



VNIVERSITAT
DE VALÈNCIA

FACULTAT DE MEDICINA Y ODONTOLOGÍA

Programa de Doctorado 3139 en Medicina

La investigación española en Ciencias de la Salud frente al uso compartido de datos: una aproximación cualitativa

Presentada por:

Andrea Sixto Costoya

Dirigida por:

Rafael Aleixandre Benavent

Antonio Vidal Infer

Valencia, junio de 2021



El Dr. D. Rafael Aleixandre Benavent, científico titular del Centro Superior de Investigaciones Científicas (CSIC) y el Dr. D. Antonio Vidal Infer, profesor titular del Departamento de Historia de la Ciencia y Documentación de la Facultad de Medicina y Odontología de la Universitat de València

CERTIFICAN:

Que la presente tesis doctoral titulada: **“La investigación española en Ciencias de la Salud frente al uso compartido de datos: una aproximación cualitativa”**, presentada por Andrea Sixto Costoya ha sido realizada bajo nuestra dirección y cumple con los requisitos para que su autora pueda obtener con ella el grado académico de Doctora por la Universitat de València.

Y para que así conste, firman el presente certificado.

Valencia, junio de 2021.

Fdo: Dr. D. Rafael Aleixandre Benavent

Fdo: Dr. D. Antonio Vidal Infer

**ALEIXANDR
E BENAVENT
RAFAEL -
20773955X** Firmado
digitalmente por
ALEIXANDRE
BENAVENT RAFAEL
- 20773955X
Fecha: 2021.06.25
15:32:29 +02'00'

**ANTONIO
MARTIN|
VIDAL|
INFER** Firmado
digitalmente por
ANTONIO
MARTIN|VIDAL|
INFER
Fecha:
2021.06.25
14:55:51 +02'00'

Agradecimientos

Es difícil escribir un apartado como este cuando una siente que tiene tanto que agradecer.

Cuando empecé esta tesis, no tenía ni idea de cómo iban a ser los siguientes años de mi vida. Ahora que me encuentro en el final del trayecto, siento que son tantas las vivencias, tantas las experiencias, que solo puedo sentir que he tenido mucha suerte.

Suerte por haber podido introducirme este mundo increíble que es el de la investigación, no siempre fácil, a veces duro y exigente, pero siempre gratificante si sabes tener paciencia y la visión a medio plazo.

Suerte por haber tenido la sensación a lo largo de estos años de estar poniendo mi granito de arena para engrandecer esta inmensa playa que es la Ciencia.

Suerte por la educación pública recibida, sin ella no habría podido llegar hasta aquí.

Suerte, muy especialmente, por haber estado rodeada del grupo de personas que han hecho esto posible. Porque al final de todas las cosas, de las horas invertidas, de las entrevistas realizadas, de los cálculos hechos y de los párrafos escritos, está la gente que, consciente o inconscientemente te ha ayudado en el camino.

Por ello, os dedico esta tesis a todos vosotros.

A Rafa y a Toni, mis directores de tesis.

A Rafa, por ser el mestre de todos, el Mestre de Mestres. La combinación perfecta de excelencia científica, humanidad y cercanía. Moltes gràcies Rafa, no n'hi ha un altre com tu.

A Toni, por darle la primera oportunidad a una trabajadora social que tenía mucho entusiasmo, pero que sabía más bien poco. Si no te hubiera conocido, si no hubieras confiado en mí, no estaría hoy aquí, literalmente.

A Lourdes, amiga i companya. Por estar a mi lado desde el primer día. Por toda la ayuda e inspiración, por todos los cafés, y todas las risas.

A Rut, por tu ejemplo y por tu apoyo siempre, por ser el espejo en el que me gustaría mirarme.

A Juan Carlos, que volviste pisando tan fuerte que casi no recuerdo cuando no estabas. Por tu confianza en mí desde el principio.

A Adolfo, que con tu capacidad de resolución haces que nunca se te escape nada. Por ayudarme todas y cada una de las veces que lo he necesitado.

A Víctor, compañero de profesión. Por las oportunidades que me has dado de seguir vinculada al mundo de lo social, nos queda mucho por hacer juntos.

A Maxi, por tu soporte y tu respuesta a mis dudas, siempre desinteresada y siempre a tiempo.

También, a la gente del Palau, los que vinieron y se quedaron un tiempo, y los que estaban ahí y ahí seguirán cuando yo me vaya. Con especial cariño a Rosa, Carmen, Conchi, Carmen. Cuatro mujeres que con su discreción y buen hacer mantienen el edificio a flote.

Al Departamento de Historia de la Ciencia y Documentación de la Facultad de Medicina, por confiar en mí y hacerme sentir siempre como una más.

Al CWTS de Leiden, por darme la oportunidad de pasar un tiempo con ellos y conocer cómo funciona un centro extranjero de su calidad y prestigio.

A mi familia quiero agradecerle todo.

Porque todo lo que soy, es gracias a ellos.

Especialmente a mis padres, mi hermano, mis abuelos y mis padrinos.

Sois mis pilares, sin vosotros no soy nada.

A Alex. Mi compañero, mi amigo, el amor de mi vida.

Tú hiciste esta tesis conmigo. No hay palabras suficientes ni precisas para describir lo mucho que me has ayudado, inspirado y soportado.

Contigo siempre me siento capaz de cualquier cosa.

“Si somos los suficientes, alrededor del mundo, no sólo enviaremos un fuerte mensaje en oposición a la privatización del conocimiento –la haremos una cosa del pasado. ¿Vas a unirtenos?”

Aaron Swartz. Manifiesto por la Guerrilla del Acceso Abierto.

ABSTRACT

Data sharing between researchers is a practice that has been gaining relevance in recent years. Although it is a practice with a long history in science, nowadays technological advances and the progressive awareness of open access have made it increasingly required by journals, publishers and funding agencies. Since researchers are the main ones called upon to carry out this practice, inquiring about their knowledge, beliefs and attitudes about data sharing is vital to understand this phenomenon and to outline its medium- to long-term projection.

The methodology used was qualitative and consisted of a semi-structured interview with open-ended questions to 23 Health Sciences researchers in the Spanish context. A purposive sampling was carried out for their selection following the snowball technique. The sample consisted of people with different years of experience and who carried out basic or applied research. The interviews were manually transcribed into documents in .doc format and analyzed with the qualitative research software Atlas.ti.

Among the main results of the study, it was found that basic researchers have more knowledge and more experience with practice, as well as less reluctance than applied researchers to carry it out. However, they also encountered more often the requirement of data sharing by journals. On beliefs, in general all interviewees recognize the importance of practice, however, fears related to the loss of competitive advantage in terms of curriculum (especially among juniors), and in terms of securing funding (especially among seniors) are evident. For this reason, attitudes towards the internship can be summarized as "yes, but with conditions", in the vast majority of cases.

The main conclusion of this work is that there is a need to address the perception of researchers about a practice such as data sharing from a qualitative perspective. It is a practice that implies modifying the habits of researchers and assuming a series of risks in a very competitive world such as the scientific one. To try to ensure some success, institutional policies, as well as those of public and private funding entities, should focus on understanding the subjective realities of researchers before making moves and launching guidelines.

RESUMEN

El intercambio de datos entre investigadores es una práctica que ha ido ganando relevancia en los últimos años. Aunque es una práctica con una larga historia en la ciencia, hoy en día los avances tecnológicos y la progresiva concienciación sobre el acceso abierto han hecho que sea cada vez más requerida por las revistas, los editores y las agencias de financiación. Dado que los investigadores son los principales llamados a llevar a cabo esta práctica, indagar en sus conocimientos, creencias y actitudes sobre la compartición de datos es vital para entender este fenómeno y perfilar su proyección a medio y largo plazo.

La metodología llevada a cabo ha sido cualitativa y consistió en una entrevista semi-estructurada con preguntas abiertas a 23 investigadores de Ciencias de la Salud en el contexto español. Para su selección se ha llevado a cabo un muestreo intencional siguiendo la técnica de la bola de nieve. Se ha buscado que la muestra estuviera conformada por personas con distintos años de experiencia y que llevaran a cabo investigación básica o aplicada. Las entrevistas fueron transcritas manualmente en documentos en formato .doc y analizadas con el software de investigación cualitativa Atlas.ti.

Entre los principales resultados del estudio se ha hallado que los investigadores de básica tienen más conocimientos y más experiencia con la práctica, así como menos reticencia que los investigadores aplicados para llevarla cabo. No obstante, también se han encontrado más a menudo con la exigencia de compartir datos por parte de las revistas. Sobre las creencias, en general todas las personas entrevistadas reconocen la importancia de la práctica, sin embargo, se evidencian miedos relacionados con la pérdida de la ventaja competitiva en cuanto a currículum (sobre todo entre los junior), y en cuanto a la consecución de financiación (sobre todo entre los senior). Por este motivo, las actitudes hacia la práctica se pueden resumir en “sí, pero con condiciones”, en la gran mayoría de los casos.

