



El desafío de compartir datos de investigación en abierto. Un estudio cualitativo sobre conocimientos, creencias y actitudes del personal investigador y/o clínico sobre las células madre cancerosas en España

The challenge of sharing open research data: a qualitative study on the knowledge, beliefs, and attitudes of research and/or clinical staff on cancer stem cells in Spain

AUTORES

- (1,2)** Antonio Vidal-Infer [ORCID: 0000-0002-7860-8652] **(4)** Teresa Samper-Gras [ORCID: 0000-0001-7944-1463]
(1,2,3) Carla Almonacid-Sánchez **(1,2,5)** Rut Lucas-Domínguez [ORCID: 0000-0002-4697-7832]

FILIACIONES

- (1)** Departamento de Historia de la Ciencia y Documentación. Universidad de Valencia. VALENCIA, ESPAÑA.
(2) Unidad de Información e Investigación Social y Sanitaria (UISYS). Universidad de Valencia. VALENCIA, ESPAÑA.
(3) Unidad asociada a INAEU (Instituto Interuniversitario de Investigación Avanzada sobre Evaluación de la Ciencia y la Universidad), UC3M-UAM. MADRID, ESPAÑA.
(4) Instituto Universitario de Estudios sobre las Mujeres. Universidad de Valencia. VALENCIA, ESPAÑA.
(5) CIBERONC. VALENCIA, ESPAÑA.

Los autores declaran que no existe ningún conflicto de intereses

CONTRIBUCIONES DE AUTORÍA

CONCEPTUALIZACIÓN, RECOPIACIÓN DE DATOS, REDACCIÓN-PREPARACIÓN DEL BORRADOR ORIGINAL, REDACCIÓN-REVISIÓN Y EDICIÓN, VISUALIZACIÓN, OBTENCIÓN DE FINANCIACIÓN: A Vidal-Infer, R Lucas-Domínguez.

METODOLOGÍA, INVESTIGACIÓN: A Vidal-Infer, C Almonacid-Sánchez, T Samper-Gras, R Lucas-Domínguez.

FINANCIACIÓN

El trabajo ha sido financiado por MCIN/AE1/10.13039/501100011033 (PID2019 108579RB100).

CORRESPONDENCIA

Rut Lucas Domínguez rut.lucas@uv.es

Avda. Blasco Ibáñez, 15. CP 46010. Valencia, España.

CITA SUGERIDA

Vidal-Infer A, Almonacid-Sánchez C, Samper-Gras T, Lucas-Domínguez R. El desafío de compartir datos de investigación en abierto. Un estudio cualitativo sobre conocimientos, creencias y actitudes del personal investigador y/o clínico sobre las células madre cancerosas en España. Rev Esp Salud Pública. 2026;100: 28 de enero e202601005.

RESUMEN

FUNDAMENTOS // La identificación de células madre tumorales (CSC), involucradas en la resistencia a terapias, progresión tumoral y metástasis, ha generado un crecimiento exponencial en la investigación oncológica. El objetivo de este trabajo fue analizar el conocimiento, creencias y actitudes del personal investigador y clínico español en CSC sobre el uso compartido de los datos de investigación.

MÉTODOS // Se realizaron entrevistas semiestructuradas a una muestra formada por los principales agentes de la investigación española en CSC en base a tres variables: tipo de investigación (básica, clínica, traslacional); experiencia investigadora (junior, con menos de diez años de experiencia; y senior, con más de diez años); y tipología de centro de trabajo (público y privado). El procedimiento de análisis utilizado se basó en la denominada *Teoría Fundamentada*. Siguiendo un criterio de saturación teórica, la muestra final fue de dieciséis personas entrevistadas.

RESULTADOS // Las personas entrevistadas manifestaron tener escaso conocimiento e identificaron poca formación ofertada sobre estos temas. Coincidieron en que la obligación de publicar los datos brutos de investigación en las revistas terminaría por crear un hábito, siempre y cuando se solucionen algunas dificultades, como el coste por procesamiento de artículos (*Article Processing Charges* [APC]). Valoraron positivamente poner los datos brutos a disposición de otros equipos, aunque detectaron gran competitividad en el área científica, que conlleva una cierta reticencia a compartir.

CONCLUSIONES // El estudio desarrollado sobre el uso compartido de datos en CSC pone de manifiesto que los conocimientos son escasos y las creencias muestran algunas amenazas percibidas, aunque las actitudes son positivas respecto a compartir datos de investigación.

PALABRAS CLAVE // Células madre neoplásicas; Difusión de la información; Publicación de Acceso Abierto.

ABSTRACT

BACKGROUND // The identification of cancer stem cells (CSCs), involved in therapy resistance, tumor progression, and metastasis, has led to exponential growth in cancer research. The aim of this study was to analyze the knowledge, beliefs, and attitudes of Spanish research and clinical staff working on CSC regarding the sharing of research data.

METHODS // Semi-structured interviews were conducted with a sample consisting of the key actors in CSC research in Spain, based on three variables: type of research (basic, clinical, translational), research experience (junior with less than ten years of experience, and senior with more than ten years), and type of workplace (public and private). The analysis procedure used was based on *Grounded Theory*. Following a criterion of theoretical saturation, the final sample included sixteen interviewees.

RESULTS // The interviewees reported having limited knowledge and identified a lack of available training on these topics. They agreed that the obligation to publish research data in journals would eventually create a habit, provided certain challenges, such as the cost of Article Processing Charges (APCs), are addressed. They valued making raw data available to other teams positively, although they identified significant competitiveness in the scientific field, leading to some reluctance to share.

