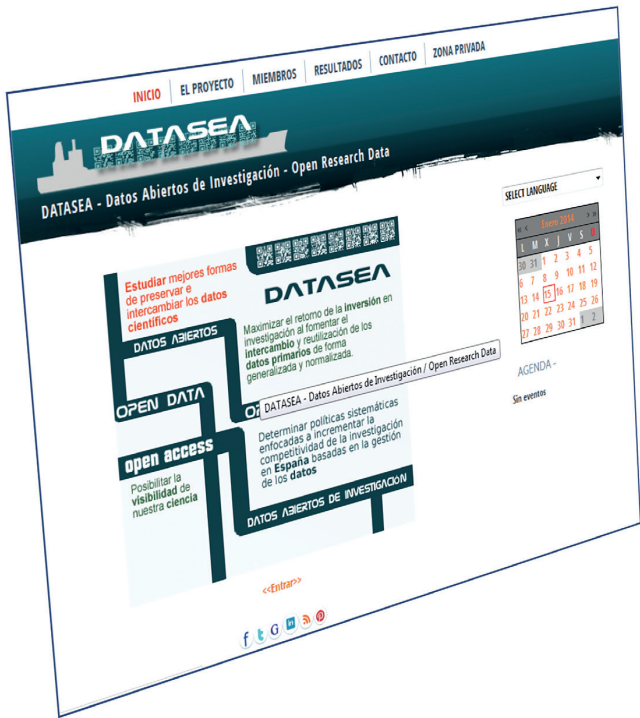
Open your minds and share your results

An open approach is the best way to maximize the benefits of research for both scientists and the public, says Geoffrey Boulton.

SCIENCE'S CAPACITY FOR SELF-CORRECTION COMES FROM ITS OPENNESS TO SCRUTINY AND CHALLENGE.

28 JUNE 2012 | VOL 485 | NATURE | 441

Web del proyecto Opendatascience



www.datasea.es

Web del proyecto Healthdatasea



www.healthdatasea.es

CAPÍTULO 3. ¿CÓMO COMPARTIR LOS DATOS?

3.1. Algunas consideraciones previas

Como es lógico, los datos que se vayan a compartir deben ser de una calidad suficientemente alta como para que tengan valor para otros investigadores. Además, deberán proporcionarse en un formato adecuado para que puedan ser reutilizados con eficacia. Las características y especificaciones sobre los formatos se tratarán en otro capítulo.

Los datos deben ser compartidos aplicando las normas establecidas que maximicen las oportunidades de vinculación y la interoperabilidad de los datos. Deberán proporcionarse también suficientes metadatos para que sea posible la utilización por otros. La asignación de estos metadatos también deberá adoptarse de acuerdo con las normas acordadas.

En definitiva, al considerar la gestión y el intercambio de datos científicos, deben tenerse en cuenta las siguientes preguntas:

¿Cómo compartir datos?

ADDITIONAL MATERIAL ON THE WEB
Rev Esp Cardiol accepts additional electronic material to support and improve the presentation of your scientific research. The publication of online-only Supplemental material will remain at the discretion of the Editorial team. Only material directly relevant to the article content will be considered for e-publication. This material will not be translated, typeset or proofread.
 To ensure that your material has the appropriate format, we recommend the following standards:

A. Plataformas digitales de las revistas

	Format	Extension	Details
Document	Word	Doc o Docx	Maximum size 300 Kb
Image	JPG	.jpg	Maximum size 10 MB
Audio	MP3	.mp3	Maximum size 10 MB
Video	MPG	.mpg	Maximum size 50 MB

Authors should submit the material in electronic form through the EES as a multimedia file along with the article. Each file should have a concise and descriptive title.
 The e-material must also meet all requirements and general ethical responsibilities described in these standards.
 The Editorial Board reserves the right to refuse electronic material it deems not appropriate.

1. ¿Qué datos generará la investigación?
2. ¿Qué datos tendrán valor para otros investigadores y, por lo tanto, pueden compartirse?
3. ¿En qué formato deben generarse los datos que desea compartir?
4. ¿Qué estándares de calidad deben aplicarse a los datos para que se puedan compartir eficazmente?
5. ¿Dónde va a depositar los datos disponibles?
6. ¿Qué límites va a poner para que otros investigadores puedan acceder a los datos?
7. ¿Qué tipo de salvaguardia o de protección de la propiedad intelectual desea?
8. ¿Cómo va a garantizar que los conjuntos de datos clave se conservan para asegurar su valor a largo plazo?
9. ¿Qué recursos requiere para poner en marcha su plan de depósito?

Cuando lo consideren apropiado, los investigadores pueden establecer una moratoria en el depósito de sus datos para facilitar el intercambio previo con otros investigadores. Cualquier restricción en este sentido debe ser razonable, transparente y en línea con las mejores prácticas establecidas.

3.2. Posibilidades existentes para el depósito de datos brutos de investigación

En España todavía no existen políticas y normativas para la gestión y reutilización de los datos. Tampoco se han desarrollado infraestructuras técnicas preparadas para el almacenamiento de datos brutos de investigación, a excepción de Digital CSIC, del Consejo Superior de Investigaciones Científicas, la Universidad Pompeu Fabra y la Fundación Juan March (Nina et al, 2013).

Las posibilidades existentes, según Nina et al. (2013) son las siguientes:

- Almacenamiento en las plataformas de las ediciones electrónicas de las revistas, habitualmente bajo la categoría de “material suplementario”.
- Almacenamiento en repositorios institucionales específicos para datos, especializados por disciplinas o de carácter multidisciplinar.
- Almacenamiento en la web del autor o en herramientas generales en la nube como Drive o Dropbox.

Los investigadores suelen preferir el almacenamiento en las plataformas de las revistas y en repositorios institucionales. Sin embargo, estas modalidades no tienen suficientes garantías de preservación a largo plazo por dos motivos. Por una parte, porque el material complementario no suele tener suficiente interés, por ahora, para las editoriales, que dudan de su capacidad para generar beneficios económicos. Por otra, porque la información se dispersa en múltiples fuentes no interoperables técnicamente. Entre las ventajas cabe destacar la mayor facilidad para localizarlos y entenderlos (Arguimbau, 2013).

Sin embargo, los expertos prefieren almacenar los datos en repositorios, sobre todo institucionales, ya que poseen dos ventajas de las que carecen los otros: su interoperabilidad y la existencia de una institución responsable que garantiza su calidad y su continuidad.

3.3. ¿En qué repositorios puedo depositar mis datos?

Todavía existen pocos repositorios de datos científicos, incluso a escala mundial. Un ejemplo en España es *Digital.CSIC* (<https://digital.csic.es/>), depósito de documentos digitales cuyo objetivo es organizar, archivar, preservar y difundir en modo de acceso abierto la producción intelectual resultante de la actividad investigadora del Consejo Superior de Investigaciones Científicas.



En estos momentos en que apenas se han desarrollado repositorios apropiados a las necesidades de todos los colectivos, la opción más práctica viable para los investi-

gadores e instituciones que no disponen de repositorios institucionales son los ya existentes de carácter multidisciplinar como Dryad, Figshare, Zenodo y Dataverse.



La descripción de las principales características de cada uno de ellos viene explicada con detalle en el artículo de Nina et al. (2013).

3.4. Características de los depósitos de datos

Los repositorios de datos tienen las siguientes características comunes:

1. Suelen aceptar tanto texto como los datos brutos que lo acompañan.
2. Estos datos son recogidos en los directorios de acceso abierto como ROAR, Openarchives.org community, OpenDOAR y Zenodo.
3. Llevan asignado un DOI, lo que los permite identificarlos bibliográficamente y citarlos.
4. Es posible su exportación a gestores de referencias bibliográficas.
5. Proporcionan información estadística basada en métricas alternativas (tuiteos en Figshare o Zenodo).
6. Utilizan herramientas de la web 2.0.
7. Mencionan las licencias de uso de los datos, generalmente del tipo Creative Commons 0 (CC0).

3.5. Consejos para elegir un depósito para los datos

Algunos consejos para elegir el depósito adecuado para los datos (Nina y cols., 2013) son:

1. Los datos asociados a los artículos deberían ser depositados en *Dryad*, que comenzó orientado a ciencias de la salud y ahora es un proyecto multidisciplinar, sólido y apoyado por la comunidad bibliotecaria y editorial.

2. Los datos sin publicación o derivados de financiación europea, deberían ser depositados en *Zenodo*, proyecto con gran proyección en la Unión Europea que garantiza la preservación de los datos a largo plazo.
3. Si una institución debe dar soporte inmediatamente a sus investigadores de forma corporativa, debería crearse una comunidad en *Zenodo* o en *Dataverse*. Una buena opción sería *Zenodo* para todas las disciplinas excepto ciencias sociales, en las que *Dataverse* tiene ya una comunidad de usuarios creada.
4. Los investigadores del CSIC pueden depositar sus datos en la plataforma de su institución, digital CSIC.

3.6. Dos ejemplos prácticos en Dryad y Figshare

En la siguiente captura se muestra la página de inicio de Dryad. En la parte izquierda se sitúa la aplicación para el envío de datos (*Submit data now*), con un acceso a una lista de preguntas frecuentes o instrucciones (*how and why?*) sobre el funcionamiento de Dryad (*how and why?*):

a) Sobre el envío de datos:

- Por qué utilizar Dryad
- Qué tipo de datos acepta
- Cuando se pueden enviar los datos
- Cómo se pueden modificar los datos previamente enviados a Dryad
- Qué preparativos son necesarios previos al envío
- Cómo escribir el archivo "Léame" (*ReadMe file*).
- Cómo citar los datos en mi propio manuscrito
- Cómo enviar los datos
- Qué costes lleva asociado el proceso



- Qué sucede con los datos una vez enviados
 - Cómo seguir las estadísticas de uso y descarga de los datos depositados
- b) Sobre el uso de los datos:
- Cómo se pueden reutilizar los datos de Dryad
 - Si hay que pagar para descargar o utilizar los datos
 - Cómo citar datos de Dryad
 - Cómo se puede descargar información a un software de gestión de referencias
- c) Información General
- Qué es un paquete de datos
 - Procedimientos que pueden seguir las revistas para incluirse en el repositorio
 - Qué es un Dryad DOI
 - Por qué utiliza Dryad Creative Commons Zero
 - Cómo Dryad puede ayudar a preparar un plan de gestión de datos
 - Qué cargos tiene el envío de datos
 - Qué revistas están integradas en Dryad

El archivo ReadMe file está destinado a ayudar a asegurar que los datos se interpreten correctamente. Puede presentarse un único ReadMe file para el paquete de datos en su conjunto o bien proporcionar el anterior junto a otros independientes para cada archivo de datos individual. El ReadMe file debe ser un archivo de texto que contenga información sobre una breve descripción de lo que incluyen los datos, definiciones, códigos de datos, unidades de medida, pasos a seguir en el procesamiento de los datos, otros lugares donde se almacenan los datos y persona o institución de contacto para resolver problemas.

Un paquete de datos es el conjunto de archivos de datos asociados a una publicación, junto con los metadatos que los describen.

Un Dryad DOI es un identificador digital, permanente y único, que se asigna a cada paquete de datos. Es un componente importante de una citación de datos pues se recomienda utilizarlo cada vez que hace referencia a un paquete de datos o archivo de datos.

En relación con los cargos por el envío de datos, Dryad es una organización sin fines de lucro que proporciona acceso a largo plazo a su contenido, sin costo para los investigadores, los educadores o los estudiantes, independientemente de su nacionalidad o afiliación institucional.

Página de inicio de Dryad (www.dryad.com)

Debajo de la aplicación para el envío de datos se sitúa el buscador Search for data con la posibilidad de activar el sistema de búsqueda avanzado que permite limitar por

los campos título, autor, materia y fecha de publicación (*Add refinement*), así como filtrar los resultados (*Refine search*).

La siguiente figura presenta capturas de una búsqueda realizada con la palabra “infarct”.