La principal conclusión de este trabajo es que existe la necesidad de abordar la percepción de los investigadores e investigadoras sobre una práctica como es el uso compartido de datos desde una perspectiva cualitativa. Se trata de una práctica que

implica modificar los hábitos de los investigadores y asumir una serie de riesgos en un mundo muy competitivo como es el científico. Para tratar de asegurar un cierto éxito, las políticas institucionales, así como de las entidades financiadoras públicas y privadas, deben centrarse en entender las realidades subjetivas de los investigadores antes de realizar movimientos y lanzar directrices.

ÍNDICE

Introducción	4
1. El movimiento de acceso abierto y los datos de investigación	4
1.1 Definición y contexto histórico del movimiento de acceso abierto.....	4
1.2 El movimiento de acceso abierto en Ciencias de la Salud	9
2. Contextualización histórica de los datos de investigación y de su uso compartido	12
2.1 Definición de los datos de investigación	15
2.1.1 Definición data sharing y principios FAIR	18
2.1.2 Características formales de los datos abiertos y los metadatos	21
2.1.3 Los planes de gestión de datos	24
2.1.4 Los metadatos	25
2.2 Data sharing en Ciencias de la Salud	27
3. Infraestructuras para compartir datos: los repositorios de datos abiertos en Ciencias de la Salud	33
○ Repositorios temáticos.....	35
○ Repositorios multidisciplinares	40
○ Repositorios institucionales	47
4. Las revistas científicas del ámbito de las Ciencias de la Salud y el uso compartido de datos	48
5. Modalidades que utilizan las revistas para compartir datos	51
5.1 Bajo petición expresa al autor/es del artículo	51
5.2 Material suplementario.....	52
5.3 El depósito en repositorios de datos.....	54
6. El personal investigador de Ciencias de la Salud y el uso compartido de datos.....	55
6.1 Perspectiva cuantitativa	55
6.2 Perspectiva cualitativa	58
Objetivos	61
• Objetivo general.....	61
• Objetivos específicos.....	61
Método	62
1. Origen de este estudio y contextualización de la metodología cualitativa	62
1.1 Origen del estudio y trabajos previos en el marco del proyecto DATASEA	62
1.1.1 Origen del estudio	62
1.1.2 Estudios complementarios en el marco del proyecto DATASEA.....	63

1.2	Contextualización de la metodología cualitativa	70
1.2.1	Definición	70
1.2.2	Investigación cualitativa y cuantitativa: diferencias y sinergias.....	71
1.2.3	Breve recorrido histórico de la metodología cualitativa y aproximación a su multidisciplinariedad.....	73
2.	Métodos y técnicas cualitativas utilizadas en este estudio	76
2.1	El universo y la muestra	76
2.1.1	El universo	76
2.1.2	La muestra	79
2.2	Recogida de información.....	83
2.2.1	La entrevista en profundidad	83
2.2.2	Las preguntas y respuestas	85
2.2	Dimensiones de análisis	85
2.3	Procedimiento.....	91
2.4	Análisis.....	92
	Resultados	94
1.	Conocimientos y experiencias.....	95
1.1.	Experiencias relacionadas con las revistas y las modalidades de éstas para compartir datos.....	95
1.2	Experiencias relacionadas con los proyectos de investigación	101
1.3	Formación y apoyo de la institución	102
2.	Creencias	104
2.1.	Sobre las ventajas.....	104
2.2.	Sobre las desventajas y los miedos	107
3.	Actitudes	114
3.1.	Disposición a compartir datos.....	115
3.2.	Cómo deberían de ser presentados los datos.....	119
	Discusión	123
1.	Conocimientos y experiencias.....	123
1.1	Experiencias.....	123
1.3	Revistas.....	127
1.4	Proyectos.....	128
1.5	Esfuerzo percibido.....	129
2.	Creencias	131
2.1	Sobre las ventajas.....	131
2.2	Sobre los inconvenientes y los miedos	134

3. Actitudes	145
3.1 Disposición a compartir datos.....	145
3.2 Cómo deberían compartirse los datos	152
Conclusions	155
Limitaciones y propuestas de futuro	158
Referencias bibliográficas	160
Anexos	181
1. El guion	181
2. Las redes.....	182

Introducción

1. El movimiento de acceso abierto y los datos de investigación

1.1 Definición y contexto histórico del movimiento de acceso abierto

En los últimos años, el movimiento de acceso abierto a la ciencia ha tenido un enorme impacto en todo el mundo, generando interés y atención en gobiernos, instituciones universitarias, políticas editoriales y científicos en general (Elliott y Resnik, 2019). Históricamente, el movimiento por el acceso abierto a la ciencia tiene sus orígenes en el deseo de un amplio sector de la ciencia por tener un acceso más libre y menos sujeto a las tradicionales restricciones de los contenidos científicos. Los antecedentes directos que encendieron la mecha para que este movimiento se iniciara fue el enfado de la comunidad científica, en los años 80, por el aumento de los precios de las revistas académicas por parte de las editoriales conocido como la “crisis de las publicaciones seriadas”. A esta subida de precios se le unió el malestar por el control total que tenían las editoriales sobre el copyright, y la obligación por parte de las mismas a bibliotecas y otras instituciones a realizar contratos por paquetes (Melero, 2005; Vargas Arbeláez, 2014). Sin embargo, el movimiento por el acceso abierto no hubiera sido posible sin un desencadenante que empieza a cobrar fuerza a finales del siglo XX y, sobre todo, a principios del XXI y que cambió totalmente la comunicación científica en formato escrito: el desarrollo de Internet y la digitalización de los contenidos científicos (Abadal, 2012).

Hasta entonces, la comunicación científica a través de las revistas había mantenido como vehículo preferente el formato papel. Esta comunicación tal y como la conocemos hoy, nace aproximadamente hace 350 años, cuando ven la luz las que se consideran las dos primeras revistas científicas de la historia: la *Philosophical Transactions of the Royal Society of London* y la *Journal des sçavans* (Banks, 2009). A partir de estas dos primeras y hasta la actualidad, miles de revistas nacen tratando de dar respuestas a las necesidades de la ciencia y de la sociedad, manteniendo, como se

ha dicho, el formato impreso como el principal para su difusión. Con el paso del tiempo, la infraestructura sociotécnica relacionada con la comunicación científica, que implica a diferentes agentes, fue mudando y adaptándose a los tiempos. Por ejemplo, aunque en un primer momento, las publicaciones periódicas se difundían por sociedades científicas, como la Royal Society, con el tiempo se le han añadido las universidades, las agencias gubernamentales y, por último, las editoriales privadas (Reed-Elsevier, Taylor y Francis, Wiley-Blackwell, etc...) (Abadal y Rius Alcaraz, 2006).

No obstante, a mediados de los años 80, varios hitos supusieron un cambio de paradigma que en la actualidad aún sigue definiéndose. Dos de esos hitos son fruto de las circunstancias que se mencionan en el primer párrafo (descontento con las políticas editoriales y aparición de internet), mientras que un tercero, muy relacionado con los anteriores, sería el nacimiento de la primera revista impresa que creó un formato digital, la denominada *New Horizons in Adult Education* editada por la Universidad de Siracusa en 1987 y la primera revista cien por cien digital, es decir, sin versión impresa previa, la *Post-modern Culture*, publicada en 1990 (Alonso Gamboa, 2017). Estas dos primeras revistas, junto con otras que surgen en esta primera etapa, son las que el autor Voutssas (Juan Voutssas, 2012) denomina “protorevistas” electrónicas, por las características comunes que tenían de formato solo en texto plano ASCII, sin edición, ni ningún elemento multimedia (imágenes, vídeos...), que se transmitieron antes de Internet como se conoce hoy, ya que se trataba de un intercambio de archivos electrónicos anexos a un mensaje por medio del FTP (File Transfer Protocol). Tras esta etapa preliminar, este autor menciona una segunda etapa intermedia, donde aparecen revistas a texto completo que son difundidas en soporte CD-ROM y CD, y una tercera, que tiene lugar a partir de mediados de los 90, que poco a poco va sustituyendo a los otros formatos y que llega hasta la actualidad: las revistas distribuidas mediante la “World Wide Web” utilizando el lenguaje HTML y los navegadores.