CONCLUSIONS // The study on data sharing in CSC reveals that knowledge is limited, and beliefs highlight some perceived threats, although attitudes toward sharing research data are generally positive.

KEYWORDS // Neoplastic stem cells; Information dissemination; Open Access publishing.

INTRODUCCIÓN

La denominada Ciencia Abierta u *Open Science* se refiere al acceso ilimitado a todos los aspectos de la investigación, con el fin de que cualquier persona pueda acceder a ellos libremente. Dentro de la *Open Science*, han aparecido otras modalidades *abiertas*, como el *Open Access*, el *Open Source* y los *Open Data* (en el caso que nos ocupa, *Research Open Data* o datos de investigación en abierto (1)).

Al mencionar este tipo de datos, los datos de investigación, se hace referencia a todo material registrado durante la investigación, una definición consensuada y reconocida por la comunidad científica (2) y por organismos como los Institutos Nacionales de Salud NIH de Estados Unidos (3) o la Organización para la Cooperación y el Desarrollo Económicos (OCDE) (4). En esta línea, el llamamiento en favor de los datos abiertos no sólo tuvo eco en la comunidad científica, sino también entre los editores (5,6) y los organismos de financiación (7,8). Se diseñaron políticas para fomentar la transparencia y la apertura de la investigación científica, como las *Directrices para la Promoción de la Transparencia y la Apertura TOP* (9) y los principios FAIR o *Findable, Accessible, Interoperable, and Reusable*, que hacen hincapié en la apertura y la reutilización de los datos de investigación a través de una gestión eficaz de los datos (10,11).

Desde el anterior programa marco de la Unión Europea, el programa *Hori-*

zonte 2020, se recoge como requisito en los proyectos financiados que los datos de investigación sean publicados en abierto (12). La Comisión Europea insta a las organizaciones, entidades e instituciones que realicen actividades investigadoras que tengan un plan de gestión de datos o DMP, entendido como el ciclo de vida de los datos que se recopilarán, procesarán y/o generarán en un proyecto de investigación, incluyendo información sobre el manejo de datos de investigación durante y después del final del proyecto, sobre qué datos se recopilarán, procesarán y/o generarán, qué metodología y estándares se aplicarán, si los datos se compartirán o se harán de acceso abierto y cómo se conservarán los datos después de la finalización del proyecto. De hecho, las convocatorias europeas de proyectos de investigación tienen como requisito que todos los proyectos financiados tengan estipulado un DMP, también conocido más recientemente como *Research Data Management* (13).

Una de las áreas más relevantes para la compartición de datos es el de las células llamadas Células Madre Cancerosas o CSC. El cáncer es un concepto genérico que engloba una gran cantidad de patologías caracterizadas por el desarrollo de células anormales, que se dividen, crecen y se diseminan sin control a cualquier parte del cuerpo (14). Los tumores evolucionan con el tiempo, produciendo diferencias genéticas y fenotípicas entre tumores encontrados en diferentes tejidos, así como entre el mismo tipo de tumor en diferentes indi-

Este artículo tiene una licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional. Usted es libre de Compartir (copiar y redistribuir el material en cualquier medio o formato) bajo los siguientes términos: Atribución (debe darse el crédito apropiado, proporcionar un enlace a la licencia e indicar si se realizaron cambios. Puede hacerlo en cualquier manera razonable, pero no de alguna manera que sugiera que el licenciente lo respalda a usted o su uso); No comercial (no podrá utilizar el material con fines comerciales); Sin derivados (si remezcla, transforma o construye sobre el material, no puede distribuir el material modificado); Sin restricciones adicionales (no puede aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros hacer cualquier cosa que la licencia permita). <https://creativecommons.org/licenses/by-nc-nd/4.0/>

viduos (heterogeneidad tumoral), lo que complica el diagnóstico, pronóstico y tratamiento (15,16). Cada tumor es único, por lo que se da una gran variedad de tipos y subtipos, y existe evidencia sólida que indica que la quimiorresistencia, la progresión tumoral y la metástasis están vinculadas a la subpoblación de CSC. Esto explica por qué la hipótesis de la CSC se encuentra en el centro de un campo en rápida evolución que puede desempeñar un papel fundamental para cambiar el modo en que los investigadores básicos de cáncer, los investigadores clínicos, los médicos y los pacientes tratan esta enfermedad, optimizando recursos y tiempo para conseguir una atención más eficiente, lo que repercutiría en una mejor toma de decisiones asistenciales y preventivas en el ámbito de la Salud Pública

Paradójicamente, mientras el personal investigador y las instituciones científicas se ven obligados a adaptarse a criterios de evaluación tradicionales basados en las patentes y el impacto de las publicaciones, los financiadores y otros agentes abogan por abrir el acceso a los datos provenientes de la investigación. Sin embargo, si no hay un reconocimiento intelectual al respecto en términos de citas e impacto, será muy difícil que el uso compartido de datos se consolide para poder obtener los beneficios potenciales esperados (17,18). Otro de los desafíos a que se enfrenta este proceso es la falta de interoperabilidad, las fronteras geográficas y todas aquellas cuestiones relacionadas con la confidencialidad y la protección de datos (19).

Con todo ello, el presente trabajo tuvo como objetivo conocer y valorar los conocimientos, creencias y actitudes del personal investigador y clínico

español en CSC sobre el uso compartido de datos, para así explorar las ventajas, barreras e inconvenientes existentes en la cultura de compartir datos de investigación en este colectivo.