The screenshot shows a search interface with the following elements:

- Buscar** (Search) section:
 - Search input: "infarct" with a green "Ir" button.
 - Advanced Search options: "Resultados/página" set to 10, "Ordenar ítems por" set to "pertinencia", and "en orden" set to "descendente".
 - Showing items 1-2 de 2.
 - Filters: "Dryad Data Packages (2)", "KNB (0)", and "TreeBASE (0)".
 - Search results:
 - Item 1: Dalal HM, Wingham J, Palmer J, Petre C, Taylor R, Lewin R (2012) Data from: Why do so few patients with heart failure participate in cardiac rehabilitation? A cross-sectional survey from England, Wales and Northern Ireland. *BMJ Open* doi:10.5061/dryad.n6661sh1
 - Item 2: van Gils PF, Over EAB, Hamberg-van Reenen HH, de Wit GA, van den Berg M, Engelfriet PM (2011) Data from: The polypill in the primary prevention of cardiovascular disease: cost-effectiveness in the Dutch population. *BMJ Open* doi:10.5061/dryad.qg54c627
- Refine Search** section:
 - Author**: Dalal, Hasnain M. (1), de Wit, G. Ardine (1), Engelfriet, Peter M. (1), Engelfriet, Peter (1), Hamberg, Heleen (1), Hamberg-van Reenen, Heleen H. (1), Lewin, Robert (1), Over, Eelco A. B. (1), Over, Eelco (1), Palmer, Joanne (1), ... View More
 - Subject**: Adult cardiology (1), Health economics (1), Heart failure (1), Myocardial infarction (1), Public health (1), Quality in health care (1), Stroke (1)

La siguiente figura presenta capturas de uno de los registros obtenidos en la búsqueda anterior.

The screenshot shows a data record with the following information:

- Data from:** The polypill in the primary prevention of cardiovascular disease: cost-effectiveness in the Dutch population. **BMJ Open**
- Files in this package:**
 - Content in the Dryad Digital Repository is offered "as is." By downloading files, you agree to the [Dryad Terms of Service](#). To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.
 - Title:** Polypill model
 - Downloaded:** 46 times
 - Download:** [Polypill model.nb \(403.2Kb\)](#)
 - Details:** [View File Details](#)

The screenshot shows the abstract for the study:

- DOI:** doi:10.5061/dryad.qg54c627
- Pageviews:** 450
- Keywords:** Myocardial infarction, Health economics, Public health, Stroke
- Date Submitted:** 2011-11-22T16:42:12Z
- Abstract:**

OBJECTIVES: The aim of the present study was to estimate the cost-effectiveness of the polypill in the primary prevention of cardiovascular disease. **DESIGN:** A health economic modeling study. **SETTING:** Primary health care in the Netherlands. **PARTICIPANTS:** Simulated individuals from the general Dutch population, aged 45-75 years of age. **INTERVENTIONS:** Opportunistic screening followed by prescription of the polypill to eligible individuals. Eligibility was defined as having a minimum 10-year risk of cardiovascular death as assessed with the SCORE function of alternatively 5%, 7.5%, or 10%. Different versions of the polypill were considered, depending on composition: 1) the Indian polycap, with three different types of blood pressure lowering drugs, a statin, and aspirin; 2) as 1) but without aspirin; 3) as 2) but with a double statin dose. In addition, a scenario of (targeted) separate antihypertensive and/or statin medication was simulated. **PRIMARY OUTCOME MEASURES:** Cases of acute myocardial infarction or stroke prevented, QALYs gained, and the costs per QALY gained. All interventions were compared with usual care. **RESULTS:** All scenarios were cost-effective with an incremental cost-effectiveness ratio between €7,900-12,300 per QALY compared with usual care. Most health gains were achieved with the polypill without aspirin and containing a double dose of statins. With a 10-year risk of 7.5% as threshold, this pill would prevent approximately 3.5% of all cardiovascular events. **CONCLUSIONS:** Opportunistic screening based on global cardiovascular risk assessment followed by polypill prescription to those with increased risk

The screenshot shows citation instructions:

- When using this data, please cite the original article:
 - van Gils PF, Over EAB, Hamberg-van Reenen HH, de Wit GA, van den Berg M, Schuit AJ, Engelfriet PM (2011) The polypill in the primary prevention of cardiovascular disease: cost-effectiveness in the Dutch population. *BMJ Open* 1(2): e000363. doi:10.1136/bmjopen-2011-000363
- Additionally, please cite the Dryad data package:
 - van Gils PF, Over EAB, Hamberg-van Reenen HH, de Wit GA, van den Berg M, Engelfriet PM (2011) Data from: The polypill in the primary prevention of cardiovascular disease: cost-effectiveness in the Dutch population. Dryad Digital Repository. doi:10.5061/dryad.qg54c627
- [Cite](#) | [Share](#)



Se presentan a continuación dos capturas que muestran los metadatos incluidos en el registro anterior.

dc.contributor.author	van Gils, Paul F.
dc.contributor.author	Over, Eelco A. B.
dc.contributor.author	Hamberg-van Reenen, Heleen H.
dc.contributor.author	de Wit, G. Ardine
dc.contributor.author	van den Berg, Matthijs
dc.contributor.author	Engelfriet, Peter M.
dc.date.accessioned	2011-11-22T16:42:12Z
dc.date.available	2011-11-22T16:42:12Z
dc.date.issued	2011-12-21
dc.identifier	doi:10.5061/dryad.qg54c627
dc.identifier.citation	van Gils PF, Over EAB, Hamberg-van Reenen HH, de Wit GA, van den Berg M, Schuit AJ, Engelfriet PM (2011) The poly pill in the primary prevention of cardiovascular disease: cost-effectiveness in the Dutch population. <i>BMJ Open</i> 1(2): e000363.

dc.relation.haspart	doi:10.5061/dryad.qg54c627/2
dc.relation.isreferencedby	doi:10.1136/bmjopen-2011-000363
dc.relation.isreferencedby	PMID:22189351
dc.subject	Myocardial infarction
dc.subject	Health economics
dc.subject	Public health
dc.subject	Stroke
dc.title	Data from: The poly pill in the primary prevention of cardiovascular disease: cost-effectiveness in the Dutch population
dc.type	Article
dc.contributor.correspondingAuthor	van Gils, Paul F.
prism.publicationName	BMJ Open
internal.workflow.submitted	true



Figshare también es un repositorio en el que se pueden alojar los datos de investigación y ponerlos a disposición de la comunidad científica de manera citable y reconocible. La siguiente figura presenta una captura de una

búsqueda en Figshare con el término “Alzheimer” que muestra algunos de los registros obtenidos. Como se ve en la captura, junto a la columna de los autores se sitúa el tipo de archivo: (dataset y figura, en este caso).

The screenshot shows the Figshare search results page. At the top, there is a search bar with the text 'alzheimer' and a 'Search' button. Below the search bar, it says 'You searched for: alzheimer Found: 1563 results'. The results are displayed in a table with columns for 'Authors', 'Type', and 'Date'. Three results are visible:

Authors	Type	Date
A. Kulminski K. Arbeev more...	Dataset	30/01/2014
N. Sasaoka M. Sakamoto more...	Fileset	29/01/2014
N. Sasaoka M. Sakamoto more...	Figure	29/01/2014

La siguiente captura muestra el resultado de una búsqueda en figshare. Los principales componentes del registro son, de izquierda a derecha: ampliar (enlarge), descargar (download), nº de visitas (views), nº de veces que se ha compartido (shares), aplicaciones en las que

se ha compartido (facebook, twitter, google +, enclavar (embed), como citar la referencia (cite this), la descripción del archivo (en este caso un póster), las áreas temáticas (categories) y los autores.

Using Real-Time fMRI to Control a Dynamical System by Brain Activity Classification

[Enlarge](#) [Download](#)

28
views

12
shares

cites
coming
soon

Share this: [f Share](#) 0 [Tweeet](#) 0 [G+1](#) 12 [Embed*](#)

Cite this: Eklund, Anders (2013): Using Real-Time fMRI to Control a Dynamical System by Brain Activity Classification. figshare.
<http://dx.doi.org/10.6084/m9.figshare.707014>

*The embed functionality can only be used for non commercial purposes... [more](#)

Description

The poster of our MICCAI paper "Using Real-Time fMRI to Control a Dynamical System by Brain Activity Classification"

Categories

- [Bioinformatics](#)
- [Neuroscience](#)
- [Applied Computer Science](#)

Authors

[Anders Eklund](#)

Published on 25 May 2013 - 20:55 (GMT)
 Filesize is 1.29 MB

Otros componentes de los registros de figshare son el resumen (abstract), las palabras clave (tags), las aplicaciones para la gestión de las referencias (Reference

Manager, Endnote y Mendeley), enlaces (links), comentarios (comments) y el código QR

Abstract

We present a method for controlling a dynamical system using real-time fMRI. The objective for the subject in the MR scanner is to balance an inverted pendulum by activating the left or right hand or resting. The brain activity is classified each second by a neural network and the classification is sent to a pendulum simulator to change the force applied to the pendulum. The state of the inverted pendulum is shown to the subject in a pair of VR goggles. The subject was able to balance the inverted pendulum during several minutes, both with real activity and imagined activity. In each classification 9000 brain voxels were used and the response time for the system to detect a change of activity was on average 2-4 seconds. The developments here have a potential to aid people with communication disabilities, such as locked in people. Another future potential application can be to serve as a tool for stroke and Parkinson patients to be able to train the damaged brain area and get real-time feedback for more efficient training.

Links

- <http://liu.diva-portal.org/smash/get/diva2:297924/FULLTEXT01>
- <https://www.youtube.com/watch?v=HJL7j-uVqxA>
- http://link.springer.com/chapter/10.1007%2F978-3-642-04268-3_123

Comments (0)


You must be [logged in](#) to post comments.

Tags

- [fMRI](#)
- [Real-time](#)
- [Brain computer interface](#)

Export

- [Export to Ref. Manager](#)
- [Export to Endnote](#)
- [Export to Mendeley](#)





GESTIÓN DE DATOS DE INVESTIGACIÓN: INFRAESTRUCTURAS PARA SU DIFUSIÓN

Luis-Millán González, Tomás Saorín, Antonia Ferrer-Sapena, Rafael Aleixandre-Benavent y Fernanda Peset



Luis-Millán González, doctor por la *Universitat de València (UV)*, es profesor desde 1998 en el *Departamento de Educación Física* de la *UV* donde imparte las asignaturas de metodología de la investigación y ejercicio físico para la calidad de vida. Sus intereses son: gestión de los datos de investigación, arquitectura de la información, estadística, software documental y ejercicio físico en la patología músculo-esquelética. Ha publicado en las principales revistas de medicina del deporte indexadas en los *Journal Citation Reports*.

<http://orcid.org/0000-0002-6478-4014>

*Universitat de València, Facultat de Ciències de l'Activitat Física i l'Esport (FCAFE)
Gascó Oliag, 3, 46010 Valencia, España
luis.m.gonzalez@uv.es*



Tomás Saorín es profesor asociado en la *Facultad de Comunicación y Documentación* de la *Universidad de Murcia* y documentalista de la *Comunidad Autónoma de Murcia* donde ha participado en la puesta en marcha de proyectos de gestión de contenidos e información institucional en las áreas de servicios sociales, trabajo, empleo y bibliotecas. Participa en el capítulo español de *Wikimedia* para el conocimiento libre y en acciones de divulgación del movimiento *GLAM-Wiki* para bibliotecas, archivos y museos. Ha investigado sobre estrategia digital y de edición electrónica, así como sistemas de gestión de contenidos y proyectos colaborativos.

<http://orcid.org/0000-0001-9448-0866>

*Servicio Regional de Empleo y Formación, Unidad de Documentación
Av. Infante Juan Manuel, 14, 30011 Murcia, España
tsa@um.es*



Antonia Ferrer-Sapena es licenciada en geografía e historia contemporánea por la *Universidad de Valencia*, y doctora en técnicas y métodos de información y documentación por la *Universidad Politécnica de Valencia (UPV)*. Es coordinadora de investigación en *Florida Centre de Formació* y miembro de su *Comité de Innovación*. Entre los proyectos que dirige destaca el portal de economía social *ESocial.com*. Es profesora de la *UPV*, y miembro del foro *Innovem Junts*, del *Instituto de la Pequeña y Mediana Industria (Impiva)* de la *Generalitat Valenciana*.