A partir de esa última etapa, las revistas electrónicas distribuidas en Internet han ido aumentando de manera gradual pero imparable. Según refieren Abadal y Rius (2006) en su artículo, en el año 2006 el prestigioso directorio Ulrichs de publicaciones seriadas recogía 19.488 revistas académicas activas en formato digital. A octubre del año 2019, este mismo directorio contiene más de 85.000 revistas activas que están en formato

online (Ulrichsweb, 2019). En cuanto a la comunidad científica, el reconocimiento a las publicaciones en línea es mayoritario, debido, fundamentalmente, a que sus estructuras son muy parecidas a la de las revistas impresas en aspectos como la revisión por pares, los consejos editoriales o las normas de publicación.

En este contexto de digitalización de la ciencia, las editoriales continuaron teniendo su cuota de poder, demostrado en las subidas de precios de las cuotas y las imposiciones a las bibliotecas, que causó el descontento antes mencionado. Sin embargo, Internet había ofrecido a los sectores descontentos una posible alternativa a la comunicación científica bajo las normas de las editoriales tradicionales. Por primera vez, se ponía en entredicho el modelo en el que las editoriales cobran a los científicos y demás usuarios por el acceso a la información científica (Laakso et al., 2011). Nace así lo que se conoce como el movimiento por el acceso abierto.

Una de las primeras definiciones del acceso abierto y también una de las más reconocidas, la ofrece el científico Peter Suber (2004) en el año 2004, quien es uno de sus precursores más reconocidos. La definición de Suber dice lo siguiente: *“Open access (OA) literature is digital, online, free of charge, and free of most copyright and licensing restrictions.”* Esta definición concuerda con las reivindicaciones de los años 80 sobre las dificultades en el acceso a las publicaciones científicas, con el proceso de digitalización de la ciencia y la aparición de Internet, y con el poder sobre el copyright que ostentaban las editoriales.

Dentro de los hitos que rodean al movimiento por el acceso abierto, es importante destacar la creación a principios de los 90 de la plataforma arXiv.org por el físico Paul Ginsparg para el depósito de trabajos de investigación en el ámbito de la física, matemáticas e informática, que también tuvo mucho que ver con la difusión de los primeros e-prints (entendido éste como el formato electrónico que cualquier documento de investigación) (McKiernan, 2000).

Otra experiencia pionera que también data de principios de los 90, antes del lanzamiento de la World Wide Web, es la de la creación de las revistas de acceso abierto *Surfaces* por Jean Claude Guedon y *Psycoloquy* por Stevan Harnard. La filosofía

de estas revistas era que los autores mantuvieran el copyright a la vez que ponían a disposición todos los contenidos de forma gratuita (Melero y Abad Garcia, 2008).

Sin embargo, fue a partir del siglo XXI cuando hubo una mayor cantidad de acontecimientos trascendentes. Entre ellos estarían, en el año 2000, los hitos que supusieron, por un lado, la decisión de la Biblioteca Nacional de Medicina de los Estados Unidos de poner en libre acceso la gran base de datos bibliográfica Medline y su motor de búsqueda PubMed; y, por otro, y la creación en el mismo año de BioMed Central, una editorial científica británica especializada en publicaciones de acceso abierto (se tratará este hito con mayor atención en el siguiente apartado 1.2 “El movimiento de acceso abierto en Ciencias de la Salud”) (Kurata, Morioka, Yokoi, y Matsubayashi, 2013; Suber, 2002). En el año 2001, de manera casi simultánea, surge una editorial de acceso abierto sin ánimo de lucro llamada PLOS (Public Library of Science), fundada por un grupo de importantes investigadores. Esta iniciativa surge cuando estos investigadores deciden enviarles una carta a otros investigadores en todo el mundo pidiéndoles que se negaran a publicar en revistas que no les concedieran el copyright de sus artículos si no al momento, al menos pasado un período de embargo no superior a seis meses (Melero y Abad Garcia, 2008).

Después de estas primeras experiencias, que empezaban a dejar entrever cómo se ponía en marcha el acceso abierto, tuvieron lugar las que se consideran las tres cristalizaciones más significativas de este movimiento, ya que le ofrecieron un contexto más formal y estructurado, que se conocen como las 3 “B”: la Declaración de Budapest (2002), la Declaración de Bethesda (2003) y la Declaración de Berlín (2003) (Da Costa y Leite, 2016). Estas tres “B” le dieron cuerpo al acceso abierto y sentaron precedente para su desarrollo, pero en distintos niveles. En la primera declaración se habla del acceso libre a los textos completos, a su distribución y su uso manteniendo el copyright y respaldando que sean los autores e instituciones quienes controlen este derecho. En la segunda, se suscribe lo dicho en la anterior añadiendo que el copyright no debe ser suponer un escollo para el acceso, por lo que estimula a los que poseen los copyright a que se comprometan a mantener un derecho de acceso irrevocable (Melero, 2005). En el caso de la tercera declaración, la Declaración de Berlín, es de gran importancia a nivel de política científica, ya que, además de ratificar lo dicho en

las otras dos declaraciones, supone que líderes políticos y científicos de distintas instituciones firmen una adhesión a los principios del acceso abierto que, aunque no es vinculante, sí que significa un compromiso y un apoyo a la misma (Max-Planck-Gesellschaft, 2003; Sociedad Max Planck, 2003).

En general, las tres declaraciones ratifican que la publicación en acceso abierto implica lo siguiente:

1. Que los propietarios del copyright confieren a los potenciales usuarios el derecho libre, irrevocable, universal y permanente de acceso y licencia para copiar, utilizar, distribuir, transmitir y presentar el trabajo públicamente y hacer y distribuir obras derivadas, en cualquier soporte digital para cualquier finalidad responsable, sujeto a la apropiada atribución de la autoría.
2. Que al menos se debe facilitar una versión completa de la obra y todos los materiales suplementarios, así como de los permisos en un formato electrónico estándar. Esta versión debe depositarse inmediatamente después de la publicación inicial en como mínimo un repositorio en línea respaldado por una institución académica, sociedad científica, agencia gubernamental, o cualquier institución debidamente acreditada con el fin de facilitar el acceso abierto, la distribución sin restricciones, la interoperabilidad y el archivado a largo plazo.

Sin embargo, es en la última declaración, la de Berlín, cuando se especifica más concretamente qué tipo de material debe ponerse en acceso abierto. Esto es relevante ya que, aunque en principio la reivindicación de acceso abierto estaba más enfocada a las publicaciones originales finales, como se puede apreciar en la primera declaración sobre acceso abierto del año 2001, la Declaración de Budapest (Open Society Institute, 2002), poco después se fue aceptando que el acceso abierto debe ir más allá y cubrir todas las fases de la investigación. De este modo, según la Declaración de Berlín, una contribución de acceso abierto debe implicar a todos los actores que contribuyen en el proceso científico, desde el personal investigador hasta las editoriales. Se considera que las contribuciones de acceso abierto deben incluir necesariamente tanto las publicaciones finales, como los datos y metadatos primarios, así como también todos los materiales, fuentes, representaciones digitales de materiales gráficos y pictóricos (Max-Planck-Gesellschaft, 2003).

Después de la definición en la Declaración de Berlín del año 2002, en la que queda estipulado que el acceso abierto comprende desde la publicación final hasta los datos primarios, numerosas instituciones y organizaciones refutan esta directriz adaptándola a sus realidades y particularidades. En el caso de la Unión Europea (UE), la definición ofrecida por la Comisión sobre acceso abierto, que afecta a todas las investigaciones financiadas con fondos de la Unión, dice lo siguiente (EuropeanCommission, 2019):

Open access (OA) can be defined as the practice of providing on-line access to scientific information that is free of charge to the user and that is re-usable.

1.2 El movimiento de acceso abierto en Ciencias de la Salud

Como ya se ha mencionado, entre los primeros hitos del acceso abierto en la ciencia fueron precisamente en el ámbito de las Ciencias de la Salud, con la puesta en libre acceso de la base de datos bibliográfica Medline y su motor de búsqueda PubMed, la creación de las editoriales BioMed Central y, en 2001, PLOS, que posteriormente empezaría a editar varias revistas relacionadas con la biomedicina, como PLOS Medicine, PLOS Genetics, o PLOS Clinical Trials.