MATERIAL Y MÉTODOS

Se realizó un estudio cualitativo basado en la *Teoría Fundamentada* (Glaser y Strauss, 2017), por el cual se recogieron entrevistas que permitieran conocer la perspectiva de los principales agentes de la investigación española en CSC, sus conocimientos, creencias y actitudes, ante el uso compartido de datos de investigación, para así comprender los significados que atribuían a sus comportamientos. En lugar de contar con hipótesis previamente establecidas, se establecieron áreas de interés y ejes analíticos que se fueron completando con otros que emergieron en el trabajo de campo. La interpretación se basó en la contrastación de similitudes y diferencias sobre diferentes categorías semánticas que permitieron construir conceptos y sus propiedades a partir de los datos recogidos.

Sujetos. Se estableció la muestra en función de tres áreas de representatividad, con tal de recoger diferentes planteamientos relevantes para el estudio: tipo de investigación (básica, clínica, traslacional); experiencia investigadora (junior, con menos de diez años de experiencia; y senior, con más de 10 años); y tipología de centro de trabajo (público, privado). Asimismo, tal y como se recomienda desde el *X Plan Estatal de Investigación Científica y Técnica y de Innovación* (21) y el *Programa Marco Horizonte Europa* (22), se consideró el sexo de las personas a entrevistar. La estrategia de selección muestral trató de evitar sesgos en las características repre-

sentadas, pero no permitió analizar por cada una de ellas, si bien para la exposición de resultados se decidió ofrecer una identidad que permitiera, durante la lectura de las citas o verbatim, recuperar fácilmente las características de las personas entrevistadas.

Una vez identificados los grupos, las personas entrevistadas se seleccionaron mediante muestreo cualitativo por designación en primera instancia y por bola de nieve en segunda. La incorporación de nuevas personas entrevistadas finalizó al producirse la saturación teórica, es decir, cuando la información comenzó a ser repetitiva y no había aportaciones significativas [TABLA 1].

Instrumento. Cada entrevista semiestructurada se inició con una cuestión general que recogía el propósito de la entrevista, combinada con preguntas específicas que guiaban hacia los temas relevantes, objetivos de la investigación. En este caso, se propuso una estructura de preguntas basada en recoger conocimientos, creencias y actitudes ante el uso compartido de datos de investigación, convenientemente ordenadas para que facilitara la comparación de las respuestas, a la vez que lo suficientemente abiertas para que permitiera la emergencia de cuestiones de interés y profundizar sobre ellas. Como medida complementaria para reducir posibles sesgos, en algunas preguntas se solici-

Tabla 1
Perfiles de las personas entrevistadas identificados con código de entrevista.

ID	Sexo	Experiencia	Tipo de investigación	Tipología de centro
E01 HSBP	Hombre	Senior	Básica	Público
E02 HJBP	Hombre	Junior	Básica	Público
E03 MSBPr	Mujer	Senior	Básica	Privado
E04 HSCPr	Hombre	Senior	Clínica	Privado
E05 MSTP	Mujer	Senior	Traslacional	Público
E06 HJBPr	Hombre	Junior	Básica	Privado
E07 MSBP	Mujer	Senior	Básica	Público
E08 MSBP	Mujer	Senior	Básica	Público
E09 MJBP	Mujer	Junior	Básica	Público
E10 HSCP	Hombre	Senior	Clínica	Público
E11 MJBP	Mujer	Junior	Básica	Público
E12 MSTP	Mujer	Senior	Traslacional	Privado
E13 MSBPr	Mujer	Senior	Básica	Privado
E14 HSBPr	Hombre	Senior	Básica	Privado
E15 MSBPr	Mujer	Senior	Básica	Privado
E16 MSBP	Mujer	Senior	Básica	Público

taba información específica (ejemplos y detalles) para que el sujeto elaborara más su respuesta.

La elaboración del guion de preguntas se estableció por un panel de

expertos con las fuentes secundarias existentes en cuanto al uso compartido de datos, los estudios científicos relacionados, la literatura científica y el resumen de resultados de estudios previos [TABLA 2].

Tabla 2
Guion de la entrevista semiestructurada.

CONOCIMIENTOS

Contextualización de los datos de investigación.

Proyectos/programas/publicaciones en revistas en que participa donde se recomienda o exige que se lleve a cabo esta práctica.

Papel de la institución/organización a la hora de ofrecer recursos en forma de guías, soporte técnico, formación o supervisión.

Diferencia entre financiación pública y privada en los datos brutos.

Conocimiento sobre infraestructuras digitales para compartir datos.

Planes y protocolos.

Incentivos.

CREENCIAS

Desde una perspectiva global de la ciencia, posible importancia de compartir datos.

En su área, posible importancia de compartir datos.

Dificultad para integrar esta práctica.

Valoración de los incentivos, coste-beneficio.

Barreras tecnológicas (accesibilidad y sencillez de las infraestructuras).

Apoyo institucional, valoración.

Aspectos positivos y negativos según su opinión. Ventajas e inconvenientes.

ACTITUDES

Nivel de disposición a la hora de compartir.

Preferencias técnicas y formales relacionadas con el proceso de compartir los datos.

Qué debería tener un *dataset* para ser comprensible, reutilizable e interoperable.

Las entrevistas semiestructuradas tuvieron una duración aproximada de treinta minutos y se realizaron de forma telemática por videoconferencia, asegurando en el contacto previo la confidencialidad y anonimato a la persona a entrevistar mediante consentimiento informado. El mismo equipo investigador, con formación en Ciencias Sociales, realizó las entrevistas.