<http://www.esocial.com>

<http://orcid.org/0000-0001-6432-917X>

*Universidad Politécnica de Valencia
Camino de Vera, s/n, 46022 Valencia, España
arferra@upv.es*



Rafael Aleixandre-Benavent es científico titular del *Consejo Superior de Investigaciones Científicas (CSIC)* y catedrático acreditado de biblioeconomía y documentación. Es doctor en medicina, especialista en documentación médica por la *Universitat de València (UV)* y codirector de la *Unidad de Información e Investigación Social y Sanitaria (Uisys)* de la *UV*. Sus principales líneas de trabajo son la evaluación de la investigación y de las publicaciones científicas y los estudios sobre el acceso abierto a los datos científicos

<http://orcid.org/0000-0002-6678-8844>

*Instituto de Historia de la Medicina y de la Ciencia López Piñero
Plaza Cisneros, 4, 46003 Valencia, España
rafael.aleixandre@uv.es*



Fernanda Peset es profesora titular de la *Universidad Politécnica de Valencia, Departamento de Comunicación Audiovisual, Documentación e Historia del Arte* y coordinadora del programa de doctorado. Es doctor por la *Universidad de Murcia* en 2002. Su docencia y publicaciones se orientan a comunicación científica, acceso abierto, implantación del protocolo *OAI-PMH*, normalización de la información, descripción de documentos, sistemas de documentación de museos, y datos abiertos y enlaces. Participa en proyectos como *IraLIS, E-LIS, Grupo CIEPI, ODISEA* y dirige el proyecto *HD Datasea*.

<http://orcid.org/0000-0003-3706-6332>

*Universidad Politécnica de Valencia
Camino de Vera, s/n, 46022 Valencia, España
mpeset@upv.es*

Artículo recibido el 04-02-2013
Aceptación definitiva: 19-06-2013

El profesional de la información, 2013, septiembre-octubre, v. 22, n. 5. ISSN: 1386-6710 415

CAPÍTULO 4. EL ACCESO ABIERTO A LOS DATOS EN LAS REVISTAS CIENTÍFICAS

Como se ha indicado antes, una posibilidad para el depósito de los datos es almacenamiento en las plataformas de las ediciones electrónicas de las revistas, habitualmente bajo la categoría de “material suplementario”.

Un aspecto interesante que está actualmente en estudio se refiere a la política de datos abiertos de las revistas, es decir, si permiten que los autores depositen en abierto los datos brutos de investigación que han servido de soporte a los artículos publicados en la revista.

Un trabajo previo de Alsheikh (2011) que analizaba la disponibilidad pública de los datos brutos de investigación en revistas de alto impacto llegaba a tres conclusiones importantes. La primera, la heterogeneidad en las instrucciones a los investigadores que publican en esas revistas, ya que algunas revistas requieren la

disponibilidad pública de los datos como condición para la publicación del artículo, mientras que otras fomentan el intercambio de datos pero no aportan ninguna instrucción para llevarlo a cabo y algunas revistas no mencionan ninguna práctica en absoluto. La segunda, que casi un tercio de los 500 artículos examinados no estaban sujetos a ninguna política de disponibilidad de datos porque se publicaron en revistas carentes de políticas específicas. La tercera, que aún cuando la investigación se publicara en revistas con instrucciones específicas sobre disponibilidad de los datos, más de la mitad de los artículos no depositaron los datos.

Algunos estudios llevados a cabo en nuestro grupo han llegado a conclusiones similares. Así, en un análisis sobre la política de datos abiertos de las revistas pediátricas incluidas en la *Web of Science*, el 82% de las revistas aceptaban material suplementario, pero solamente el 30% permitía la reutilización de los datos y el 23% permitía su depósito en una website, ya fuera institucional o del autor (Peset et al., en prensa). Ninguna revista

pediátrica española hacía mención a políticas de datos abiertos. En otro estudio que analizaba 13 revistas pediátricas españolas (Aleixandre et al., 2013a), cinco ofrecían indicaciones sobre la posibilidad de reutilizar los datos y de depositarlos en repositorios o páginas web personales o institucionales. Únicamente una revista, editada por la Asociación Española de Psiquiatría del Niño y el Adolescente, promociona explícitamente la apertura, aunque sin mencionar específicamente el depósito de datos, indicando que “esta revista provee acceso libre inmediato

a su contenido bajo el principio de que hacer disponible gratuitamente investigación al público apoya a un mayor intercambio de conocimiento global”. Además, menciona que “se permite y se anima a los autores a difundir sus trabajos electrónicamente (por ejemplo, en repositorios institucionales o en su propio sitio web) antes y durante el proceso de envío, ya que puede dar lugar a intercambios productivos, así como a una citación más temprana y mayor de los trabajos publicados”.

Cartas al Editor/Rev Esp Cardiol. 2013;64(12):1006-1008 1007

Compartir datos de investigación en cardiología

Sharing Research Data in Cardiology

Sra. Editora:

Desde hace más de una década, el movimiento Open Access ha permitido poner a disposición de los investigadores el texto completo de numerosas revistas, facilitando con ello el acceso universal a los avances científicos. REVISTA ESPAÑOLA DE CARDIOLOGÍA apostó desde el principio por esta modalidad de publicación, así como la edición bilingüe en español y en inglés de su versión electrónica, lo que ha mejorado su difusión y su impacto internacionales¹. Un paso más en este sentido es la apertura de los datos brutos que sirven de soporte a la investigación para compartirlos y reutilizarlos, lo que en terminología sajona se conoce como *data-sharing*. Este movimiento ya está bien implantado en algunas áreas como la genética. Así, las dos mayores entidades financiadoras del Proyecto Genoma Humano, el Wellcome Trust² y el National Institutes of Health³, han invertido en la infraestructura necesaria para compartir datos a gran escala y a largo plazo y han desarrollado normativas al respecto. En el área de las enfermedades cardiovasculares también se han hecho llamamientos en este sentido, poniendo un énfasis especial en la necesidad de compartir los datos de los ensayos clínicos controlados, una vez consideradas las necesarias medidas para proteger la intimidad de los pacientes⁴⁻⁶.

Las ventajas de compartir los datos de las investigaciones son bien conocidas: los investigadores tienen mayores oportunidades de descubrir nuevos conocimientos; aumenta el número de análisis estadísticos enmascarados; estimula los descubrimientos adicionales; evita la repetición de costosos proyectos que utilizan los mismos medicamentos o aparatos que se utilizan previamente; muchos fraudes podrían reducirse si otros investigadores pudieran comprobar los datos. En síntesis, permite que muchas investigaciones se puedan hacer con costes mínimos aprovechando los datos ya existentes, con lo que se consigue un mayor aprovechamiento de los recursos⁷.

Sin embargo, en la mayor parte de las áreas biomédicas, salvo en genética, la revolución que supone compartir datos se mueve muy lentamente y los investigadores suelen ser reacios a hacerlo, por lo que los datos de investigación están infratutilizados⁸. Muchos investigadores desconfían de liberar los datos detallados por miedo a que análisis externos de otros puedan contradecir las interpretaciones iniciales. La publicación de estudios adicionales incorrectos crearía una espiral de confusión que desorientaría a la comunidad científica internacional y la distraería de otras labores. Otros argumentan que han dedicado mucho tiempo y esfuerzo en el diseño y la recolección de los datos y, por lo tanto, desean ser ellos quienes difundan los resultados y no otros que no han participado en este proceso. Por otra parte, piensan que en el futuro pueden obtener nuevas observaciones adicionales de sus propios datos, por lo que ofrecerlos en abierto les impediría explotarlos si investigadores oportunistas han hecho uso de ellos previamente. La solución en este caso pasaría por proteger la autoría de los investigadores originales durante un periodo de tiempo y por establecer una política de incentivos, de manera que quienes comparten sus datos deberían tener ciertas ventajas frente a quienes no⁴⁻⁶.

A pesar de que todavía hay muchos asuntos sin resolver, la comunidad científica ha empezado el proceso de desarrollo de normativas y soluciones a los problemas más comunes que van surgiendo. En España, el Grupo de Trabajo de «Depósito y Gestión de datos en Acceso Abierto» del proyecto RECOLLECTA ha elaborado un informe⁹ que refleja los aspectos relevantes que intervienen en la gestión de los datos de investigación en los repositorios científicos. Por otra parte, el Ministerio de Economía y Competitividad subvenciona el proyecto «OpenDataScience, centro de recursos para la preservación y gestión de datos abiertos de investigación», que tiene como objetivo fomentar el intercambio y la reutilización de los datos primarios de investigación de manera generalizada y normalizada¹⁰. También las fundaciones privadas están interesadas en investigar en este campo y, por ejemplo, la Fundación Mapfre subvenciona el proyecto «Estrategias para la gestión y la promoción del uso compartido de datos científicos sobre la salud», con propósitos similares al anterior pero centrados en el campo de las ciencias de la salud¹¹.

En definitiva, pensamos que es necesario concienciar a los profesionales para que liberen sus datos de investigación, pero también crear infraestructuras que faciliten su depósito y preservación y las reglas de juego que aclaren muchas de las implicaciones que una estrategia de estas características supondría. Las revistas científicas con versiones electrónicas poseen la tecnología apropiada para permitir el depósito y la consulta de estos datos, y este es el caso de REVISTA ESPAÑOLA DE CARDIOLOGÍA, que acepta el envío de material electrónico adicional en varios formatos para apoyar y mejorar la presentación de la investigación científica (aunque desde abril de 2013 ha restringido el acceso abierto de algunos de sus contenidos en la web). Por otra parte, también es necesario crear repositorios institucionales y temáticos de datos para que los investigadores puedan depositar allí los suyos con garantías de salvaguarda. Queda mucho por hacer, pero el proceso de apertura y uso compartido de los datos científicos se presenta imparable.

FINANCIACIÓN

Plan Nacional de I+D+i del Ministerio de Economía y Competitividad. Fundación MAPFRE.

Rafael Aleixandre-Benavent^{*}, Antonia Ferrer-Sapena⁸, Adolfo Alonso Arroyo⁷ y Fernanda Peset⁹

¹UISYS-Unidad de Información e Investigación Social y Sanitaria, Instituto de Historia de la Medicina y de la Ciencia López Piñero, CSIC-Universitat de València, Valencia, España

²Departamento de Comunicación Audiovisual, Documentación e Historia del Arte, Universidad Politécnica de Valencia, Valencia, España

³Departamento de Historia de la Ciencia y Documentación, Universitat de València, Valencia, España

^{*}Autor para correspondencia:
Correo electrónico: Rafael.Aleixandre@uv.es
(R. Aleixandre-Benavent).

On-line el 26 de octubre de 2013

BIBLIOGRAFÍA

1. Bosch X, Villacastán JP, Alfonso F. Edición en inglés por Internet. Un nuevo gran paso adelante de Revista Española de Cardiología. *Rev Esp Cardiol*. 2002; 55:1-3.
2. Wellcome Trust. Data sharing [citado 12 Jul 2013]. Disponible en: <http://www.wellcome.ac.uk/About-us/Policy/Open-access/Data-sharing>
3. NIH Data Sharing Policy [citado 10 Jun 2013]. Disponible en: http://grants.nih.gov/grants/policy/data_sharing/
4. Gotzsche PC. Strengthening and opening up health research by sharing our raw data. *Circ Cardiovasc Qual Outcomes*. 2012;5:236-7.
5. Krumholz HM. Open Science and data sharing in clinical research: basing informed decisions on the totality of the evidence. *Circ Cardiovasc Qual Outcomes*. 2012;5:41-2.
6. Spertus JA. The double-edged sword of open access to research data. *Circ Cardiovasc Qual Outcomes*. 2012;5:141-4.
7. Ferrer Sapena A, Peset Fernandez F, Aleixandre-Benavent R. Acceso a los datos científicos y su reutilización: open data y open government. *Prof Inf*. 2011; 20:260-8.

En el área de las drogodependencias, donde se analizaron con la misma metodología 43 revistas, el porcentaje de revistas que aceptaban material complementario fue del 47%. El 68% aceptaba el almacenamiento en repositorios temáticos o institucionales y el 65% permitían la reutilización de los datos. La variable que contemplaba la posibilidad de publicar los datos en una website era aceptado en el 41% de las revistas.

En estos trabajos también se subraya la necesidad de realizar campañas de sensibilización mediante la edición de guías de buenas prácticas destinadas a convencer a autores y editores de la conveniencia de liberar y compartir los datos brutos de investigación para un aprovechamiento más óptimo. También se pone de relieve la

necesidad de poner al servicio de los investigadores el asesoramiento técnico y las infraestructuras necesarias para que puedan llevar a cabo estas prácticas, así como la elaboración de una normativa que regule todo este proceso para evitar el uso indebido o fraudulento.