Tras las declaraciones de Budapest, Bethesda y, sobre todo, Berlín, distintas instituciones en todo el mundo relacionadas con Ciencias de la Salud se han ido uniendo al movimiento de acceso abierto. Un ejemplo paradigmático, que marcó un antes y un después en la adhesión de instituciones de Ciencias de la Salud al acceso abierto a las publicaciones científicas, fue la política de acceso abierto del National Institute of Health (NIH) de EEUU, publicada en diciembre del 2007 y puesta en marcha en abril de 2008. Esta política, recogida en el *US Consolidated Appropriations Act, 2008*, sección 218 (Government Publishing Office, s.f.), estipula que todo el personal investigador financiado por el NIH debe poner a disposición en la Library of Medicine's PubMed Central una versión final de los manuscritos, ya revisados por pares y aceptados para su publicación. Deben estar públicamente disponibles no más tarde de 12 meses después de la fecha oficial de publicación. Además de ser un hito para el ámbito de Ciencias de la Salud, lo es también para el movimiento de acceso abierto en

general por dos motivos: 1) se trata de la primera institución financiada con dinero público en el mundo que obliga por ley estatal, en vez de por iniciativa propia de la institución, a que los manuscritos estén en acceso abierto, y 2) el NIH es el mayor financiador de investigación científica a nivel internacional por lo que tiene una amplia cobertura (Suber, 2008). A este mandato de obligatoriedad le precedía una política lanzada por el NIH en el 2005, según la cual se alentaba al personal investigador a que pusieran sus trabajos en acceso abierto. No obstante, fue poco el personal investigador que se animó a seguir esta recomendación, hecho que conllevó en el 2007 a la obligación por ley que se menciona anteriormente (English y Joseph, 2008).

Además del NIH, otra iniciativa relacionada con el acceso abierto a las publicaciones en Ciencias de la Salud, prácticamente contemporánea a la del NIH, es la del Canadian Institutes of Health Research (CIHR). Se trata de una política que también aboga por que los trabajos financiados por esta institución se publiquen en acceso abierto. La política del CIHR tiene en común con la del NIH el establecer un máximo de 12 meses de embargo para poner en acceso abierto.

A la par de estas dos iniciativas, y conforme el OA se iba estableciendo y aceptando, numerosos centros relevantes de investigación y otras instituciones, como las propias universidades, comenzaron a solicitar que los trabajos financiados por las mismas fueran depositados bien en repositorios o directamente publicarlos en OA. En el caso de EEUU, en el año 2009 más del 90% de la investigación en el área médica financiada con dinero público estaba cubierta por el mandato de OA (Chan, Arunachalam, y Kirsop, 2009).

En el contexto de la Unión Europea, una parte importante de la investigación en Ciencias de la Salud se encuentra bajo el paraguas de los proyectos Horizonte 2020. Estos proyectos, que abarcan el período de 2016 a 2020, siguen una política de acceso abierto similar a los del NIH y el CIHR en cuanto a las publicaciones científicas, ya que obligan a que todas las publicaciones derivadas de los proyectos financiados en este programa deben estar en acceso abierto (European Commission, 2019).

En una aproximación más reciente del año 2017, a nivel de país, se observan iniciativas como la Declaración de Sant Joan de Alacant en defensa del Acceso Abierto a las

publicaciones científicas. Esta declaración fue firmada por los editores españoles de revistas de Ciencias de la Salud, conscientes de los cambios sobre el OA y las cada vez más frecuentes peticiones, como la de los proyectos H2020. En la declaración se proponen una serie de recomendaciones y de peticiones, entre las que se encuentran el “Adherirse a los criterios emanados de la reunión de marzo de 2017 de la European Open Science Policy Platform” o “Instar a las agencias de investigación a nivel nacional a poner en marcha políticas científicas que requieran a sus investigadores que depositen sus publicaciones en repositorios institucionales” (Grupo de editores de revistas españolas sobre Ciencias de la Salud, 2018).

Debido a los impulsos que las instituciones le han dado al acceso abierto a las publicaciones desde hace algo más de una década, la disponibilidad en abierto de diferentes tipos de publicaciones científicas del ámbito biomédico ha ido en aumento. Un ejemplo de ello se da en el trabajo de Kurata et al., (2013), en el que estudian el aumento de publicaciones en acceso abierto en la base de datos PubMed entre los periodos de 2006 a 2010. Los resultados que obtienen muestran que, mientras en 2006 los documentos en acceso abierto eran un 25% aproximadamente del total, en el año 2010 habían aumentado a el doble, con un 50% de los documentos en acceso abierto. Por otro lado, si se atiende al repositorio de acceso abierto de la base de datos Medline, PubMed Central, desde su lanzamiento en el año 2000 ha ido progresivamente aumentado su contenido hasta llegar a los 6 millones de artículos a texto completo (PubMed Central, 2020).

A la vez que cada vez más instituciones promovían el acceso abierto y que esto se veía reflejado en la cantidad de publicaciones mediante esta modalidad, otros autores comenzaron a estudiar beneficios más concretos del acceso abierto, más allá de bien general de que la ciencia esté al alcance de todos. En el caso del estudio de Cui Huang et al., (2017) sobre el efecto del acceso abierto en el Factor de Impacto (FI) de las revistas médicas entre los años 2011 y 2015, se encuentra que las publicaciones en OA tienen un FI más elevado en el área médica. Estos autores se basan en el principio de que el OA facilita la distribución de publicaciones académicas e incrementa las posibilidades de exposición de un trabajo ante un mayor número de personas, y, por ello, sus posibilidades de ser citado. Además, un aspecto interesante al que hacen

referencia son las posibilidades que el OA ofrece a los países con menos desarrollo científico y tecnológico para dar visibilidad a sus trabajos. De hecho, en su estudio, destacan que en países como Colombia, Serbia o Uganda el optar por OA había mejorado considerablemente el impacto de sus trabajos, lo que también se refuta en el estudio de Smith et al., (2017), donde defienden que el OA es un recurso útil en países con menos posibilidades.

A modo de síntesis, se podría decir que, desde que comienza el movimiento por el OA, en el ámbito de las Ciencias de la Salud son numerosas tanto las instituciones que deciden unirse, como las iniciativas que surgen por parte del personal investigador, de las revistas y de las editoriales para promoverlo, debido a las ventajas que se ha ido descubriendo que posee esta práctica.

2. Contextualización histórica de los datos de investigación y de su uso compartido

Según Stanley y Stanley (1988), compartir datos es, históricamente, una práctica común sobre todo en contextos donde ya existía una relación de colaboración, bien entre dos investigadores o entre grupos. En su trabajo publicado en el año 1988, estos autores ahondaban en temas relacionados con el uso compartido de datos, como las ventajas e inconvenientes que suscitaban entre el personal investigador, o la demanda que empezaban a hacer algunas instituciones, como la National Academy of Science, de la publicación de estos datos. De hecho, incluso hay referencias de principios del siglo XX, como la de Galton (citado en Robinson-García, Mongeon, Jeng, y Costas, 2017), que hace alusión a la importancia de compartir los datos brutos, bajo las restricciones razonables, para aumentar la transparencia de las investigaciones. En otra prueba más reciente, aunque también de varias décadas atrás, el uso compartido de datos se define en los siguientes términos:

Sharing also has many meanings. Generally, sharing is defined as giving or receiving a part of something, or enjoying or assuming something in common. As with any other sharing that persons do, scientists may share because they: wish to do so and explicitly

set out to do so, are forced by law to do so, are shamed or coerced into it, do so out of enlightened self interest, or explicitly set up the sharing agreement with a valuable quid pro quo (J. E. Sieber, 1988).

Según esta misma autora, pero en un trabajo del año 2006 (Knoppers y Joly, 2018), la discusión y la exposición de dudas sobre el uso compartidos de datos tienen lugar desde hace unos 50 años, por lo que manifiesta que no es ninguna novedad. En su opinión, el intercambio de datos a nivel internacional ha existido desde hace al menos 150 años y ha ido evolucionando a la vez que incluía el intercambio de multitud de datos científicos. De esta manera, las diferentes disciplinas han ido tomando nota de los riesgos y de los beneficios que tiene el uso compartido de datos para sus investigaciones. Por ejemplo, en el caso de las disciplinas donde se invierte una gran cantidad de dinero como la astronomía o la exploración del espacio, esta autora defiende que han avanzado muy rápido precisamente debido al uso compartido de datos entre los científicos. No obstante, otras como las relacionadas con las Ciencias Sociales y las Ciencias de la Salud también han sabido sacar provecho, a pesar de que deben lidiar con aspectos tan delicados como la gestión y protección de los datos personales (el caso concreto de las disciplinas del ámbito de las Ciencias de la Salud se trata con más detenimiento en el apartado 5.2). Entre las disciplinas con mayor trayectoria en el uso compartido de datos, Sieber destaca la meteorología y la economía, ambas fruto de la necesidad de unificar información a gran escala que permita extraer resultados significativos. En una línea similar, autores como Milham (Nature Communications, 2018) también refieren que el uso compartido de datos tiene una larga trayectoria en la ciencia en campos como la meteorología y la economía, que basan sus investigaciones desde hace mucho tiempo en datos compartidos. Además, también cabe destacar iniciativas como la del National Center for Biotechnology Information, que ya en el año 1990 inaugura un repositorio de datos (el tema de los repositorios se tratará con mayor detalle en el apartado 3) llamado GenBank que, en el año 2019, manejaba alrededor de 30 terabytes de datos (Popkin, 2019).