Análisis de datos. Como se ha comentado anteriormente, el procedimiento de análisis utilizado se basó en la denominada *Teoría Fundamental* (20). Se trata de una técnica de análisis de datos cualitativos en la que, a partir de las notas de campo y las transcripciones literales de los discursos, se ha procedido a codificar la información obtenida, intentando dar una denominación común a un conjunto de segmentos de texto que comparten una misma idea. La segmentación y posterior codificación de las transcripciones de las entrevistas realizadas se llevó a cabo con el *software Atlas.TI* por parte de una investigadora, miembro del equipo, con formación en Ciencias Sociales y en el uso de esta herramienta. Las transcripciones y la codificación se llevaron a cabo de manera secuencial a medida que las entrevistas estaban siendo realizadas, con el fin de determinar cuándo se alcanzaba la saturación teórica y, por tanto, finalizar la realización de las entrevistas. Una vez finalizada la codificación, se integraron las categorías obtenidas para afinar la comprensión del tema de estudio en el contexto en que habían sido recogidas. Según el enfoque de la *Teoría Fundamental*, luego del análisis, se elaborarán conceptos y proposiciones partiendo directamente de los datos y no de supuestos *a priori*. En el apartado de discusión, contrastaremos las respues-

tas obtenidas con estudios precedentes a modo de comparativa para concluir con una propuesta de modelo teórico de compartición de datos.

Se resolvió que, para la exposición de los resultados, se presentarán citas textuales de las personas entrevistadas para ilustrar el concepto e ideas localizadas en el análisis; para así, en el apartado de discusión, valorar la oportunidad de dichos conceptos e ideas.

RESULTADOS

Conocimientos. Cuando se trató el tema del conocimiento de infraestructuras digitales especializadas en compartir datos en abierto, las personas entrevistadas percibieron un conocimiento limitado, al menos, en concreto, para el área oncológica [TABLA 3]. Por ejemplo, solo se mencionó uno de los repositorios más relevantes, en concreto el *Gene Expression Omnibus* (23). Una dificultad añadida fue la interfaz gráfica de los repositorios, la cual no todo el personal entrevistado la percibieron sencilla; ahora bien, la mayoría de estas experiencias fueron referidas a la búsqueda de datos.

Con respecto a la función de la institución sobre información, formación y apoyo acerca de la gestión de datos de investigación y el uso de datos compartidos, las respuestas fueron semejantes entre sí. Destacó la inexistencia o la no constancia de que existieran protocolos de gestión de datos o DMP.

La mayoría de las personas entrevistadas, en concreto doce de las dieciséis, coincidieron en el hecho de que cada vez es más frecuente la recomendación, en ocasiones obligación, de publicar los datos brutos de inves-

Tabla 3
Conocimientos y experiencias en compartición de datos sobre Células Madre Cancerosas.

1.1. Relación con el uso de repositorios y datos compartidos.

- "...no conozco ninguna base de datos de células madre [...] cuando he buscado algo de células madre lo he buscado todo en bibliografía, desconozco esos portales." (E10 HSCP)*
- "No somos un grupo bioinformático, con lo cual a veces descargamos en colaboradores para que lo hagan o incluso a una empresa." (E01 HSBP)*
- "No es un proceso sencillo. Nosotros hemos intentado depositar datos de secuenciación, y yo en mi caso he necesitado ayuda para saber dónde depositarlos, cómo hacerlo." (E02 HJBP)*
- "...estoy contento porque veo que son herramientas super útiles para los investigadores [...] mi experiencia es positiva dentro de mi ignorancia bioinformática." (E01 HSBP)*
- "No tengo demasiado conocimiento [...] yo la verdad, conozco GEO y The Cancer Genoma Atlas y poco más." (E11 MJBP)*
- "...muchas veces están en formato que no es el que, si no te dedicas a informática o específicamente a trabajar con datos es bastante complicado [...] a veces no sabes muy bien cómo han llegado a obtener ese valor." (E07 MSBP)*
- "Mi experiencia no ha sido mala [...] la información de los datos está muy plana puesta para que todo el mundo lo entienda." (E03 MSBPR)*
- "No son difíciles encontrarlos si sabes lo que buscas." (E08 MSBP)*

1.2. Formación y apoyo de la institución

- "Es muy posible que no lo tenga [...] Yo creo que no hay posicionamiento como tal." (E04 HSCP)*
- "No tenemos infraestructura propia, entonces no tenemos unos servidores que yo pueda ofrecer a mis grupos para que me suban la información y yo pueda controlar el acceso." (E05 MSTP)*
- "La institución como fundación de investigación que yo sepa no [...] yo llevo cinco años y ya voy para el sexto y no hay ningún tipo de formación [...] cada grupo trabaja en su carpeta compartida conforme le parece." (E11 MJBP)*

1.3. Compartir datos bajo demanda

- "Cada vez es un requisito mayor, tanto de las revistas como de entidades financiadoras." (E02 HJBP)*
- "Creo que está aumentando en los últimos años. [...] Pero creo que la obligatoriedad debe ir acompañada de la formación." (E02 HJBP)*
- "[...] sí, es obligatorio. Creo que es lo que debe de ser." (E03 MSBPR)*
- "Por una parte, estoy de acuerdo, porque me parece bien que se comparta toda la información, pero por otra también me parece excesivo que tengas que publicar resultados que ya te han costado un esfuerzo, un trabajo y un dinero, y tengas que pagar todo lo que se está pagando en las revistas [...] lo veo un negocio... esto tendría que estar de alguna forma también regulado." (E07 MSBP)*

tigación en las revistas, como forma de crear un hábito para generar un cambio, es decir, que como el requerimiento cada vez será mayor, ello hará que el personal investigador se acostumbre a publicar sus datos en abierto y cada vez sea más habitual.