En algunos trabajos publicados en revistas también se ponen de manifiesto las reticencias y temores de los investigadores a compartir sus datos, por lo que los datos de investigación están infratutilizados (Pisani y AbouZahr, 2010). Algunos manifiestan su recelo a que los análisis externos de otros puedan contradecir las interpretaciones iniciales; otros argumentan que han dedicado mucho tiempo y esfuerzo en el diseño y recolección de los datos y, por lo tanto, desean ser ellos quienes difundan

los resultados y no otros que no han participado en este proceso; también indican que es posible que en el futuro pueden obtener nuevas observaciones adicionales de

sus propios datos, por lo que ofrecerlos en abierto les impediría explotarlos si investigadores oportunistas han hecho uso de ellos previamente (Aleixandre et al, 2013b).



www.elsevier.es/anpediatr

ORIGINAL BREVE

Disponibilidad en abierto de los artículos y de los datos brutos de investigación en las revistas pediátricas españolas

R. Aleixandre-Benavent^{a,*}, A. Vidal-Infer^b, A. Alonso-Arroyo^b, J. González de Dios^c, A. Ferrer-Sapena^d y F. Peset^d

^a Instituto de Historia de la Medicina y de la Ciencia López Piñero (Consejo Superior de Investigaciones Científicas-Universitat de València), Unidad de Información e Investigación Social y Sanitaria-UYIS, Valencia, España
^b Departamento de Historia de la Ciencia y Documentación, Universitat de València, Valencia, España
^c Servicio de Pediatría, Hospital General Universitario de Alicante, Departamento de Pediatría, Universidad Miguel Hernández, Alicante, España
^d Departamento de Comunicación Audiovisual, Documentación e Historia del Arte, Universidad Politécnica de Valencia, Valencia, España

Recibido el 4 de noviembre de 2013; aceptado el 16 de noviembre de 2013

<p>PALABRAS CLAVE</p> <p>Publicación en abierto; Datos brutos de investigación; Revistas pediátricas españolas; Compartir; Reutilización</p>	<p>Resumen</p> <p>Introducción: La publicación en abierto de los artículos y de los datos brutos que han servido de soporte a la investigación permite su reutilización y mejoran el avance de la ciencia. El objetivo de este trabajo es identificar estas prácticas en las revistas pediátricas españolas. Método: Se han revisado las instrucciones de 13 revistas pediátricas españolas, identificando su política sobre acceso abierto y depósito. Resultados: Ocho revistas permiten el acceso abierto sin restricciones y 5 ofrecen indicaciones sobre la reutilización de los datos y el depósito en repositorios o páginas webs personales o institucionales.</p>
---	--


La revista *Science* señala que, "los conjuntos de datos deben ser depositados en una base de datos autorizada apropiada". Las revistas del grupo *Nature* establecen que "los autores deben poner a disposición de los demás y sin condiciones previas los materiales, datos y protocolos asociados. Los conjuntos de datos deben ponerse libremente a disposición de los lectores desde la fecha de publicación, y se deben proporcionar a los editores y a los revisores cuando se envía el trabajo a los efectos de evaluar el manuscrito".

British Medical Journal pide a los autores incluir una declaración sobre el intercambio de datos al final de cada artículo de investigación original, argumentando que mucha gente lo considera una obligación moral porque la mayoría

de la investigación se financia con fondos públicos e involucra al público como participantes.

Annals of Internal Medicine también tiene una declaración sobre el uso compartido de datos y debe publicarse no sólo el protocolo de estudio, sino también los datos en bruto y el código estadístico necesario para analizar los datos.

Las revistas de *Public Library of Science* (PLOS), requieren que los datos de los artículos sean puestos a disposición de los lectores y sin restricciones. También alienta a los lectores a contactar con los editores si tienen dificultades para obtener los datos o contactar con los autores, y en casos extremos, retirar la publicación.



"An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims.

Therefore, a condition of publication in a Nature journal is that authors are required to make materials, data and associated protocols promptly available to readers..."

▶ **Nature Policies: "Availability of data and materials"** ◀

Estas consideraciones nos llevan a la conclusión de que es necesario, por una parte, convencer a los investigadores de que las ventajas y los beneficios de compartir los datos son mayores que los perjuicios e inconvenientes, como se ha visto en el capítulo 1. Por otra, es necesario tener en cuenta diversos aspectos éticos y legales, lo que nos lleva a los dos capítulos que siguen.

CAPÍTULO 5. LA NECESIDAD DE PROTEGER A LAS PARTES IMPLICADAS

Como se ha indicado antes, algunos obstáculos al desarrollo e implantación del movimiento de acceso libre a los datos brutos de investigación proceden de los propios investigadores, es decir, de sus potenciales beneficiarios.

Suber (2012), director del Harvard Open Access Project (HOAP), ha descrito las siguientes barreras existentes para la reutilización de los datos: financieras (principalmente suscripciones a las revistas, pero también el pago de honorarios por la reutilización); legales (fundamentalmente, derechos de autor y contratos editoriales); y técnicas (logins, visibilidad limitada, duraciones limitadas, etc.).

Una solución ante la necesidad de proteger la autoría pasaría por embargar los datos originales durante un periodo de tiempo y por establecer una política de incentivos, de manera que aquellos que comparten sus datos deberían tener ciertas ventajas frente a quienes no lo hacen (Gotzsche, 2012; Krumholz, 2012; Spertus, 2012).

Otro de los motivos por los que los investigadores se muestran reticentes a compartir la información que han

recolectado está motivado por el sistema de evaluación de su actividad científica, que tiene como criterio principal de éxito académico la publicación de trabajos de investigación en revistas biomédicas (Field, 2009), la tan conocida política de “publish [documents] or perish”. Si esta política fuera reemplazada por “publish [data] or perish”, el panorama podría cambiar rápidamente, como lo hizo hace ya algunos años en el área genómica (Pisani, 2010).

En ciencias de la salud existen importantes razones prácticas y éticas que deben reconocerse para proteger a las partes implicadas en las investigaciones y en la publicación de los resultados: pacientes (o participantes en los estudios) e investigadores interesados en los resultados obtenidos.

Otro de los obstáculos irresueltos que frena el establecimiento de una cultura de intercambio de datos es la falta de reconocimiento y de incentivos. Para vencer este obstáculo, los comités de evaluación deberían premiar a los investigadores que incluyen en sus currículos la lista de conjuntos de datos compartidos en las solicitudes de subvenciones. Para premiar el intercambio de datos, sería interesante desarrollar un *Data Sharing Citation Index*, es decir, un sistema métrico para el seguimiento de la reutilización y citación, tal y como sugiere el Cancer Biomedical Informatics Grid (caBIG) Data Sharing (Heather y cols., 2008; Altman M and King G, 2007). Esta iniciativa contribuiría a fomentar el cambio de mentalidad desde el “publica artículos o perece” al “publica también los datos o perece”

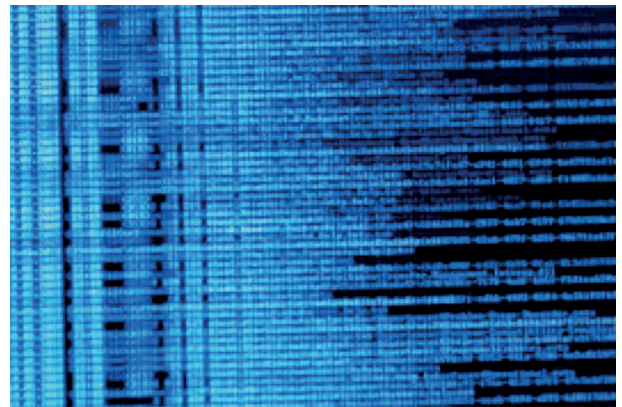


Photo by kjetikar

CAPÍTULO 6. ASPECTOS ÉTICOS Y LEGALES

Una combinación de consentimiento informado para compartir datos, junto con el anonimato y la regulación del acceso a los datos aumentará el potencial para que cada vez haya más investigadores implicados en hacer que sus datos de investigación estén fácil y ampliamente disponibles.

Cuando la investigación consiste en obtener datos de las personas, se espera que los investigadores mantengan los principios éticos y morales que rigen la realización de experimentos, tanto durante la investigación como durante el intercambio de datos.

Los datos de investigación, aunque sean personales, confidenciales y sensibles, pueden compartirse ética y legalmente si los investigadores prestan atención desde el principio de la investigación a tres aspectos importantes:

- a) Obtener el consentimiento informado para el intercambio de datos;
- b) Cuando sea necesario, proteger la identidad de las personas mediante el anonimato;
- c) Considerar establecer un control de acceso a los datos.

6.1. Consentimiento informado

La protección a los participantes debe incluir el consentimiento informado, la revisión institucional por un comité, la protección técnica y estadística de las bases de datos y los requerimientos propios a los investigadores para que respeten la intimidad de los participantes cuyos datos se incluyen (Guttmacher, 2009). En España, la Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica, establece, en sus principios básicos, que

La dignidad de la persona humana, el respeto a la autonomía de su voluntad y a su intimidad orientarán toda la actividad encaminada a obtener, utilizar, archivar, custodiar y transmitir la información y la documentación clínica, y que toda actuación en el ámbito de la sanidad requiere, con carácter general, el previo consentimiento de los pacientes o usuarios. El consentimiento, que debe obtenerse después de que el paciente reciba una información adecuada, se hará por escrito en los supuestos previstos en la Ley.

La genómica y las ciencias sociales, que tienen un historial de intercambio de datos mayor que la mayoría de las ciencias biomédicas, han desarrollado técnicas para hacer frente a dos de los principales obstáculos para el intercambio de datos públicos de investigación en salud: la confidencialidad y el consentimiento. Los procedimientos generales de consentimiento se están convirtiendo en una norma, debido en parte al desarrollo de bancos de tejidos de investigación (biobancos). El anonimato elimina algunos de los obstáculos asociados con el consentimiento, y las técnicas para la protección de la identidad están mejorando constantemente (Pisani, 2010).

El papel de los revisores en este control es crucial, pues deben asegurar que se han cumplido las conductas éticas y evitar así conflictos de interés.

Siempre que sea posible, el consentimiento debería tener en cuenta los usos futuros de datos, tales como la conservación y uso a largo plazo de los datos de investigación.

Los investigadores deben:

- a) Informar a los participantes cómo se van a almacenar los datos de investigación, cómo se van a conservar y cómo se van a utilizar a largo plazo.
- b) Informar a los participantes cómo se va a mantener la confidencialidad y el anonimato.
- c) Obtener el consentimiento informado para el intercambio de datos.

¿Cómo proceder a la anonimización

Antes de que se publiquen los datos de investigación sobre personas y de que sean compartidos con otros investigadores externos, puede ser necesario proceder a su anonimato para que no se pueden identificar.

- a) Eliminando identificadores “directos” como nombres o direcciones.
- b) Reduciendo la precisión de la información como, por ejemplo, reemplazando la fecha de nacimiento por grupos de edad.
- c) Utilizando seudónimos.
- d) Restringiendo los rangos superiores o inferiores de las variables y ocultando los valores extremos.

**Consentimiento informado para la donación voluntaria
de muestras biológicas anonimizadas para investigación
al biobanco del...**
(2 de 5)

EJEMPLAR PARA EL BIOBANCO

Declaración de consentimiento

D./Dª de años de edad, con domicilio en

DNI y nº de SIP

D./Dª de años de edad, con domicilio en

DNI en calidad de representante (en caso de minoría legal o incapacidad) del paciente con DNI y nº de SIP

DECLARO

He leído la hoja de información que se me ha entregado.

He sido informado por el profesional de salud abajo mencionado sobre la donación de muestras a un biobanco.

He comprendido las explicaciones que se me han facilitado en un lenguaje claro y sencillo.

He podido realizar observaciones y me han sido aclaradas todas las dudas que he planteado.

He comprendido que la donación de muestras a un biobanco es voluntaria y puedo revocar mi consentimiento en cualquier momento, sin tener que dar explicaciones y sin que esto repercuta en mis cuidados médicos.

Que libre y voluntariamente acepto la realización de la donación voluntaria de:

Una/s muestra/s de

Que puedo incluir restricciones sobre el uso de las mismas:

CONSIENTO

Que el Hospital u otros centros de investigación, públicos o privados, utilicen mis datos y las muestras donadas en las condiciones establecidas en la hoja de información.

Que el ente coordinador de la Red Valenciana de Biobancos pueda acceder a mis datos, en la medida en que sea necesario y manteniendo siempre su confidencialidad.