Sin embargo, el gran cambio en la concepción de la práctica de compartir datos tal y como la describían Sieber y Stanley y Stanley en sus inicios, viene dado por la irrupción

del movimiento por el OA. Con este movimiento, la práctica de compartir datos adquiere una nueva dimensión, al empezar a considerarse en las declaraciones y demás manifiestos que se fueron sucediendo como una parte importante del proceso investigador que debe estar en libre acceso. Esta concepción se puede observar en declaraciones como la de Berlín (Sociedad Max Planck, 2003), que manifiestan que los datos primarios y los metadatos forman parte del material que debe ponerse en abierto para el resto de la comunidad científica (Gewin, 2016), y en las políticas de otras instituciones importantes que lo refrendan, como la Unión Europea, que en su definición de OA (European Commission, 2019) manifiesta que:

In the context of RyD, open access to 'scientific information' refers to two main categories:

- *Peer-reviewed scientific publications (primarily research articles published in academic journals)*
- *Scientific research data: data underlying publications and/or other data (such as curated but unpublished datasets or raw data)*

En la misma línea, en un documento de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) del año 2004 firmado por 34 países, se relatan una serie de afirmaciones sobre las ventajas del acceso abierto a los datos de investigación y un compromiso de tenerlas en cuenta para fomentar lo máximo posible la transparencia, la interoperabilidad y la eficiencia (entre otros) en la ciencia. Por ejemplo, el primer reconocimiento de una lista de 8, sería el siguiente: “Recognising that an optimum international exchange of data, information and knowledge contributes decisively to the advancement of scientific research and innovation.” (OECD, 2004).

Además de la Comisión Europea y de la OCDE, otras instituciones de muy alto nivel, como la US Academy of Sciences, Engineering and Medicine, se han unido a la petición de una ciencia más abierta, incluyendo expresamente a los datos en esta ecuación. En general, son muchas las agencias financiadas por gobiernos, por ejemplo, de EEUU, Australia o de países europeos, que requieren a los autores que realicen planes de gestión de datos, y, siempre que sea posible, que los pongan a disposición del público en general. En el caso de las revistas y de las editoriales, también se ha observado un

incremento de interés en el uso compartido de datos (que se explicará más detalladamente en el apartado 4).

Además de la expansión del movimiento de OA, otro factor fue clave en el cambio de concepción y el impulso que tuvo la práctica de compartir datos a finales de los 80-90 y principios de los 2000: el uso de internet. Como sucedió con el propio OA, la generalización del uso de Internet en la ciencia y el asentamiento de la era digital, tuvo una fuerte repercusión a la hora de promover el uso compartido de datos. El hecho de que la información pueda ser compartida al instante con cualquiera que tenga conexión a Internet permitió dar un salto de un data sharing (en adelante, se utilizarán en todo el documento los conceptos “data sharing” y “uso compartido de datos” indistintamente) en el que los datos eran requeridos y compartidos entre investigadores que se conocen, a un concepto mucho más amplio y global (Popkin, 2019). Por ello, cuando se habla del uso compartido de datos en el contexto de finales de la segunda década del 2000, se refiere a compartir a través de los medios digitales, ya que son éstos los que proveen a la ciencia de mecanismos y herramientas para manejar grandes cantidades de datos en un mundo donde la investigación es cada vez más global (Sayogo y Pardo, 2013).

2.1 Definición de los datos de investigación

Actualmente, existen varias definiciones de los datos de investigación ofrecidas por diferentes organizaciones e instituciones, que varían ligeramente según necesidades y particularidades. A continuación, se muestran cuatro definiciones correspondientes a distintas instituciones, con las que se pueden observar tanto notables similitudes como algunos matices que las diferencian. Además, se ponen una serie de ejemplos concretos relativos al área de Ciencias de la Salud.

La primera definición, ofrecida por la Office of Management and Budget (Office Management and Budget, 1999) de EEUU, se trata de punto de vista gubernamental con intención de uniformizar los requerimientos administrativos para acuerdos y

subvenciones con instituciones universitarias, hospitales y organizaciones sin ánimo de lucro, dice lo siguiente:

Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples). Research data also do not include:

a) Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and

b) Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.

Published is defined as either when:

a) Research findings are published in a peer-reviewed scientific or technical journal; or

b) A Federal agency publicly and officially cites the research findings in support of an agency action that has the force and effect of law.

Otra, dada por la Deutsche Forschungsgemeinschaft (Sociedad Alemana de Investigación) en su *Guidelines on the Handling of Research Data* (Deutsche Forschungsgemeinschaft, 2019), más enfocada a la gestión de la información por parte de un organismo dedicado la investigación, se refiere a los datos de investigación como:

Research data is an essential foundation for scientific work. The diversity of this data reflects the wide range of different scientific disciplines, research interests and research methods. Research data might include measurement data, laboratory values, audiovisual information, texts, survey data, objects from collections, or samples that were created, developed or evaluated during scientific work. Methodical forms of testing such as questionnaires, software and simulations may also produce important results for scientific research and should therefore also be categorised as research data.

La tercera, facilitada desde el ámbito de la Unión Europea con fines de definir qué tipo de información forma parte de los proyectos que financia (European Commission, 2019), define los datos de investigación de la siguiente forma:

In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form.

El cuarto concepto de datos de investigación, que pone a disposición de sus usuarios la biblioteca de la University of Leeds (University of Leeds, s.f.), además de definir qué son los datos de investigación, también facilita un listado más concreto de los tipos de datos que puede haber, como se puede observar en la siguiente definición:

Research data is any information that has been collected, observed, generated or created to validate original research findings.

Although usually digital, research data also includes non-digital formats such as laboratory notebooks and diaries.

Types of research data:

- *Documents, spreadsheets*
- *Laboratory notebooks, field notebooks, diaries*
- *Questionnaires, transcripts, codebooks*
- *Audiotapes, videotapes*
- *Photographs, films*
- *Test responses*
- *Slides, artefacts, specimens, samples*
- *Collections of digital outputs*
- *Data files*
- *Database contents (video, audio, text, images)*
- *Models, algorithms, scripts*
- *Contents of an application (input, output, logfiles for analysis software, simulation software, schemas)*
- *Methodologies and workflows*

- *Standard operating procedures and protocols*

En referencia a los datos en áreas relacionadas con las Ciencias de la Salud y la biomedicina, los datos sin procesar coinciden en líneas generales con las descripciones que se han hecho en este apartado. Siguiendo a Arning (2015), se pueden encontrar algunos ejemplos más concretos de los datos sin procesar en estas áreas, como son los datos estadísticos, por ejemplo los procedentes de estudios diagnósticos anonimizados; clasificaciones y códigos de enfermedades; imágenes, como las procedentes de la resonancia magnética; datos de sensores de bioseñales o mediciones de parámetros vitales, como los procedentes de electrocardiogramas o electroencefalografías; datos de biomateriales de pruebas de laboratorio, como muestras de sangre o datos genómicos; información administrativa de pacientes, como la que se almacena en los sistema de recogida de información de hospitales; o datos procedentes de recursos audiovisuales, modelos y visualizaciones.

2.1.1 Definición data sharing y principios FAIR

Según Torres-Salinas et al., (2012), el data sharing puede definirse como el acceso abierto a los datos primarios de las investigaciones para que puedan ser reutilizados por el resto del personal investigador para diversos fines. Dentro del movimiento por el acceso abierto, autores como Klump et al., (2006), defienden que el uso compartido de datos debe tratarse de manera análoga a como se hace con las publicaciones tradicionales (artículos originales, revisiones, etc.), es decir, que los datos deben poder descargarse y compartirse libremente mientras sean correctamente citados. Sin embargo, para que los datos puedan ser compartidos, numerosas instituciones y autores llegaron a la conclusión de que no era suficiente con la sola voluntad de querer hacerlo, sino que era necesario seguir una estructura y sistematización de los conjuntos de datos. Para ello, se establecen cuatro grandes principios que los datos y los conjuntos de datos deben seguir para que la práctica de compartir datos pueda ser

realmente útil. Estos cuatro principios se denominan los “Principios FAIR”, (Nature Genetics, 2016).