Creencias. En el apartado de creencias se trataron diversos bloques temáticos. Por un lado, se habló de la importancia que le otorga el personal investigador al hecho de compartir datos en abierto, aspectos que se entrelazaron con otro de los temas a tratar: aspectos positivos y ventajas de la ciencia abierta [Tabla 4]. Todos los discursos expresados en las entrevistas consideraron importante poner los datos brutos a disposición de otros equipos investigadores, fundamentalmente por la utilidad a la hora de comprobar y contrastar resultados de estudios y ensayos clínicos, así como el ahorro de recursos y la prevención de duplicidades en los estudios, asunto remarcable para el área oncológica. De los dos aspectos positivos recogidos anteriormente se llega a una de las ventajas más destacadas por las personas entrevistadas: entender la ciencia como colaborativa para un avance significativo.

Por otro lado, en este apartado se abordaron los inconvenientes, las dificultades y las barreras, uno de los apartados más extensos de las entrevistas. Para comenzar, se percibió una gran competitividad en el área científica, que conlleva una cierta reticencia a compartir los datos que se hayan podido obtener, ya que se destacó el papel de la información como fuente de poder, como una herramienta para crecer, conseguir publicaciones relevantes y recursos económicos. Frente a la ventaja de una ciencia colaborativa,

emergió la individualización de la ciencia como barrera y aspecto a mejorar dentro del ámbito científico. Asimismo, en términos materiales, las limitaciones de los recursos tecnológicos y la infraestructura computacional fue otra de las barreras con las que se encuentra el personal investigador, pues el hecho de no contar con herramientas adecuadas para la gestión de los datos complicaba el proceso. Se desprendió de las entrevistas que compartir datos suponía un esfuerzo añadido.

Una de las barreras más repetidas a lo largo de las entrevistas fue la heterogeneidad de los datos, los metadatos y su formato. La poca estructuración y homogeneización hacen que el proceso, tanto de compartir datos como de su reutilización, sea más complicado, ya que en ocasiones se necesita un *software* específico para la recuperación de datos o para cambiar el formato de éstos y poder compartirlos en un formato recomendado y libre.

Cuando se habla de inconvenientes y barreras, las personas entrevistadas también reconocieron miedos y preocupaciones ante el robo de datos o la preocupación de que los datos fueran utilizados de una manera poco ética.

Se preguntó por la percepción del personal investigador ante el apoyo por parte de las instituciones en lo referido a Ciencia Abierta y por si tenían información o conocían iniciativas. Algunas de las personas entrevistadas percibían que sí que se están llevando cabo algunas iniciativas, pese a que anteriormente habían indicado no tener constancia de que en su propia institución fuera así. De hecho, no se refirieron a su entorno laboral, sino al ámbito institucional oficial.

Tabla 4
Creencias en compartición de datos sobre Células Madre Cancerosas.

2. CREENCIAS

2.1. La importancia de compartir datos en abierto: ventajas

- "...todo lo que se genere se pueda compartir, para poder sumar y avanzar." (E14 HSBPR)*
- "Creo que sí, por supuesto, a mí me ha servido muchísimo tener el repositorio del TCGA [The Cancer Genoma Atlas] porque he podido realizar mis análisis, comprobar que me estaba saliendo en mis pacientes [...] es muy útil." (E11 MJBP)*
- "Creo que con fondos públicos los datos deberían de ser obligatoriamente abiertos, porque si no corremos el riesgo de hacer cinco veces el mismo estudio en cinco sitios diferentes [...] creo que en el cáncer es una de las que más." (E15 MSBP)*
- "...es básico porque el tiempo que tu tardas en reunir datos de este tipo de pacientes, es una cantidad ingente de tiempo..." (E16 MSBP)*
- "...cuando tú haces un estudio con 100 muestras, no es lo mismo que tener 100.000 y si lo que quieres es buscar patrones de respuesta..." (E03 MSBP)*
- "Si cedes tus datos vas a estar citado en otras publicaciones, que también te interesa mucho, y posiblemente vas a poder tener colaboraciones que si no hubieras enviado esos datos a un repositorio no los ibas a tener." (E08 MSBP)*

2.2. La importancia de compartir datos en abierto: amenazas, barreras e inconvenientes

- "¿Cuál es la principal barrera? Yo no quiero compartir la información contigo, por si tú publicas una cosa relevante antes que yo." (E04 HSCPR)*
- "...tienes que estar produciendo constantemente, tienes un elevado nivel de burocracia que te absorbe una cantidad de tiempo enorme y yo creo que uno de los motivos por los que no se comparten datos es por no tener que hacer el esfuerzo de compartirlos." (E02 HJBP)*
- "...vale, yo comparto mis datos, pero cómo, cuándo y dónde" (E02 HJBP)*
- "...cada uno tiene que quedarse con sus datos hasta que los publica... cada uno está mirando por sus intereses propios" (E10 HSCP)*
- "... si te adelantas mucho en poner los datos en abierto pues estas dando una ventaja a los que están trabajando en tu mismo campo." (E12 MSTPr)*
- "Cada persona comparte los datos de una manera distinta, con un formato diferente..." (E11 MJBP)*
- "...te encuentras que tienes problemas para interpretar o para el material que estas usando, salen resultados y no tienen sentido, y es porque no tenías la información adecuada desde el principio." (E09 MJBP)*
- "No son user friendly y no son intuitivas o fáciles de acceder." (E09 MJBP)*
- "...que se puedan emplear de manera incorrecta los datos que tú has cedido y al final puedas verte reflejado en estudios que no se han hecho correctamente..." (E08 MSBP).*