A completar por el donante:

Fdo.: D./Dª

DNI:



RED VALENCIANA

Documento de consentimiento informado sobre la donación voluntaria de muestras biológicas anonimizadas para investigación. El donante, tras declarar que ha leído la información, que ha sido informado por el profesional de salud, que ha comprendido las explicaciones que se le han facilitado en un lenguaje claro y sencillo y que le han aclarado todas las dudas que ha planteado, consiente “libre y voluntariamente” que “el Hospital u otros centros de investigación, públicos o privados, utilicen mis datos y las muestras donadas en las condiciones establecidas en la hoja de información”. También consiente que “el ente coordinador de la Red Valenciana de Biobancos pueda acceder a mis datos, en la medida en que sea necesario y manteniendo siempre su confidencialidad”.

6.2. Copyright

El copyright es un derecho de propiedad intelectual asignado automáticamente al creador, que impide la copia y la publicación no autorizada de un trabajo original. Se aplica a los datos de investigación y juega un papel muy importante en la creación, uso compartido y reutilización de los datos.

Los usuarios secundarios de los datos deben obtener la autorización del titular de los derechos antes de que estos se pueden reproducir. Los datos se pueden utilizar para su uso en la enseñanza o para propósitos de investigación sin infringir los derechos de autor, bajo el concepto de trato justo, siempre que sea reconocido el propietario de los datos.

Las licencias que se utilizan están basadas en Creative Commons, y en las que Open Knowledge Foundation recomienda para las bases de datos. Dado que los datos se distribuyen en forma de conjuntos, Datasets o bases de datos, las tres licencias ScienceCommons están referidas a esas bases de datos. Lo que declaran las licencias es qué se permite hacer con esos conjuntos de datos. Es decir, regulan la publicación y reutilización de los resultados y datos de la investigación científica, referidos a bases de datos:

- ODC Attribution License (ODC-BY), obliga a atribuir/reconocer el propietario original.
- PDDL, Public Domain Dedication and License, similar a la CC más permisiva (CC0) sin atribución ni limitación geográfica de uso.
- Open Database License, ODbL, es más restrictiva, pero permite compartir –copiar, distribuir y usar la base de datos-, producir trabajos sobre ella y modificar, transformar y construir otros productos sobre la base de datos. Con esta última, se mantiene la mención del origen (Attribution) y obliga a ceder el nuevo producto con esta misma licencia (Share-Alike).

Panton recomiendan el uso de la licencia *Public Domain Dedication and License*-PDDL, equivalente a la Creative-Commons Cero, para los datos científicos derivados de una subvención pública.

CAPÍTULO 7. ENCUESTA SOBRE LA GESTIÓN DE LOS DATOS DE INVESTIGACIÓN DE LOS INVESTIGADORES ESPAÑOLES EN CIENCIAS DE LA SALUD

En un estudio llevado a cabo en investigadores que habían publicado artículos en la revista *Annals of Internal Medicine* en 2008 y se les preguntó bajo qué condiciones pondrían sus protocolos, datos estadísticos y datos a disposición de otros, solamente el 13% manifestó que sus protocolos estaban disponibles sin ninguna condición, el 3% que estaban disponibles los datos estadísticos y el

4% los datos. Además, La mayoría de los autores que indicaron que dichos materiales estaban disponibles, requerían que los interesados contactaran con ellos primero, y muchos establecieron condiciones específicas para proceder al intercambio de estos materiales (Lane, 2009). En definitiva, los investigadores estaban más dispuestos a compartir plenamente sus protocolos y menos dispuestos a compartir datos. Otros estudios han mostrado que la falta de financiación y la falta de tiempo para destinarlo a estas tareas son los principales obstáculos para poner los datos a disposición de otros (Tenopir, 2011). Por otra parte, se pone de manifiesto que para proceder a la apertura de datos y para compartirlos es necesario establecer tempranamente un plan de gestión de los datos en el que se describa todo el proceso.

Para conocer las prácticas de los investigadores que trabajan en centros de investigación españoles en biomedicina se diseñó un cuestionario compuesto por 40 preguntas. El cuestionario se basó en otros ya implementados en otras instituciones, convenientemente adaptado a las particularidades del caso español. Los cuestionarios se han implementado en el Hospital Severo Ochoa (Leganés, Madrid), la Facultad de Medicina de Valencia, el Hospital Clínico de Valencia y el Hospital General de Castellón, entre otros.

Las preguntas del cuestionario se dividen en tres bloques: A: Datos personales (edad, sexo, nombre del centro de investigación, área, puesto, años de investigación); B: Creación y reutilización de los datos; C: Preservación de los datos.

En el apartado A de datos personales se investiga, sobre todo, las variables edad, sexo, nombre del centro de investigación, área, puesto, años de investigación. El apartado B sobre “Creación y reutilización de los datos” investiga la existencia de políticas o planes de gestión de los datos en la institución, los motivos por los que se desarrolló esta política o por los que no existe, quién es la persona en el grupo que tiene la responsabilidad de gestionar los datos, dónde se almacenan, de qué tipo de datos se hacen copias de seguridad. También se pregunta a los investigadores si les gustaría utilizar los datos de investigación no publicados de otros investigadores con los que no colabora y qué medios utiliza para localizarlos; qué datos externos suelen utilizar; si alguna vez le han solicitado los datos otros investigadores o si le han solicitado que estén en abierto; cuáles son los problemas o prejuicios que tiene para compartir. El apartado C está destinado a indagar sobre las prácticas de preservación de los datos una vez terminado el proyecto. Se pregunta sobre cuáles son las amenazas que consideran que pueden derivarse de compartir los datos, cómo los almacenan para preservar su acceso y utilización futura, dónde es más adecuado que se preserven, la existencia de planes en la institución para la creación de un repositorio de datos para su almacenamiento y preservación, la existencia en su disciplina de guías o

recomendaciones para preservar datos, la importancia que tiene la preservación y, por último, si la institución presta alguna ayuda para llevar a cabo la gestión de los datos en todo su ciclo.

Gestión de datos de investigación*

Este cuestionario se inscribe en un proyecto de investigación I+D+i cuyo objetivo es fomentar el intercambio de los datos científicos en formato digital.

Entendemos por "datos de investigación todo aquel material que ha sido registrado durante la investigación, reconocido por la comunidad científica y que sirve para certificar los resultados de la investigación que se realiza. Además de estas características también se especifica que debe provenir de una fuente única y deben ser difíciles o imposibles de obtener de nuevo" (Torres-Salinas; Robinson-García; Cabezas-Clavijo, 2012). Ejemplos: vídeos, gráficos, tablas, rawdata, datos numéricos...

Por reutilización de datos entendemos la posibilidad de crear nuevo conocimiento a partir de datos existentes, combinándolos de nuevas maneras, no imaginadas para su forma original. En un sentido estricto son más reutilizables los datos en bruto que un vídeo, por ejemplo.

Solicitamos su colaboración rellenando la encuesta que figura a continuación, que consiste en 40 preguntas divididas en 3 apartados y debería durar entre 10 y 15 minutos.

Siendo conscientes de la multitud de encuestas que llegan por correo, agradecemos que dedique unos minutos a la nuestra

* Los resultados de esta encuesta serán compartidos con proyectos similares

*Obligatorio



A. Datos personales

1. Edad *

2. Sexo *

- Hombre
 Mujer

3. Escoja de la lista la Universidad o Centro de Investigación donde haya realizado la mayor parte de su trabajo científico en los últimos 5 años *

- Universitat Politècnica de València
 Universitat de València

CAPÍTULO 8. CONCLUSIONES

- Es necesario adquirir una conciencia colectiva de la importancia de compartir datos como práctica que mejora la transmisión del conocimiento y, en definitiva, la eficiencia de las inversiones en investigación.
- En algunas ciencias biomédicas, como la genética, y en algunas ramas de las ciencias sociales, existe una mayor concienciación sobre la necesidad de compartir datos.
- La comunidad de investigadores debería contemplar el data sharing como una oportunidad, más que como una imposición o una amenaza.
- Es necesario reemplazar la política del “publish [documents] or perish”, por la de “publish [data] or perish”.
- Es necesario un cambio en la cultura de la ciencia que comprometa no sólo a los investigadores, sino que también requiere compromisos importantes de las agencias e instituciones que financian la investigación. Se podría ampliar la actual ley de la ciencia de manera que no solamente promueva el acceso abierto a las investigaciones financiadas con fondos públicos, sino también el depósito de los datos originarios en repositorios especiales que permitan que estos sean compartidos y, potencialmente, reutilizables.
- Las convocatorias públicas de proyectos de investigación deberían incluir no solo la obligatoriedad de compartir los datos derivados, sino los mecanismos para llevarlo a la práctica, incluyendo las subvenciones o infraestructura necesarias para ello.
- Es necesario realizar campañas de sensibilización dirigidas a los profesionales para que liberen sus datos de investigación, pero también crear infraestructuras que faciliten su depósito y preservación y las reglas de juego que aclaren muchas de las implicaciones que una estrategia de estas características supondrían. Deben ser los organismos oficiales los impulsores de este movimiento.
- La mayor parte de los investigadores españoles de las ciencias de la salud carecen de políticas o planes de gestión de los datos. Los principales motivos por los que no existen estas políticas son la ausencia de una obligación institucional, la falta de tiempo, el esfuerzo que se requiere para llevarla a cabo y la falta de apoyo y de orientación. Desconocen la existencia de repositorios, instituciones y otros sistemas de gestión y preservación de los datos y suelen utilizar por lo general ordenadores locales o memorias digitales portátiles. Sin embargo, a una amplia mayoría les gustaría utilizar los datos de investigación recogidos por otros investigadores. La principal amenaza a sus datos es la restricción de acceso y uso relacionada con la propiedad intelectual.

Aspiraciones a largo plazo

- Extender al máximo la disponibilidad de los datos abiertos a todos los investigadores de una manera clara y transparente.
- Conseguir los recursos técnicos y humanos y la sinfraestructuras necesarias para dar soporte a la gestión de los datos y a su sostenibilidad.
- Contribuir a la internacionalización del uso compartido de datos y apoyar a las comunidades menos desarrolladas para que puedan hacerlo.
- Contribuir a inculcar una doctrina de valor añadido al conocimiento existente que suponen los datos compartidos, especialmente en la mejora de la salud de la población.

CAPÍTULO 9. GLOSARIO

Compartir datos. Es la práctica utilizada de poner a disposición de otros investigadores los datos utilizados para cualquier investigación.

Consentimiento informado. Es la conformidad libre, voluntaria y consciente de una persona, manifestada en el pleno uso de sus facultades después de recibir la información adecuada, para participar en una investigación, después de haber comprendido la información que se le ha dado acerca de los objetivos del estudio, los beneficios, las molestias, los posibles riesgos y las alternativas, sus derechos y responsabilidades.

CNI. Coalition for Networked Information, Reino Unido. Organización dedicada a apoyar la tecnología de la información digital para el avance de la comunicación académica y el enriquecimiento de la productividad intelectual (<http://www.cni.org/>).

DataCite. Consorcio internacional que tiene como objetivo mejorar la citación de los datos con el fin de establecer un acceso más fácil a los datos de investigación científica a través de Internet, aumentar la aceptación de los datos de la investigación como legítimas contribuciones citables para el registro científico, y apoyar el archivo de datos que permita verificar los resultados y reutilizados en futuros estudios (<http://www.datacite.org/>).

Dataverse. Es un repositorio de datos de investigación. El proyecto Dataverse se encuentra alojado en la Universidad de Harvard. Disponen de un software de código abierto utilizado por la red Dataverse que permite compartir datos de investigación. La red presta ayuda en la resolución de problemas con el intercambio de datos.

Datos brutos. Son llamados también datos sueltos o desorganizados. Se trata de los datos recogidos y que no han sido organizados numéricamente.

Datos confidenciales. Datos privados o aquéllos que se acordó que fueran confidenciales o secretos, es decir,

que estuvieran fuera del dominio público, como los que tratan de sobre negocios, salud y opinión política, entre otros.

Datos personales. Datos que se refieren a personas que pueden ser identificadas a partir de ellos.

Datos personales sensibles. Son datos sobre la raza de una persona, etnia origen, opiniones políticas, creencias religiosas o similares, el comercio, la afiliación sindical, la salud física o mental, la vida sexual, la comisión presunta de delitos o las sentencia de cualquier tribunal respecto de éstos.