Las siglas FAIR (*Findable, Accessible, Interoperable, Reusable*), son pues las que rigen cómo deben de ser los datos una vez se decide que sean compartidos. Si alguno de estos principios no se cumple correctamente, el hecho de compartir los datos podría perder su potencial. El que las siglas estén ordenadas de esta forma concreta responde a una secuencia lógica. Lo primero, los datos deben ser fáciles de encontrar (*findable*); lo segundo, que además de encontrarlos, puedas obtenerlos (*accessible*); lo tercero, que una vez los encuentras y los obtienes, sea posible comprenderlos y trabajar con ellos (*interoperable*); y cuarto y último, que sean reutilizables (*reusable*), para que una vez se encuentren, se obtengan, se comprendan y se pueda trabajar con ellos, también sea posible reutilizarlos para otros fines estando clara su procedencia y las condiciones de reuso (Leung y Dumontier, 2019)

En el artículo de Wilkinson et al., (2016) definen las características de cada principio de la siguiente forma:

Para ser “Findable”:

- Los datos y metadatos deben tener asignado un doi
- Los datos tienen que estar descritos con suficientes metadatos
- Los metadatos claramente tienen que poder identificar a los datos que describen
- Los datos y metadatos deben estar registrados o indexados en un recurso que sea fácil de localizar

Para ser “Accessible”:

- Los datos y metadatos pueden ser recuperables con su identificador, utilizando un protocolo estandarizado de comunicación
 - El protocolo debe ser abierto, gratis y universalmente implementable
 - El protocolo debe permitir, cuando sea necesario, un proceso de autenticación y autorización

- Los metadatos deben ser accesibles, incluso cuando los datos ya no están a disposición

Para ser “Interoperable”:

- Los datos y metadatos deben utilizar un lenguaje formal, accesible, compartible y ampliamente aplicable que sea útil para la representación del conocimiento,
- Los datos y metadatos deben utilizar vocabularios que sigan los principios FAIR
- Los datos y metadatos pueden incluir referencias a otros datos o metadatos

Para ser “Reusable”:

- Los datos y metadatos deben ser descritos de manera rica, con una pluralidad de atributos precisos y relevantes
 - Los datos y metadatos deben publicarse con una licencia de sobre su uso y reutilización clara y accesible
 - Los datos y metadatos tienen que estar asociados con una información sobre su procedencia detallada
 - Los datos y metadatos deben seguir los estándares que utiliza la comunidad del dominio concreto que se esté usando

Es importante recalcar que los principios FAIR, pensados para establecer un marco para el uso compartido de los datos en abierto, no implican que absolutamente todos los datos deban de tener el mismo nivel de apertura ni que todos los datos de todas las disciplinas sigan las mismas reglas. Se trata más bien de un cambio de paradigma, que persigue el objetivo de que los datos estén abiertos por defecto en vez de cerrados, como venía sucediendo hasta. Esto se ve muy claramente con el lema de los principios FAIR *“tan abiertos como sea posible, tan cerrados como sea necesario”* que se menciona en el documento *“Guidelines on FAIR Data Management in Horizon 2020”* de la Comisión Europea (European Commission, 2016a).

2.1.2 Características formales de los datos abiertos y los metadatos

Según la Open Knowledge Foundation (Fountadion, s.f.), los datos abiertos son *“data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike.”*. Para que los datos efectivamente puedan ser libremente compartidos y reutilizados y los principios FAIR cumplidos, es necesario que tanto datos como metadatos cumplan una serie de características formales. El primer punto sobre los formatos de los datos es que se debe de asegurar que vayan en la línea de ser lo más fácilmente reutilizables posible. Buscando un ideal, los datos compartidos más apropiados serían aquellos que tienen un formato estructurado, libre, abierto y sin costes. En esta línea, Peset et al., (2017) mencionan que el hecho de que los datos estén disponibles, no significa automáticamente que puedan ser reutilizados, para ello, se deben cumplir al menos dos condiciones: que tengan un formato interoperable y que tengan una descripción asociada, es decir, metadatos. Sobre estas dos cuestiones, formato de los datos y descripción de los metadatos, se hablará en este apartado.

En primer lugar, se hará un breve repaso de los conceptos “libre”, “abierto” y “gratuito”, que a veces se usan indistintamente, pero que no significan lo mismo. Cuando se dice que el formato más adecuado es aquel que es libre, abierto y sin costes, esto es porque es la manera en la que puedes acceder a ellos (abiertos), puedes obtenerlos y modificarlos (libres) y puedes hacer todo ello sin costes (gratuitos). Cuando se habla del movimiento o filosofía de Datos Abiertos, nos estamos refiriendo a datos que persiguen estar en este tipo de formatos.

En general, la idea principal es que los datos deben tener formatos operables para que otros programas puedan abrirlos sin limitaciones y que el usuario pueda gestionarlos de manera gratuita. Un ejemplo de formatos de este tipo sería .csv o .xls, teniendo su opuesto, por ejemplo, en el formato .pdf (Generalitat de Catalunya, s.f.). En la misma dirección se manifiesta Michener (2015), cuando dice que será la mejor elección los formatos que no tengan propietario y que sean conocidos y aceptados entre la comunidad científica, y nuevamente ponen de ejemplo el formato .csv, situándolo en idoneidad por encima de Excel. No obstante, se debe hacer un inciso cuando hablamos

de Excel, ya que .xls o .xlsx es un formato muy utilizado en la comunidad científica a la hora de compartir datos y, aunque no sea libre, sí que permite la posibilidad de exportar datos en otros formatos (Aleixandre-Benavent, Lucas-Domínguez, Sixto-Costoya, y Vidal-Infer, 2018; Aleixandre-Benavent et al., 2014). La razón por la que este autor sitúa .csv por encima de .xlsx es que éste último utiliza formatos propietarios (protegido por una patente o derechos de autor, en este caso, Microsoft) y no libres (sin dueño).

Un sistema muy interesante para entender el grado de interoperabilidad de los datos es el esquema de desarrollo de 5 estrellas de Tim Berners-Lee (considerado padre de la World Wide Web) para clasificar los datos abiertos (Berners-Lee, 2009), que consta de 5 escalones donde el primer escalón tiene 1 estrella (lo mínimo) y el último 5 (lo máximo). En este esquema, se puede observar que en el primer escalón (1 estrella) sería hacer publicar algún tipo de material en la web con licencia abierta, pero que no puede ser modificado, como un imagen escaneada; el segundo (2 estrellas) es ponerlo disponible, pero con datos estructurados, como podría ser un archivo en el formato propietario .xlsx; el tercero (3 estrellas), es que esté disponible en un formato no propietario, por ejemplo .csv en vez de .xlsx; el cuarto (4 estrellas) es que se utilicen URI (uniform resource identifier) para identificar recursos de una determinada red; y quinto y último, (5 estrellas), que los datos estén ligados unos con otros para proveer de un contexto.

Otra visión muy interesante es la que se ofrece a través de Portal Europeo de Datos (<https://www.europeandataportal.eu/es/>), recurso financiado por la UE para promover la accesibilidad de los datos abiertos y el valor de los mismos, sobre el formato de los datos y el cómo escoger el formato correcto. Desde su punto de vista, el formato más utilizable para otros investigadores que no sean los creadores originales, es el formato con el que el conjunto de datos se generó y gestionó por primera vez. La justificación es que ese formato original ofrecerá un contexto mucho más fácil de entender y que, además, se sabe que funciona porque ya ha sido utilizado. Aunque entienden que se puede dar el caso de que ese formato no sea el más libre y abierto, refieren que siempre existe la posibilidad de compartir los datos en dos formatos, el original y uno libre, lo que serviría también como práctica para la

preservación de los datos a largo plazo en el caso de que el formato original fuera propietario y se quedara obsoleto (Commission, s.f.).

Por otra parte, otras dos recomendaciones que hacen desde esta página con respecto a los datos están relacionadas con la estructura y con la forma de entrega. En cuanto a la estructura, recomiendan tener en cuenta que, dependiendo del formato de los datos, así debe ser la estructura, ya que no todos los datos se pueden representar de la misma forma. Aunque la estructura más frecuentemente utilizada es la tabular, los datos también se pueden representar de manera jerárquica (relaciones entre puntos de datos de manera vertical) o en red (relaciones entre cualquier combinación y cualquier dirección). Sobre la forma de entrega de los datos, las indicaciones que ofrecen son claras: si los datos abiertos se ofrecen en formato descargable, deben ser fácilmente descargables, es decir, pensar en detalles como comprimir el tamaño del dataset, para que no pese demasiado y no lleve mucho tiempo descargarlo, o utilizar una terminología que sea fácilmente comprensible. A continuación, en la Tabla 1 se facilita una tabla donde se nombran algunos ejemplos de opciones de formatos libres utilizados actualmente dependiendo del tipo de archivo y a los que se puede recurrir para asegurar que los datos compartidos están 100x100 accesibles para cualquier persona.