2.3. Incentivos

- "...algún tipo de reconocimiento en la propia base de datos... que figurara tu nombre de alguna manera o tu institución o tu grupo de investigación como contribuyente." (E10 HSCP)*
- "Creo que el incentivo a estos niveles es la colaboración top... que la persona que tienes al otro lado o que la red en la que vas a participar sea de prestigio." (E05 MSTP)*
- "...el mayor incentivo es que te penalice no poner la gestión de datos en un proyecto." (E12 MSTPR)*
- "...debería ser inherente a la investigación, es decir, yo es una cosa que no me planteo..." (E16 MSBP)*

En este bloque también se abordó el tema de los incentivos al personal investigador por compartir sus datos de investigación y, cuando se preguntó por ello, la respuesta es que dichos incentivos no existían. En el caso de las personas entrevistadas en este estudio, apuntaron también a que el incentivo ha de ser como el inicio para crear una cultura de la Ciencia Abierta. Finalmente, algunas de las personas entrevistadas entendieron la práctica de compartir datos como una práctica inherente a la investigación, que se debe de interiorizar y hacerlo como una parte más del proceso de investigación.

Actitudes. Por último, se trataron las actitudes y se preguntó por dos cuestiones, fundamentalmente: disposición a compartir datos y preferencias tanto de infraestructura como metodológicas para llevar a cabo esta práctica, es decir, cómo les gustaría que fuese el modelo. Si bien durante la entrevista fueron surgiendo indicios sobre la disposición a compartir datos de cada persona entrevistada, no obstante se realizó la pregunta de una manera directa en el apartado de actitudes. Así se pudo contrastar el discurso entre las creencias con respecto a la Ciencia Abierta y la disposición real que existía a compartir los datos de investigación.

Los argumentos para una mayor tendencia y disposición a compartir datos de investigación estuvieron relacionados con las ventajas citadas anteriormente y la importancia otorgada al hecho de compartir; para las mismas personas entrevistadas fue destacable el argumento de una ciencia colaborativa para avanzar, especialmente en el área oncológica. Quienes se situaron con una menor disposición a compartir sus datos de investigación, argumenta-

ron con la competitividad y el miedo al robo o mal uso de los datos.

En general, entendieron que una mejor infraestructura para gestionar los datos haría el proceso más sencillo y ordenado, por lo que tomaron fuerza los argumentos referidos a una mayor financiación para *Open Science* en los centros de investigación. En esa misma línea también, se aseguró contar con personal especializado en gestión de datos, así como personal de ayuda en caso de necesitarla para resolver dudas del proceso. En este punto, se indicó que una mayor formación del personal investigador sobre la gestión de datos puede hacer crecer la práctica de compartir datos de investigación y darles autonomía.

Si bien en el apartado de barreras se habló de lo complejo que es manejar en algunos casos los repositorios, lo poco intuitivos que resultan, en esta sección se hizo alusión a que las interfaces fueran sencillas, fáciles de utilizar y con pocas barreras de acceso. Con respecto a la estandarización y homogeneización de los datos y las variables, así como sus formatos, las personas entrevistadas remarcaron que los datos compartidos estuviesen siempre en los mismos formatos y que fueran libres, esto es, que no hubiese la necesidad de *softwares* privativos para su transformación. Con respecto a las nomenclaturas se solicitó que fuesen estándares, así como el establecimiento de un conjunto mínimo de variables para que los datos puedan ser interoperables. Con respecto a los metadatos y la información que contienen, se hizo referencia a que la información debía de ser completa para que los datos puedan ser reutilizables, información referida a los métodos utilizados en el estudio, junto con las variables explicadas.

Algunas de las personas entrevistadas hicieron referencia al acceso e identificación a las plataformas para compartir datos. Por una parte, se situaron quienes manifiestan una tendencia a que el acceso sea lo más libre posible y, por otra, algunos de los discursos apuntaron a aportar una mayor información de quien facilita sus datos o que pueda haber constancia de un registro de quien descarga los datos

[TABLA 5].

DISCUSIÓN/CONCLUSIONES

En buena parte de centros especializados en el área oncológica, por ejemplo, el Centro Nacional de Investigaciones Oncológicas (CNIO), no se recoge una sección de recursos en sus respectivos sitios web donde se mencionen estos repositorios. Esta idea es relevante para tratar de comprender que los motivos que alegan las personas entrevistadas para haber aprendido el funcionamiento de estas infraestructuras es la experiencia investigadora y la necesidad de su uso.

La poca mención a uso de repositorios pone de manifiesto que, en la actualidad, se requiere de una formación específica por parte del usuario, tal y como detallan González-Pérez *et al.* (24)

Las ventajas de la compartición de datos para el área de las CSC que destacan las personas entrevistadas ya fueron resumidas por Sixto-Costoya *et al.* (25) El *data sharing* en la investigación oncológica podría suponer un ahorro de recursos, tanto económicos como de tiempo, tal y como se recoge en Rung y Brazma (26), de manera que dichos recursos podrían ser destinados a la obtención de un mayor avance. De hecho, como se ha podido constatar

en el estudio, la compartición de datos supone un avance significativo para entender la ciencia como colaborativa. Sin embargo, se advirtió, en alguno de los discursos, una disonancia entre la importancia de compartir datos o las ventajas que esto conlleva, con la disposición a llevar a cabo esta práctica, sobre todo la idea de compartir los propios datos.