Digital CSIC. Repositorio de publicaciones y de datos del Consejo Superior de Investigaciones Científicas. <http://digital.csic.es/politicas/politicaDatos.jsp>

DOI. Digital Object Identifier. Identificador único y permanente para las publicaciones electrónicas.

Drive. Es un servicio de Google para alojar y compartir archivos. Actualmente este servicio ofrece 15 gigabytes de espacio gratuito por usuario.

Dropbox. Es un servicio para alojar y compartir archivos en la nube de la empresa Dropbox.

Dryad. Es un repositorio de datos de investigación de publicaciones científicas. Este proyecto se creó a partir de la iniciativa de un grupo de editores de revistas y sociedades científicas de biología evolutiva y ecología.

Figshare. Es un repositorio donde se pueden alojar los datos de investigación y ponerlos a disposición de la comunidad científica de manera citable y reconocible. Permite la carga de cualquier tipo de archivo y hacerlo visualizable a través del navegador. Permite la difusión de: cifras, datos, medios de comunicación, documentos, carteles, presentaciones y catálogos de archivos.

Horizon 2020. Horizonte 2020 es el programa que financia proyectos de investigación e innovación de diversas áreas temáticas en el contexto europeo para el periodo 2014-2020 destinado a investigadores, empresas, centros tecnológicos y entidades públicas.

Interoperabilidad. El Instituto de Ingenieros Eléctricos y Electrónicos (IEEE) define interoperabilidad como la habilidad de dos o más sistemas o componentes para intercambiar información y utilizar la información intercambiada.

JISC. *Joint Information Systems Committee.* Es el Comité de Sistemas de Información de Reino Unido, es un organismo público que tiene como labor el apoyo de

la explotación eficaz de los sistemas de información en la educación superior y en la investigación, proporcionándoles además la infraestructura de alta calidad necesaria.

NIH. *National Institutes of Health.* Estas siglas reúnen al grupo de instituciones del gobierno americano dedicadas a la investigación médica.

ODISEA. International Registry on Research Data. Herramienta web desarrollada en la Universidad Politécnica de Valencia que registra bancos, repositorios y bibliotecas de datos destinadas a la preservación y reuso de los datos de investigación (beneficios, ciclo de vida, proyectos, normativas, etc.). Ofrece resultados sobre número y tipología temática de este tipo de depósitos a escala mundial (<http://odisea.ciepi.org>).

Openarchives.org community. Reune a la comunidad de usuarios que comparten información a través del protocolo OAI (Open Archive Initiative)

OpenDOAR. Es un directorio de repositorios de acceso abierto académicos (www.opendoar.org).

Orcid. Orcid (Open Researcher & Contributor ID) es un proyecto sin ánimo de lucro que ofrece un sistema para la identificación inequívoca de los autores de publicaciones científicas, y a la vez un espacio para registrar sus datos y trabajos, y si lo desean, compartirlos. (https://biblioteca.ulpgc.es/noticia_orcid).

ReadMe file. Archivo "Léame". Archivo que acompaña normalmente a los programas, en el cual hay información de relevancia sobre éste en cuanto a su uso, características y requisitos de hardware. Generalmente se distribuye en formato de texto plano, aunque también puede estar en otros formatos.

ROAR. <http://roar.eprints.org> Es un registro de repositorios. Tiene como objetivo la promoción del acceso abierto y lo hace a través del suministro de información acerca del crecimiento y estado de los repositorios de todo el mundo.

Wellcome Trust. Es una fundación benéfica que tiene como objetivo la mejora de la salud humana y animal, por lo que apoya la investigación en estos campos.

Zenodo. Es un repositorio europeo de datos de investigación creado por OpenAire y el CERN con el apoyo de la Comisión Europea. El repositorio recoge datos multidisciplinares de investigación de cualquier comunidad científica, investigador o institución de investigación.



CAPÍTULO 10. RECURSOS EN INTERNET

DataCite. <http://www.datacite.org/>

Dataverse. <http://thedata.org>

Digital CSIC. <http://digital.csic.es/politicas/politicaDatos.jsp>

Drive. <http://drive.google.com>

Dropbox. <https://www.dropbox.com>

Dryad. <http://datadryad.org>

Figshare. <http://figshare.com>

Harvard Open Access Project (HOAP). http://cyber.law.harvard.edu/hoap/Main_Page

Horizon 2020: <http://ec.europa.eu/programmes/horizon2020/>

JISC. <http://www.jisc.ac.uk/>

ODISEA. <http://odisea.ciepi.org>

Open access (Digital Agenda site): <http://ec.europa.eu/digital-agenda/en/open-access-scientific-knowledge-0>

Open access (Science in Society site): http://ec.europa.eu/research/science-society/open_access

OpenAIRE: <http://www.openaire.eu/>

OpenDOAR. www.opendoar.org

Participant Portal: <http://ec.europa.eu/research/participants/portal/>

ROAR. <http://roar.eprints.org>

UK Data Archive (UKDA). <http://www.data-archive.ac.uk>

Zenodo. <http://zenodo.org>

CAPÍTULO 11. BIBLIOGRAFÍA

Aleixandre-Benavent R, Ferrer-Sapena A, Alonso Arroyo A, Peset F. Sharing research data in Cardiology. *Rev Esp Cardiol*. 2013; 66:1007-8.

Aleixandre-Benavent R, Alonso-Arroyo A, Vidal-Infer A, Catalá-López F. Fomento del acceso público y el uso compartido de los datos brutos procedentes de la investigación científica. *Med Clí (Barc)*. 2015; 144(6):283-4.

Aleixandre-Benavent R, Vidal-Infer A, Alonso-Arroyo A, Valderrama-Zurián JC, Bueno-Cañigral F, Ferrer-Sapena A. Public availability of published research data in substance abuse journals. *Int J Drug Policy*. 2014; 25: 1143-6.

Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JPA (2011) Public Availability of Published Research Data in High-Impact Journals. *PLoS ONE* 6(9): e24357.

Altman M, King G (2007) A proposed standard for the scholarly citation of quantitative data. *D-Lib Magazine* 13: 3/4. Available: <http://gking.harvard.edu/files/cite.pdf>.

Biotechnology and Biological Sciences Research Council (2010) BBSRC data sharing policy. Acceso: www.bbsrc.ac.uk/organisation/policies/position/policy/datasharing-policy.aspx

Clarke MT. Open sesame? Increasing access to medical literature. *Pediatrics*. 2004;114:265-8.

Davis PM. Public accessibility of biomedical articles from PubMed Central reduces journal readership--retrospective cohort analysis. *FASEB J*. 2013 Apr 4.[Epub ahead of print] PubMed PMID: 23554455.

Dealing with data. Challenges and opportunities. Introduction. *Science*, 331 (February (6018)) (2011), pp. 692-693

Economic and Social Research Council (2010) ESRC researchdata policy. Acceso: www.esrc.ac.uk/about-esrc/information/data-policy.aspx

Evangelou E, Trikalinos TA, Ioannidis JP. Unavailability of online supplementary scientific information from articles published in major journals. *FASEB J*. 2005; 19: 1943-4.

Field D, Sansone SA, Collis A, Booth T, Dukes P, Gregurick SK, Kennedy K et al. Omics data sharing *Science* 2009. 326:234-236. doi:10.1126/science.1180598

González LM, Saorín T, Ferrer A, Aleixandre-Benavent R, Peset F. Gestión de datos de investigación: infraestructuras para su difusión. *Prof Inf*. 2013;22:414-23.

Gotzsche PC. Strengthening and opening up health research by sharing our raw data. *Circ Cardiovasc Qual Outcomes* 2012; 5:236-237.

Guttmacher AE, Nabel EG, Collins FS. Why data-sharing policies matter. *Proc Natl Acad Sci U S A*. 2009 Oct 6;106(40):16894. doi: 10.1073/pnas.0910378106. Epub 2009 Sep 18.

Heather A. Piwowar and Wendy W. Chapman. Public sharing of research datasets: a pilot study of associations. *J Informetr*. 2010 April ; 4(2): 148-156. doi:10.1016/j.joi.2009.11.010.

Kaye J. The tension between data sharing and the protection of privacy in genomics research. *Annu Rev Genomics Hum Genet*. 2012;13:415-31. doi:10.1146/annurev-genom-082410-101454. Epub 2012 Mar 9. Review. PubMed PMID:22404490.

Krumholz HM. Open Science and data sharing in clinical research: basing informed decisions on the totality of the evidence. *Circ Cardiovasc Qual Outcomes* 2012; 5:141-142.

Medical Research Council (n.d.) MRC policy on data sharing and preservation. Acceso: www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Datasharinginitiative/Policy/index.htm

Mennes M, Biswal BB, Castellanos FX, Milham MP. Making data sharing work: The FCP/INDI experience. *Neuroimage*. 2012; pii: S1053-8119(12)01067-1.

Molloy JC. The open knowledge foundation: open data means better science *PLoS Biol.*, 9 (December (12) (2011), p. e1001195

Murray-Rust, Peter. Open Data in Science. Available from Nature Precedings <<http://hdl.handle.net/10101/npre.2008.1526.1>> (2008).

Natural Environment Research Council (2011) NERC data policy. www.nerc.ac.uk/research/sites/data/policy.asp

Nature Publishing Group (2009) Nature journals' policy on availability of materials and data. Acceso: www.nature.com/authors/policies/availability.html

NIH Data Sharing Policy. [página web]. [citado 10 Junio 2013]. Disponible en: http://grants.nih.gov/grants/policy/data_sharing/.

Organisation for Economic Co-operation and Development (2007) OECD principles and guidelines for access to research data from public funding. [página web]. [citado 8 Noviembre 2013]. Disponible en: www.oecd.org/dataoecd/9/61/38500813.pdf

Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. *Bull World Health Organ*. 2010 Jun;88(6):462-6. doi: 10.2471/BLT.09.074393.

Rural Economy and Land Use Programme (n.d.). Project communication and data management plan. Acceso: www.relu.dataarchive.ac.uk/DMP2010.doc

Rural Economy and Land Use Programme Data Support Service(2011) Example data management plans. www.relu.dataarchive.ac.uk/DMPexample.asp

Spertus JA. The double-edged sword of open access to research data. *Circ Cardiovasc Qual Outcomes* 2012; 5:143-144.

Spertus JA. The double-edged sword of open access to research data. *Circ Cardiovasc Qual Outcomes* 2012; 5:143-144.

Suber P. Harvard Open Access Project (HOAP) (<http://cyber.law.harvard.edu/research/hoap>).

Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu E, Read E *et al.* Data sharing by scientists: practices and perceptions. *PLoS ONE*, 6 (6) (2011), p. e21101

Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, Manoff M, Frame M. Data sharing by scientists: practices and perceptions. *PLoS One*. 2011;6(6):e21101. doi: 10.1371/journal.pone.0021101. Epub 2011 Jun 29. PubMed PMID: 21738610.

UK Data Archive (2011) Activity-based data management costing tool for researchers. www.dataarchive.ac.uk/media/257647/ukda_jiscdmcosting.pdf

Van den Eynden, V., Bishop, L., Horton, L. and Corti, L. (2010). Data management practices in the social sciences. Acceso: www.dataarchive.ac.uk/media/203597/datamanagement_socialsciences.pdf

Walport M, Brest P. Sharing research data to improve public health. *Lancet*, 377 (February (9765)) (2011), pp. 537–539.

Wellcome Trust (2010) Wellcome Trust policy on data management and sharing. www.wellcome.ac.uk/Aboutus/Policy/Policy-and-position-statements/WTX035043.htm

Wellcome Trust. Data sharing. [página web]. [citado 12 Julio 2013]. Disponible en: <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/>.

ANEXO 1. ORGANISMOS INTERNACIONALES



National Institutes of Health

Los NIH fueron los primeros organismos gubernamentales que establecieron la obligatoriedad en compartir datos. Desde 2003 reclaman a los investigadores que solicitan financiación para proyectos de investigación que declaren cuál es su plan para poner en acceso abierto los datos que derivaran del mismo. Para dar cobertura a los datos proporcionados por los investigadores como consecuencia de estos requerimientos, los NIH han creado diversos repositorios y herramientas. Los artículos deben depositarse en PubMed Central.

http://grants.nih.gov/grants/policy/data_sharing/

Medical Research Council-MRC



El Medical Research Council (MRC) del Reino Unido, a través de su Data Support Service, ayuda a mejorar las capacidades de uso compartido de datos poblacionales y otros estudios basados en los pacientes. El servicio se centra en:

- Descubrimiento de conjuntos de datos de investigación con metadatos y documentación esencial.
- Gestión del acceso y uso de los datos para la investigación secundaria de calidad.
- Planificación y ejecución de la gestión efectiva de datos en la investigación.
- Preservación para permitir la reutilización informada de los datos conservados.