TIPO DE ARCHIVO	FORMATOS LIBRES
IMAGEN	Joint Photographic Experts Group (.jpeg) Portable Network Graphics (.png)
TEXTO	OpenDocument Text (.odf)
VIDEO	XviD (.avi) Ogg Theora (.ogv)
HOJA DE CÁLCULO	Código Seguro de Verificación (.csv) OpenDocument Spreadsheet (.ods)
AUDIO	Vorbis (.ogg) Opus (.opus)

Tabla 1. Resumen informativo de algunos de los formatos libres existentes

2.1.3 Los planes de gestión de datos

Para finalizar este apartado sobre los datos, es interesante mencionar los llamados “Planes de gestión de datos” (“Data Management Plan”, en inglés), por su importancia en la organización de los datos en proyectos y otros trabajos como las tesis doctorales. Siguiendo la definición que da el grupo Datasea (Datasea, s.f.) para el plan de gestión de datos, éste es básicamente un documento sobre cómo se van a gestionar los datos en una investigación. Para elaborarlo, se debe hacer un ejercicio de reflexión antes de comenzar cualquier trabajo (una investigación sobre algún tema concreto, una tesis doctoral, la redacción de un proyecto, etc.) de qué datos se van a generar y para qué, ya que el plan de gestión de datos se elaborará en función del propósito final de cada investigación. En el caso de la Comisión Europea (European Commission, s.f.), se considera que los planes de gestión de datos son un elemento clave para una buena gestión de los datos, ya que describen cómo es ciclo de vida de los datos recolectados y su posterior procesado. En el ejemplo concreto de los proyectos H2020 que opten por la participación en el Open Research Data Pilot, se estipula que para los datos cumplan los principios FAIR, los planes de gestión de datos deben incluir la siguiente información:

- ✓ Cómo será el manejo de los datos durante y después del proyecto
- ✓ Qué datos se recopilarán, procesarán y/o generarán
- ✓ Qué metodología y estándares se aplicarán
- ✓ De qué modo los datos serán compartidos en acceso abierto
- ✓ Cómo será el proceso de curación y preservación de los datos

En el trabajo “Ten Simple Rules for Creating a Good Data Management Plan” Michener (2015) detalla diez reglas para elaborar un plan de gestión de datos completo. Estas directrices tienen en cuenta aspectos como la identificación de los datos que serán recolectados, relacionado con el tipo de datos, las fuentes, el formato y el volumen; cómo va a ser la organización de los datos, por ejemplo, qué programas se van a utilizar para su tratamiento; de qué manera serán descritos los metadatos para que

sean comprensibles; cómo se asegurará la calidad y preservación; qué políticas se llevarán a cabo, por ejemplo, relacionadas con las licencias de uso y apertura; y de qué manera se planteará un presupuesto realista de los costes (humanos y materiales) que supondrá la gestión de los datos.

A nivel de universidades, también hay diferentes iniciativas que abogan por facilitar tanto a estudiantes, como a profesores e investigadores explicaciones detalladas y plantillas para elaborar planes de gestión de datos. Por ejemplo, el servicio de biblioteca de la University of Bath ofrece un recurso web focalizado en los datos de investigación, en el que informan de aspectos relacionados con los planes de gestión de datos que van desde el nivel teórico de qué son los planes de gestión de datos y qué ventajas pueden tener para la investigación, hasta una explicación de qué requerimientos tienen entidades financiadoras como el NIH o la UE. Además, ofrecen una serie de plantillas para facilitar la realización de los planes (University of Bath, s.f.).

2.1.4 Los metadatos

Los metadatos que acompañan a los datos son uno de los soportes imprescindibles cuando se habla del uso compartido de datos. Básicamente, se trata de información que acompaña a los datos para otorgarles un contexto entendible (Torres-Salinas et al., 2012). Los metadatos tienen que ser lo suficientemente completos como para que la persona que recibe los datos tenga toda aquella información necesaria para comprender qué contienen, cómo están organizados y cuál es la procedencia del material compartido (Mayernik, 2019). Sin esta explicación, compartir datos puede pasar de ser una práctica muy útil, a una pérdida de tiempo tanto para la persona que comparte como para la que recibe.

En el libro *Understanding metadata. What is Metadata, and what is for?* Riley (2017) hace una clasificación muy útil de los distintos tipos de metadatos y sus propiedades. Aunque ella lo enfoca más en datos relacionados con museos y patrimonio cultural, su clasificación es también válida cuando se habla de datos de investigación. En primer lugar, ella destaca los denominados metadatos “descriptivos”, que son aquellos que se

refieren a la información que sirve para entender el contenido del conjunto de datos. Por otra parte, estarían otros tipos de metadatos más centrados en la forma y en la estructura, como los metadatos “técnicos” o los metadatos sobre los derechos de autor. Para una comprensión más visual, a continuación, se presenta en la Tabla 2 los tipos de metadatos, ejemplos sobre su contenido y sus usos principales, inspirada en la descripción de Riley, pero adaptada a los datos de investigación.

TIPO DE METADATOS	EJEMPLOS DE SU CONTENIDO	USOS PRINCIPALES
Metadatos descriptivos	Título Autores Fecha de publicación del estudio Resumen de la temática del estudio Protocolos para obtención de los datos	Dar un contexto al contenido en sí de los datos
Metadatos técnicos	Tipo de archivo Tipo de formato Tamaño del archivo Fecha de creación de los datos (o de las versiones)	Detallar o los tipos de archivos en los que los datos se encuentran disponibles, cuanto pesan e informar sobre la antigüedad de su creación
Metadatos sobre la preservación	El doi otorgado en caso de haberlo Indicación del repositorio de datos donde está depositado	Informar sobre qué modos de preservación existen para el conjunto de datos
Metadatos sobre los derechos	Derechos de copyright Licencias Creative Commons	Indicar qué derechos se han estipulado para el conjunto de datos

Tabla 2. Resumen informativo de los tipos de metadatos, con ejemplos de su contenido y sus usos principales

A continuación, se pone un ejemplo de qué metadatos se necesitan para un dataset procedente de un estudio realizado en un laboratorio donde se hacen ensayos preclínicos con animales. En este caso, los metadatos descriptivos deberían incluir, como mínimo, qué protocolos se llevaron a cabo para la obtención de esos datos, es decir, la lista de instrucciones que se siguieron, así como el tipo de animales que se

utilizaron, cuáles eran las condiciones del laboratorio, en definitiva, todo lo referente a cuál fue el método seguido y en qué consistía el estudio realizado. En cuanto a los metadatos técnicos, se refiera a qué tipo de archivo (por ejemplo, texto, imagen, hojas de cálculo) y en qué formato se encuentran esos datos, en este caso, por ejemplo, podrían ser archivos en formato .xlsx y/o .csv creados en “X” fecha. Además, deberían indicar qué derechos se tienen para la reutilización de esos datos y, por supuesto, el doi o el enlace que corresponda para poder consultarlos, compartirlos o citarlos.

Finalmente, hay que mencionar que, en la actualidad, existen numerosos estándares de metadatos elaborados por organizaciones e instituciones para adaptarse a la necesidad de describir los datos de manera sistematizada y efectiva. Algunos ejemplos de iniciativas relacionadas con estándares de metadatos, explicados muy bien por los autores Kim y Burns (2016a) en su trabajo centrado en el campo de la biología, son la Data Documentation Initiative (DDI), el Darwin Core, el Ecological Metadata Language (EML) o el Content Standard for Digital Geospatial Metadata (CSGDM).

2.2 Data sharing en Ciencias de la Salud

A nivel de políticas y de mandatos, el uso compartido de datos en Ciencias de la Salud está todavía en construcción y varía considerablemente entre países e instituciones.

Atendiendo a recomendaciones y posicionamientos de organizaciones de alcance internacional, destaca el documento lanzado por la OECD titulado *Recommendation on Health Data Governance*. Este documento, publicado por primera vez en el 2016, recoge una serie de puntos que giran en torno a la necesidad de los gobiernos de eliminar barreras y facilitar la interoperabilidad de las infraestructuras de los gobiernos en cuanto a los datos del campo de la salud (OECD, 2019).

Si se toma como referencia el posicionamiento de instituciones de prestigio, un ejemplo paradigmático podría ser el caso del National Institute of Health (NIH), por ser la mayor institución financiadora de investigación en Ciencias de la Salud a nivel mundial. En este caso, aunque todavía no tiene un mandato de que las investigaciones financiadas con sus fondos deban poner los datos en abierto de manera obligatoria,

como sí sucede con las publicaciones finales, ha trazado un camino en este sentido (Suber, 2008). Esto se puede comprobar cuando, en el año 2015, la institución lanzó un plan titulado *“National Institutes of Health Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research”*, en el que se sientan las bases para que los datos de investigación sigan el mismo camino que las publicaciones finales. En este documento, se pone en valor tanto los datos en sí como su potencial si son compartidos relacionados, por ejemplo, con el aumento de la transparencia, la replicabilidad y el ahorro de costes humanos y materiales (National Institutes of Health, 2015). En la misma línea a este plan del año 2015, un anuncio en la web del NIH en el 2018 publicaba una convocatoria en la que pedían a los stakeholders propuestas y feedback para la futura publicación de una nueva política de datos de investigación, que venía acompañada de una explicación cronológica de los pasos que ha dado el NIH con respecto a la publicación de los datos abiertos desde el plan del 2015 hasta el año 2018, en lo que de momento han denominado NIH Data Commons Pilot Phase (National Institutes of Health, 2018). Esta información de los avances del NIH deja entrever que una política de acceso abierto a los datos por parte de esta entidad puede estar cerca de ver la luz.