Por otro lado, las barreras que encontraron las personas entrevistadas son diversas y estuvieron en consonancia con las ya descritas en estudios anteriores (27,28). Entre los principios *FAIR*, se encuentra la interoperabilidad, es decir, que los conjuntos de datos no requieran poseer *software* propietario, que los datos sean legibles, bien definidos por metadatos y que, por tanto, sea posible manejarlos de manera adecuada (10). Otro aspecto que considerar como posible barrera para compartir datos de investigación tiene que ver con la industria editorial y las denominadas publicaciones *Gold Open Access*. Esta tipología de revistas se caracteriza por cobrar APC a los autores para publicar en acceso abierto, en ocasiones con costes bastante elevados. Este gasto, unido al esfuerzo extra para depositar los datos en abierto, podría ser disuasorio, tal y como se muestra en el análisis de las entrevistas, ya que, en opinión de las personas entrevistadas, este proceso debería regularse, bien económicamente, bien con formación específica, para facilitar el depósito de datos en abierto.

En línea con el presente trabajo, la encuesta llevada a cabo por Goodey *et al.* (27) muestra que el 75% de los investigadores aseguraron recibir muy poco crédito por compartir sus datos abiertamente y proponían como posi-

Tabla 5
Actitudes en compartición de datos sobre Células Madre Cancerosas.

3. ACTITUDES

3.1 Disposición a compartir datos

"...con los datos que tú tienes hay gente que pueda sacarles más partido que tú o darle un enfoque distinto al que le has dado tú." (E04 HSCPR)

"Si los comparte, todos podremos sacar beneficio y avanzar más rápidamente, o sea, que yo estoy totalmente a favor." (E07 MSBP)

"...la investigación en Oncología, que considero un problema importante, no se puede hacer solo [...] por eso intentamos transferir la información que yo he obtenido a otros grupos para que ellos me den también información." (E03 MSBPR)

"En los entornos más de investigación básica... sí que están concienciados de compartir datos... en entornos clínicos, es diferente, porque en muchas ocasiones ha habido una tendencia a atesorar los datos y a no compartíroslos, como una herramienta de poder." (E14 HSBPR)

"...si no nos lo piden las revistas en las que publicamos, no los compartimos. [...] te cuesta un poco porque cuesta mucho obtener datos. [...] lo haces porque te están obligando porque, si no, no vas a poder enviarlo a esa revista". (E08 MSBP)

"...mi colección de muestra, mi paciente, que me ha costado muchos años y mucho sudor conseguir, entonces soy muy reticente para eso... si tengo una cohorte super valiosa porque este tipo de cáncer es muy minoritario y tengo muchas muestras, la cuido con muchísimo mimo, entonces miro mucho qué hago con eso y cómo lo hago." (E05 MSTP)

3.2. Preferencia de modelo para compartir de datos

"...a los centros de investigación les das dinero para eso sí o sí, no que haga falta pedir un proyecto específico para conseguirlo [...] contratar personal específico para gestión de datos en los centros... o tener personal en las agencias financiadoras que te ayuden." (E01 HSBP)

"...es importante que se desarrolle formación en este ámbito para poder dar autonomía a los investigadores." (E02 HJBP)

"Requiere unos niveles de inversión que escapa a los simples mortales, con lo cual tiene que ser una propuesta de país." (E05 MSTP)

"...estándares... cuál es el conjunto mínimo de variables que vamos a exigir para que puedan ser precisamente los datos compartibles." (E14 HSBPR)

"Una única base de datos donde estuviera todo... donde de una manera muy sencilla pudieras buscar absolutamente todo, sin usuarios ni contraseñas, ni registros, abierta, libre y muy sencilla." (E10 HSCP)

"Creo que la infraestructura hace falta y es lo crítico y es lo primero". (E05 MSTP)

"...contactar con alguien que te lo pueda explicar o que puedas comentar las dudas." (E07 MSBP)

"...que hubiera alguien siempre que pudiera contestar sobre ello." (E15 MSBPR)

"...consultas con una interfaz sencilla y que el output sea alcanzable para cualquiera" (E12 MSTPR)

"...variables básicas que deberían de estar siempre en todos los estudios con diseño claro de que cual era el objetivo del estudio cuando se hicieron estos datos a que conclusiones se llegó." (E15 MSBPR)

"Que te tuvieses que identificar... quién eres, donde estás, tus datos... y que alguien lo revisase también para comprobar que todo eso está correcto." (E07 MSBP)

bles incentivos la citación de sus artículos de investigación (67%), el mayor impacto y visibilidad de sus artículos (61%) y alguna forma de beneficio público o mandato de la revista/editor (ambos, 56%).

Si bien hay encuestas que han tratado y recogido los comportamientos del personal investigador en CSC acerca de la compartición de datos, el presente estudio en base a entrevistas semiestructuradas nos permite valorar cómo entienden sus conductas, una información útil para diseñar y programar políticas de compartición de datos.

A modo de conclusiones, señalar que el modelo teórico para compartir datos en CSC debería tener como objetivo facilitar la colaboración y el intercambio de información entre investigadores, mejorando así la eficiencia y el impacto de la investigación oncológica. Por ello, se propone un modelo para compartir datos que contemple una mayor finan-

ciación para *Open Science* en centros de investigación, acompañada de infraestructuras más adecuadas para la gestión de datos. El modelo debe incluir un plan de capacitación sobre gestión de datos dirigido al personal investigador y unos recursos con una mayor usabilidad y accesibilidad, promoviendo interfaces sencillas y fáciles de utilizar, así como la estandarización y compatibilidad de los conjuntos de datos a través del uso de formatos libres. Estas características apuntalarán un modelo teórico de compartición de datos en CSC más accesible, estandarizado y bien respaldado en términos de recursos y conocimientos, lo que implica una optimización de recursos en Salud Pública y, por tanto, una mejor atención a la ciudadanía. En esta línea, y como fruto de los resultados de la investigación, se ha generado una guía de recomendaciones para el uso compartido de datos de investigación en CSC, que se encuentra depositada y es accesible gratuitamente desde el repositorio Zenodo (29). ©