- Preservación a largo plazo de los conjuntos de datos de alto valor financiados por el MRC.
- Herramientas, normas y directrices para apoyar estas actividades.

<http://www.mrc.ac.uk/index.htm>

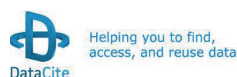
Wellcome Trust



El Wellcome Trust se compromete a garantizar que los resultados de la investigación que financia, incluyendo datos de la investigación, se gestionen y utilicen de forma que maximicen el beneficio público. Poniendo los datos de investigación a disposición de la comunidad de investigación de manera oportuna y responsable se asegura que estos datos pueden ser verificados, utilizados y aprovechados para avanzar en el conocimiento y para generar mejoras en la salud. Para maximizar el valor de los datos de investigación es necesario fomentar una cultura en la que tanto los generadores como los usuarios de datos adopten buenas prácticas de investigación y actúen con integridad y transparencia en la gestión, el uso y el intercambio de datos. Tanto los investigadores, como las instituciones de investigación, los proveedores de recursos y los editores tienen un papel importante que desempeñar para desarrollar los recursos y sistemas necesarios.

<http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm>

Data Cite



Bajo el lema *Helping you to find, access and data*, Data Cite es una organización sin ánimo de lucro cuyo objetivo es:

- Establecer un acceso más fácil a los datos de investigación en Internet
- Aumentar la aceptación de los datos de investigación como aportes legítimos y citables en el expediente académico
- Dar soporte al archivo de datos para que los resultados puedan ser verificados y re-utilizados para estudios futuros.

Proporciona apoyo a los investigadores, ayudándoles a encontrar, identificar y citar datos de investigación fiables, así como a los centros y editores de revistas para que puedan establecer links entre los artículos de investigación y los datos subyacentes.

Se basa en la idea de que se debe citar de la misma manera otras fuentes de información, tales como artículos, libros y datos, pues ello permite una fácil reutilización y verificación y que se pueda hacer un seguimiento de su impacto.

<http://datacite.org/>

OECD (Organisation for Economic Co-operation and Development)



La OECD (Organización para la Cooperación y el Desarrollo Económicos) ha sido una gran impulsora del data sharing. En 2004 (OECD, 2007; Arzberger et al., 2004) publicó OECD Principles and guidelines for access to research data from public funding (2007), donde se fijan recomendaciones y directrices para gobiernos y agencias públicas de investigación, que sirven como base de sus políticas de acceso abierto.

www.oecd.org

La Unión Europea



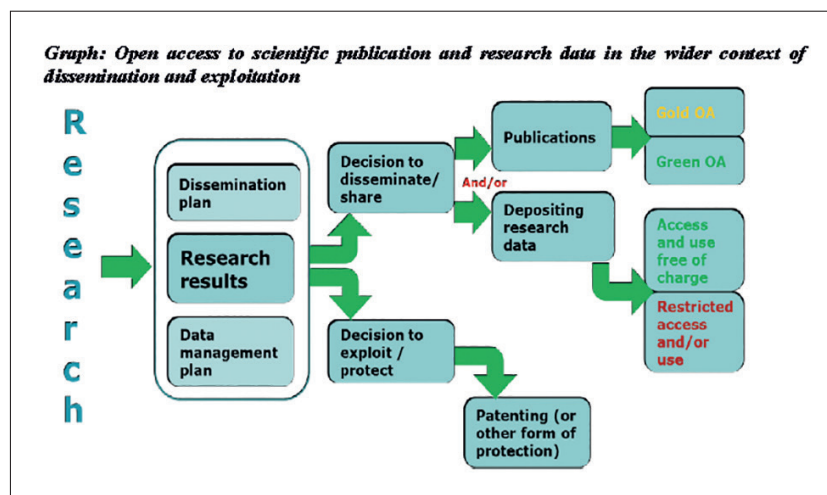
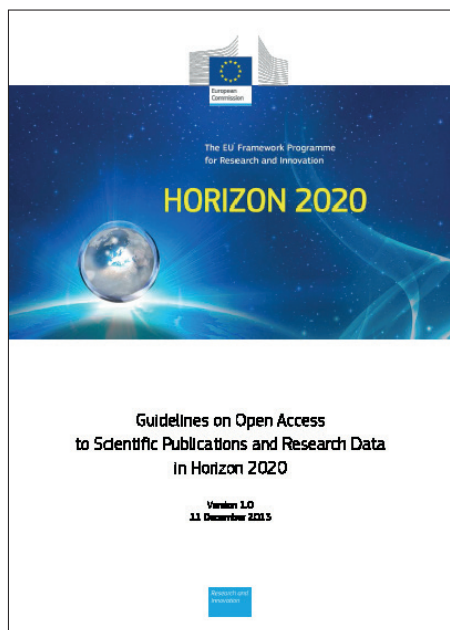
En 2007, la Unión Europea publicó *On scientific information in the digital age: access, dissemination and preservation*, donde se exponen proyectos específicos como Caspar (Cultural, artistic and scientific knowledge for preservation, access and retrieval), que estudia cómo se acceden y preservan los datos, o Seadatanet, que permitió la creación de la infraestructura para compartir datos marinos de más de 40 países.

El VI Programa marco contemplaba la cofinanciación de repositorios convencionales y de datos, dando luz verde a proyectos como ODE (*Opportunities for data exchange*), que analiza las opiniones de expertos y percepciones de los investigadores sobre data sharing; Aparsen (*Alliance for permanent access*), una red de excelencia que centraliza los esfuerzos investigadores y garantiza la preservación de los datos científicos; Parse Insight (*Permanent access to the records of science in Europe*), que establece el plan de actuación que deberá seguir la Unión Europea en esta materia (Torres-Salinas y cols., 2012).

El VII Programa Marco de la Unión Europea, Horizon 2020, establece las condiciones en las que se deberá ofrecer el acceso abierto a los resultados de los proyectos en los acuerdos de subvención de los proyectos financiados por Consejo Europeo de Investigación. Los beneficiarios de estos proyectos financiados deben:

- a) Depositar en un repositorio abierto de datos de investigación, es decir, los datos necesarios para validar las publicaciones de investigación, con sus metadatos, tan pronto como sea posible; y otros datos, con sus metadatos, dentro del plazo previsto en sus planes de gestión de datos así como el momento de liberación que indicaron.
- b) ofrecer información en el repositorio sobre las herramientas para validar la investigación e incluso, si es posible, proporcionarlas.

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf



“Open access to research data refers to the right to access and re-use digital research data under the terms and conditions set out in the Grant Agreement. Openly accessible research data can typically be accessed, mined, exploited, reproduced and disseminated free of charge for the user.”

Fuente: Horizon 2020.

ANEXO 2. RECOMENDACIONES PARA LOS CENTROS DE INVESTIGACIÓN SANITARIOS PARA FOMENTAR EL USO COMPARTIDO DE DATOS (PIOWAR, 2008)

1. Comprometerse a compartir los datos de investigación de la forma más abierta posible, una vez solventadas las restricciones de privacidad. Optimizar las tecnologías de la información y de la comunicación para hacerlo posible.
2. Reconocer el intercambio de datos en las decisiones de contratación y de promoción académica, gratificando esta práctica incrementando el factor de impacto. Usar las medidas concretas que esté disponibles para poder proceder a esta gratificación.
3. Educar a los alumnos y a los investigadores en el intercambio de datos responsable y en las prácticas de reutilización a través del trabajo en clase, tutorías y del desarrollo profesional. Promover un marco para la adopción de mecanismos adecuados para el intercambio de datos.
4. Fomentar las prácticas de intercambio de datos como parte de las políticas de publicación. Integrar estas políticas explícitamente en las instrucciones a los autores y a los revisores de las revistas científicas y congresos
5. Fomentar el intercambio de datos como parte de las políticas de financiación. Las entidades financiadoras deben establecer los requisitos oportunos para compartir los datos y evaluarán el plan de intercambio de datos como parte de su contribución científica.
6. Financiar los costes derivados del intercambio de datos, su depósito en repositorios, y la adopción de infraestructuras para el depósito y las métricas apropiadas, así como la investigación sobre las mejores prácticas posibles.
7. Publicar las experiencias en el intercambio de datos para facilitar el intercambio de buenas prácticas.

ANEXO 3. ARGUMENTOS COMÚNMENTE ACEPTADOS EN NOMBRE DE OPEN DATA (MURRAY-RUST)

- Los datos pertenecen a la raza humana”. Ejemplos típicos son los genomas, los datos sobre organismos vivos, los de las ciencias biomédicas y los datos sobre el medio ambiente.
- Si se utilizó dinero público para financiar el proyecto, éste y los datos derivados deben estar disponibles universalmente.
- Si el proyecto fue desarrollado por o en una institución gubernamental, éste y los datos derivados deben estar disponibles universalmente.
- Los datos no pueden ser protegidos legalmente por copyright.
- Los patrocinadores de la investigación no tienen pleno valor a menos que los datos resultantes sean de libre disposición.
- Las restricciones en la reutilización de datos producen su infrutilización.
- Algunos datos son necesarios para ejecutar actividades de la comunidad (datos de las instituciones públicas, datos geográficos, etc.), por lo que deben estar disponibles.
- En la investigación científica, la tasa de descubrimiento se acelera si existe un mejor acceso a los datos.

COMPARTIR DATOS ACELERA EL AVANCE DE LA CIENCIA

“Los datos abiertos son un recurso sin explotar con un enorme potencial para fomentar la construcción de sociedades más fuertes, más interconectadas que satisfagan mejor las necesidades de nuestros ciudadanos y permitir que florezca la innovación y la prosperidad.”

Líderes del G8 en la carta pública de junio de 2013

“Como dijo el estadístico profesor Deming: ‘En Dios confiamos, todos los demás traigan los datos’. Vamos a construir un hogar para esos datos –de manera que todos puedan traerlos y usarlos-, y lancémonos a un próximo gran descubrimiento.”

Kroes, Neelie (2013). *Opening up scientific data*. Speech/13/236. Event date: 18/03/2013. http://europa.eu/rapid/press-release_SPEECH-13-236_en.htm

Los datos suelen tener una vida útil más larga que el proyecto de investigación que los crea. Los investigadores pueden continuar trabajando con ellos e integrarlos y reutilizarlos en trabajos futuros.

Los datos científicos bien organizados, bien documentados, preservados y compartidos son muy valiosos para el avance de la investigación científica y permiten aumentar las oportunidades para el aprendizaje y la innovación.

6. ANEXO IV. VOCABULARIO DE MATERIAL SUPLEMENTARIO

1. DECLARACIÓN DE DEFINICIONES

Los datos de investigación constituyen un apoyo a la publicación de los artículos científicos, y se adjuntan como material suplementario, pueden ser datos observacionales, experimentales y computacionales que resultan del proceso de investigación. Las tipologías son múltiples y varían en función de la disciplina científica.

Una definición adoptada por entidades como los *National Institutes of Health* (NIH) de Estados Unidos o la Organización para la Cooperación y el Desarrollo Económicos (OCDE 2007), se recoge en Torres-Salinas (2012): que considera datos de investigación todo aquel material que ha sido registrado durante la investigación, reconocido por la comunidad científica y que sirve para certificar los resultados de la investigación que se realiza.

El Informe para la conservación y reutilización de los datos científicos en España del grupo RECOLECTA recoge la definición de la Universidad de Melbourne: “Los datos de la investigación son hechos, observaciones o experiencias en que se basa el argumento, la teoría o la

prueba. Los datos pueden ser numéricos, descriptivos o visuales. Los datos pueden ser en estado bruto o analizado, pueden ser experimentales u observacionales. Los datos incluyen: cuadernos de laboratorio, cuadernos de campo, datos de investigación primaria (incluidos los datos en papel o en soporte informático), cuestionarios, cintas de audio, vídeos, desarrollo de modelos, fotografías, películas, y las comprobaciones y las respuestas de la prueba. Las colecciones de datos para la investigación pueden incluir diapositivas; diseños y muestras. En la información sobre la procedencia de los datos también se podría incluir: el cómo, cuándo, donde se recogió y con qué (por ejemplo, instrumentos). El código de software utilizado para generar, comentar o analizar los datos también pueden ser considerados datos”

2. ACUÑACIÓN DE TÉRMINOS

A continuación se recogen las variantes terminológicas con las que se hace referencia al concepto de datos de investigación:

Additional material	Material adicional
Auxiliary materials	Materiales auxiliares
Datasets	Datasets
Electronic supplementary information	Información electrónica suplementaria
Electronic supplementary material	Material electrónico suplementario
Multimedia materials	Materiales multimedia
On line material	Material en línea
On line supporting material	Material suplementario en línea
Online Data Supplements	Datos suplementarios en línea
Supporting data	Datos de apoyo
Supporting information	Información de apoyo
Supporting material	Material de apoyo
Supplemental data	Datos suplementarios
Supplemental Digital Content	Contenido digital suplementario
supplemental information	Información suplementaria
Supplemental material	Material suplementario
Supplementary data	Datos suplementarios
Supplementary material	Material suplementario
Supplementary materials online	Materiales suplementarios en línea
Supplements	Suplementos

3. ONTOLOGÍAS

La Open Knowledge Foundation (OKFN) recoge un listado con las ontologías hasta ahora creadas en *Linked Open Vocabularies* (LOV) por áreas de conocimiento. En el área de Life Science (http://lov.okfn.org/dataset/lov/details/vocabularySpace_Life.html) las ontologías existentes son:

- The Translational Medicine Ontology (TMO)
- DICOM files (Digital Imaging and Communications in Medicine)
- BIOL - Biological Taxonomy Vocabulary
- BIOPAX - format for biological pathway data
- BIOTOP- ontology for Biology and Medicine
- BOTANY - Biological Taxonomy Vocabulary
- OBSM - Observation Method Ontology
- TAXON - TaxonMap Ontology
- TXN - TaxonConcept Ontology
- UNIPROT - Uniprot Core Ontology
- WLO - Wildlife Ontology

4. BIBLIOGRAFÍA

- Grupo de Trabajo de “Depósito y Gestión de datos en Acceso Abierto” del proyecto RECOLECTA. Informe preliminar para la conservación y reutilización de los datos científicos en España [en línea] Madrid: Fundación Española para la Ciencia y la Tecnología, FECYT (2012). Disponible en WWW.FECYT.ES
- NIH (National Institutes of Health). Final NIH Statement on sharing data. National Institutes of Health, 2003. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm
- OECD (Organisation for Economic Co-operation and Development). OECD Principles and guidelines for access to research data from public funding. Paris: OECD Publications, 2007. <http://www.oecd.org/dataoecd/9/61/38500813.pdf>
- Torres-Salinas, Daniel; Robinson-García, Nicolás; Cabezas-Clavijo, Álvaro (2012). “Compartir los datos de investigación: introducción al data sharing”. *El profesional de la información*, v. 21, n. 2, pp. 173-184. B. <http://hdl.handle.net/10760/16786>

7. ANEXO V. PUBLICACIONES CIENTÍFICAS SOBRE LOS ASPECTOS TÉCNICOS Y PROCEDIMENTALES

Las publicaciones científicas sobre los aspectos técnicos y procedimentales se han clasificado en tres grupos: Ciencias de la Información, Ciencias de la Salud y Ciencias del Deporte.

CIENCIAS DE LA INFORMACIÓN

- Peset, F y González, LM (2014). Gestión de datos científicos: RDM research data management. SEDIC, marzo 2014. http://www.sedic.es/f_cursosvirtuales-14-2tr1.asp
- González, Luis-Millán and Peset, Fernanda. *Datos de investigación: reflexiones sobre su acceso abierto.*, 2013. In International Open Access Week 2013, 21 de octubre de 2013. <http://hdl.handle.net/10760/22576>
- Peset, Fernanda; Fernández-López, Antonio-Lázaro (2013) Carencias informativas de los datos abiertos en España. *Notas ThinkEPI 2014*. <http://www.thinkepi.net/carencias-informativas-de-los-datos-abiertos-en-espana>
- Nina-Alcocer, Victor; Blasco-Gil, Yolanda; Peset, Fernanda (2013). Los inicios del datasharing: guía práctica para compartir datos de investigación. *El profesional de la información*, noviembre-diciembre, v. 22, n. 6, pp. 562-. http://eprints.rclis.org/209711/datas_haring.pdf
- González, Luis-Millán; Saorín, Tomás; Ferrer-Sapena, Antonia; Aleixandre-Benavent, Rafael; Peset, Fernanda (2013). Gestión de datos de investigación: infraestructuras para su difusión. *El profesional de la información*, vol. 22, n. 5, pp. 415-423. <http://eprints.rclis.org/20912/>
- Saorín, Tomás; Peset, Fernanda; Ferrer-Sapena, Antonia. Factores para la adopción de linked data e implantación de la web semántica en bibliotecas, archivos y museos. *Information Research* [en línea]. Marzo 2013, vol. 18, nº 1, paper 570. Disponible en: <http://eprints.rclis.org/21005/>
- Ferrer-Sapena, Antonia; Sánchez-Pérez, Enrique-A. Open data, big data: ¿hacia dónde nos dirigimos? *Anuario ThinkEPI 013*. <http://eprints.rclis.org/21006/>
- Aleixandre-Benavent, Rafael; Ferrer-Sapena, Antonia. Cómo aumentar la difusión y el impacto de los trabajos participando en la ciencia abierta. En: *X Foro Internacional sobre la evaluación de la calidad de la investigación y de la educación superior (FECIES)* 25-28 junio 2013 Granada, España. Tipo de participación: sesión invitada.
- García-García, Alicia; García-Massó, Xavi; Ferrer, Antonia; González-Moreno, Luis M.; Fernanda Peset (2013). ¿Es reutilizable el material adicional en las revistas más citadas? = Is supplementary material reusable in the most cited journals? *CRECS 3ª Conferencia sobre revistas en ciencias sociales y humanidades*, Sevilla, 9 mayo 2013. <http://eprints.rclis.org/21001/> y <http://tv.us.es/%C2%BFes-reutilizable-el-material-adicional-en-las-revistas-mas-citadas/>
- García-García, Alicia; Peset, Fernanda (2013) Tres en uno, tres informes sobre la gestión de datos de investigación. *Blok de bid*. <http://www.ub.edu/blokdebid/es/content/tres-en-uno-tres-informes-sobre-la-gestion-de-los-datos-de-investigacion>
- Ferrer-Sapena, Antonia; Villamón, M guel; González, Luis-Millán; Peset, Fernanda; Aleixandre-Benavent, Rafael; García-García, Alicia; Morales Aznar, José. Gestión de los datos de investigación como medida de calidad. En: *IX Foro Internacional sobre la evaluación de la calidad de la investigación y de la educación superior (FECIES)* 12-15 junio 2012 Santiago, España [en línea]. Granada: Asociación Española de Psicología Conductual, 2012. p. 455. ISBN 978-84-695-3701-5. Tipo de participación: comunicación. Disponible en: <http://www.ugr.es/~aepc/IXFORO/LIBRORESUMENESIXFORO.pdf>
- Ferrer-Sapena, Antonia; Peset, Fernanda La reutilización de datos culturales. *Anuario ThinkEPI 2012*. <http://eprints.rclis.org/21004/>
- García-García, Alicia; García-Massó, Xavi; Ferrer-Sapena, Antonia; González, Luis-Millán; Peset, Fernanda; Rodríguez-Gairín, Josep-Manuel; Saorín, Tomás. O iSEA: International Registry on Research Data. En: *5as Jornadas OS-Repositorios "La motricidad de los repositorios de acceso abierto"* 23 al 25 de mayo 2012 Universidad de País Vasco, Bilbao, España [en línea]. Tipo de participación: comunicación. Disponible en: <http://e-spacio.uned.es/ocs/index.php/osrepositoriosBilbao/motricidad/paper/view/15>
- García-García, Alicia; García-Massó, Xavier; Ferrer-Sapena, Antonia; González-Moreno, Luis-Millán; Peset, Fernanda; Aleixandre, Rafael. Mejores práctica en reuso de conjuntos de datos publicados online como material adicional a los artículos. En: *2a Conferencia sobre calidad de revistas de ciencias sociales y humanidades (CRECS 2012)* PI y Grupo ThinkEPI, 10 de mayo de 2012, Valencia, España [en línea]. Tipo de participación: comunicación. Disponible en: <http://eprints.rclis.org/20999/>
- García-García, Alicia; Rodríguez-Gairín, Josep-Manuel; Saorín, Tomás; González, Luis-Millán; García-Massó, Xavi; Ferrer-Sapena, Antonia; Peset, Fernanda

ODiSEA: International Registry on Research Data. *BiD: textos universitarios de biblioteconomía i documentació* [en línea]. Diciembre 2012, nº 29. Disponible en: <http://eprints.rclis.org/21000/>

- Grupo de Trabajo de “Depósito y Gestión de datos en Acceso Abierto” del proyecto RECOLECTA (2012). *Informe preliminar para la conservación y reutilización de los datos científicos en España*. Madrid: Fundación Española para la Ciencia y la Tecnología, FECYT (2012) http://www.recolecta.net/buscador/documentos/info_rme_datos_cientificos_en_esp.pdf
- Ferrer-Sapena, Antonia; Peset, Fernanda; Aleixandre-Benavent, Rafael. Acceso a los datos públicos y su reutilización: open data y open government. *El profesional de la información* [en línea]. Mayo-junio 2011, vol. 20, nº 3, pp. 260-269. Disponible en: <http://www.elprofesionaldelainformacion.com/contenidos/2011/mayo/03.html>
- Peset, Fernanda; Aleixandre Benavent, Rafael; García-García, Alicia; Villamón, Miguel; González-Moreno, Luis-Millán; Ferrer-Sapena, Antonia. Open Data en el mundo científico: proyecto OpenDataScience. En: *VIII Foro sobre la evaluación de la calidad de la investigación y de la educación superior FECIES*, 31 mayo a 3 junio 2011 Santander [en línea]. Granada: Asociación Española de Psicología Conductual, 011. pp. 481-482. ISBN 978-84-694-3488-8. Tipo de participación: comunicación. Disponible en: <http://www.ugr.es/~aepc/VIIIIFORO/Documentos/Libros/librosresumenesviiiiforo.PDF>
- Peset, Fernanda; Ferrer-Sapena, Antonia; Subirats-Coll, Imma. Open data y Linked open data: su impacto en el área de bibliotecas y documentación. *El profesional de la información* [en línea]. Marzo-abril 2011, vol. 20, nº 2, pp. 165-173. Disponible en: <http://eprints.rclis.org/21003/>

CIENCIAS DE LA SALUD

- Aleixandre-Benaven, Rafael and Ferrer-Sapena, Antonia and Alonso Arroyo, Adolfo and Peset, Fernanda Compartir datos de investigación en cardiología. *Rev Esp Cardiol*, 2013, vol. 66, n. 12, pp. 1006-008. <http://hdl.handle.net/10760/22578>
- Aleixandre-Benavent, R; Ribeiro-Schneider, D; Bueno-Cañigral, F.J; Ferrer Sapena, A.; Peset Mancebo, F; Garci Garcia, A; Valderrama Zurián, J.C. Disponibilidad de los datos públicos de investigación publicados en revistas sobre drogodependencias. *Congreso Internacional sobre Drogas. Brasilia. Brasil*, 3-5 de mayo de 013. <http://eprints.rclis.org/21009/>. Descargar pdf
- García-García, Alicia; Aleixandre-Benavent, Rafael; Ferrer-Sapena, Antonia; Villamón, Miguel; González, Luis-Millán; Peset, Fernanda; Morales Aznar, José. ¿Cómo se gestionan los datos de investigación en ciencias de la salud? En: X Foro Internacional sobre la evaluación de la calidad de la investigación y de la educación superior (FECIES) 25-28 junio 2013 Granada, España. Granada: Asociación Española de Psicología Conductual, 2013. ISBN 978-84-695-3701-5. <http://eprints.rclis.org/21010/>

CIENCIAS DEL DEPORTE

- Peset, F.; Ferrer-Sapena, A; Villamón M; González LM; Toca-Herrera JL; Aleixandre-Benavent R. Scientific literature analysis of Judo in Web of Science® *Archives of budo* 2013 Vol. 9, issue 2. <http://eprints.rclis.org/21008/>