BIBLIOGRAFÍA

1. Lucas-Dominguez R, Alonso-Arroyo A, Vidal-Infer A, Aleixandre-Benavent R. *The sharing of research data facing the COVID-19 pandemic*. *Scientometrics* 2021;126:4975-4990. <https://doi.org/10.1007/s11192-021-03971-6>
2. Peset F, Aleixandre-Benavent R, Blasco-Gil Y, Ferrer-Sapena A. *Datos abiertos de investigación: camino recorrido y cuestiones pendientes*. *Anales de Documentación*. 2017;20.
3. US National Institutes of Health. *Final NIH statement on sharing research data*. 2003. Disponible en: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>
4. OECD. *OECD principles and guidelines for access to research data from public funding*. Paris: OECD Publishing; 2007. <https://doi.org/10.1787/9789264034020-en-fr>
5. Nature. *Calling all coronavirus researchers: keep sharing, stay open*. *Nature*. 2020;578(7793):7.
6. Science. *Editorial policies: Data and code deposition*. 2023. Disponible en: <https://www.science.org/content/page/science-journals-editorial-policies#data-and-code-deposition>
7. US National Institutes of Health. *Data management & sharing policy overview*. 2023. Disponible en: <https://sharing.nih.gov/data-management-and-sharing-policy/about-data-management-and-sharing-policies/data-management-and-sharing-policy-overview#before>
8. Research Councils UK. *Open research data concordat*. 2016. Disponible en: <https://russellgroup.ac.uk/news/open-research-data-concordat/>
9. Center for Open Science. *Top-Guidelines*. 2020. Disponible en: <https://www.cos.io/initiatives/top-guidelines>
10. European Commission. *Open Research Europe. Open Data, Software and Code Guidelines. The FAIR Data Principles*. 2024. Disponible en: <https://open-research-europe.ec.europa.eu/for-authors/data-guidelines#fairdata>
11. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A et al. *The FAIR guiding principles for scientific data management and stewardship*. *Sci Data*. 2016;3:160018.
12. European Union. *H2020 Open Access & Data Management*. 2024. Disponible en: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm
13. OpenAire. *How to comply with Horizon Europe mandate for Research Data Management*. 2024. Disponible en: <https://www.openaire.eu/how-to-comply-with-horizon-europe-mandate-for-rdm>
14. Hanahan D, Weinberg RA. *Hallmarks of cancer: the next generation*. *Cell*. 2011 Mar 4;144(5):646-674. doi: <https://doi.org/10.1016/j.cell.2011.02.013>. PMID: 21376230.
15. Burrell RA, McGranahan N, Bartek J, Swanton C. *The causes and consequences of genetic heterogeneity in cancer evolution*. *Nature*. 2013;501(7467):338-345.
16. Greaves M, Maley CC. *Clonal evolution in cancer*. *Nature*. 2012; 481(7381):306-313.
17. Ball A, Duke M. *How to track the impact of research data with metrics*. Edinburgh: Digital Curation Centre; 2015. Disponible en: <http://www.dcc.ac.uk/resources/how-guides>
18. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ et al. *Promoting an open research culture*. *Science*. 2015;348(6242):1422-1425.
19. Komorowski M, Kraemer MUG, Brownstein JS. *Sharing patient-level real-time COVID-19 data*. *Lancet Digit Health*. 2020;2(7).
20. Glaser B, Strauss A. *Discovery of grounded theory: Strategies for qualitative research*. Routledge. 2017.
21. Ministerio de Ciencia e Innovación. *Estrategia Española de Ciencia, Tecnología e Innovación 2021-2027*. [Internet]. Disponible en: <https://www.ciencia.gob.es/Estrategias-y-Planes/Planes-y-programas/PEICTI.html>. Consultado el 16 de octubre de 2024.
22. *Horizonte Europa. Información general sobre el Programa Marco de la UE*. [Internet]. Disponible en: <https://www.horizonteeuropa.es>. Consultado el 16 de octubre de 2024.
23. US National Institutes of Health. *Gene Expression Omnibus-GEO*. 2024. Disponible en: <https://www.ncbi.nlm.nih.gov/geo/>

- 24.** González-Pérez LI, Ramírez-Montoya M, García-Peñalvo FJ. *User experience in institutional repositories: a systematic literature review*. *Int J Hum Cap Inf Technol Prof*. 2018;9(1):70-86.
- 25.** Sixto-Costoya A, Aleixandre-Benavent R, Vidal-Infer A, Lucas-Domínguez R, Castelló-Cogollos L. *Data sharing: qué son y cómo se pueden compartir los datos de investigación. Manual de recomendaciones para gestores de la información*. Sociedad Española de Documentación e Información Científica. 2019.
- 26.** Rung J, Brazma A. *Reuse of public genome-wide gene expression data*. *Nat Rev Genet*. 2013;14(2):89-99.
- 27.** Goodey G, Hahnel M, Zhou Y, Jiang L, Chandramouliswaran I, Hafez A *et al.* *The state of open data 2022 (Version 5)*. *Digital Science*. 2022. <https://doi.org/10.6084/m9.figshare.21276984.v5>
- 28.** Aleixandre-Benavent R, Vidal-Infer A, Alonso-Arroyo A, Peset F, Ferrer-Sapena A. *Research data sharing in Spain: exploring determinants, practices, and perceptions*. *Data*. 2020;5(29).
- 29.** Lucas-Domínguez R, Vidal-Infer A. *Compartir datos de investigación sobre células madre tumorales. Guía de recomendaciones*. Zenodo. 2024. <https://doi.org/10.5281/zenodo.10726769>