Otro caso a nivel de instituciones de investigación en Ciencias de la Salud es el del Canadian Institutes of Health Research, que ha añadido la publicación de los datos como parte de su política. En este caso, la institución reconoce el avance y el aumento de calidad para la ciencia que supone el uso compartido de datos y decidió implantar un sistema de depósito y publicación de los datos en recursos que sean abiertos y accesibles. Por ello, requiere que, después de la publicación de los resultados, los datos generados sean depositados en bases de datos, para lo que facilita un listado de algunas de ellas y ejemplos de material depositado. Además, refiere que los beneficiarios de la financiación deben conservar los datos por un mínimo de 5 años, tanto si el dato se publica como si no (Canadian Institutes of Health Research, 2015).

En el contexto de las investigaciones financiadas con fondos de la UE, y teniendo en cuenta los proyectos H2020 como los que mayor fondo poseen, no hay a día de hoy una obligación expresa de compartir datos. Sin embargo, se promueve la publicación de los mismos y existen mecanismos y directrices para ello, principalmente mediante

el Open Research Data Pilot¹, una medida piloto y, de momento, flexible. Esta medida indica que los proyectos financiados después del 2017 tienen incluida por defecto la indicación de compartir datos pero, en todo caso, el personal investigador puede optar por no hacerlo y no se les penalizará por ello (European Commission, 2016b).

Sin embargo, a pesar del marco de la UE y de sus políticos para compartir datos, las visiones y los mandatos pueden variar entre los Estados miembros. Por ejemplo, en el caso de Dinamarca, la Faculty of Health and Medical Sciences de la Universidad de Copenhague, basada en el mandato sobre datos en el Danish Code of Conduct for Research Integrity², lanzó en el año 2016 un documento en el que se especifica que el personal investigador debe diseñar un plan de gestión de datos en el que se explique qué tratamiento se le dará a los datos recolectados en sus investigaciones (Schmitt et al., 2016a). Por otro lado, países como España estipulan para sus proyectos de investigación financiados una política más orientada a las recomendaciones, tal y como indican en el Plan Estatal de Investigación Científica y Tecnológica y de Innovación 2017-2020 (Ministerio de Industria Economía y Competividad, 2017) en las siguientes líneas:

Con el fin de impulsar el acceso a datos de investigación, los proyectos de I+D+i financiados podrán incluir, con carácter optativo, un plan de gestión de los datos de investigación que se depositarán en repositorios institucionales, nacionales y/o internacionales tras la finalización del proyecto y trascurrido el plazo establecido en las correspondientes convocatorias.

Estas dos perspectivas fueron clasificadas por el SPARC Europe y el Data Curation Center como “duras” o “suaves” en relación a la implantación de los datos de investigación en el contexto de la UE. Mientras que la perspectiva dura sería la que concibe la publicación en abierto de los datos como un imperativo y una obligación, la

¹ Se trata de una serie de directrices piloto en el marco de los proyectos financiados con el programa Horizonte 2020, que se empiezan a poner en marcha en el año 2017 y que son aplicables a algunos proyectos anteriores al 2017 y a todos los posteriores a ese año hasta el 2020. Estas directrices están formadas por indicaciones para el uso compartido de los datos que sustentan las publicaciones derivadas de estos proyectos. En el caso de los proyectos financiados a partir de 2017, llevan incluida estas directrices por defecto, pero no son de obligado cumplimiento y se abre la posibilidad de que los miembros del proyecto argumenten en contra de poner los datos en abierto.

² Se trata de un código de conductos estipulado por el Gobierno de Dinamarca y que busca asegurar la credibilidad, integridad y calidad en la investigación danesa a través de unos principios y estándares comunes para una investigación responsable (Schmitt et al., 2016b).

suave es la que aboga por promover y alentar, pero sin obligatoriedad (Digital Curation Center y SPARC Europe, 2019).

Por otra parte, desde la perspectiva del personal investigador de Ciencias de la Salud, también existen distintas posturas y puntos de vista, que va desde posiciones defensoras del data sharing más “puro”, a otras más discrepantes, encontrándose también una gama amplia de grises.

Como ejemplo de la posición de defensa del data sharing más puro, se encuentra la opinión de Gøtzsche (2012). Este autor manifiesta que, en el campo de la biomedicina, existe un no cumplimiento de la Declaración de Helsinki³, en el sentido de que tanto investigadores como la industria privada dificultan el acceso a datos brutos relativos a ensayos aleatorios, lo que, en última instancia, provoca que no se puedan escoger los mejores tratamientos para los pacientes en relación de costos y efectividad. Desde su punto de vista, las inversiones con dinero público sacan poco rendimiento de las investigaciones en salud, ya que existe demasiado sentido de la propiedad de los datos. En esta línea, apela al papel de las personas que participan en los ensayos clínicos, con los riesgos que a veces conlleva, y que las obligaciones con ellos deben anteceder a los intereses comerciales. Estas obligaciones consistirían en más acceso a los datos brutos y a los protocolos de los estudios, para evitar duplicidades y permitir la realización de nuevas investigaciones con los mismos datos, pero para obtener resultados diferentes.

Sin embargo, otras opiniones como la de Longo y Drazen (2016) plantean cuestiones diferentes sobre el uso compartido de datos en Ciencias de la Salud. En su opinión, por debajo de los imperativos morales que detectan en los argumentos para compartir datos, tales como que compartir datos es lo justo de cara a los pacientes, subyacen otras preocupaciones para el personal investigador. Entre éstas, destacan la posibilidad de que investigadores ajenos a la investigación puedan hacer un uso incorrecto o descontextualizado de los datos, o que se combinen datos que, por sus

³ La Declaración de Helsinki, desarrollada por la World Medical Association, es el estamento que rige cuáles son los principios éticos que deben seguirse para la investigación médica relacionada con sujetos humanos (World Medical Association, 2013).

diferentes características, no deben metodológicamente mezclarse. Además, otra de las preocupaciones que destacan es la posible aparición de un nuevo grupo de investigadores, a los que denominan *research parasites* que utilizan los datos de otros para sus propios fines, obteniendo productividad investigadora que debería pertenecer al grupo original, o que incluso podrían utilizar los datos para refutar al grupo generador. En sus planteamientos, ellos abogan por que el data sharing sea una especie de simbiosis o trabajo combinado donde el grupo que llega después ofrezca nuevos planteamientos de explotación de los datos, pero contando en todo momento, en términos de participación y reconocimiento, con el grupo inicial.

La perspectiva de Longo y Drazen concuerda en ciertos aspectos con los resultados obtenidos por Tenopir et al., (2015), en su estudio sobre las prácticas y percepciones de investigadores de varias disciplinas sobre el uso compartido de datos. Este estudio, que utiliza un cuestionario como herramienta, es una continuación de otro de las mismas características realizado tres años antes, por lo que se pretende establecer una comparativa de la evolución del fenómeno. Aunque en general encuentran que el personal investigador está más abierto a compartir que tres años antes, es interesante destacar los datos concretos del personal investigador de Ciencias de la Salud. La primera información relevante es que encuentran una tendencia claramente distintiva entre el personal investigador que trabajan con humanos, como las Ciencias de la Salud/Medicina, Psicología o Ciencias Sociales, y los que no, en términos de que los que investigan con personas perciben con mucha menos fuerza que la falta de acceso a los datos sea un impedimento para la ciencia. Además, son los que menos metadatos utilizan para describir y contextualizar a los datos. En esta misma línea de pensamiento, el personal investigador de Ciencias de la Salud son los que significativamente menos estarían dispuestos a compartir una parte de sus datos o todos sus datos, y entre los que más piensan que existe una carencia de tiempo para realizar esta práctica. Sin embargo, cabe destacar que, en el anterior estudio publicado en el año 2011, Tenopir et al., (2011) encuentran que la gran mayoría de las personas encuestadas del área de la medicina perciben que compartir los datos es una buena oportunidad de colaboración y son también los que hicieron más hincapié en

que los creadores iniciales de los datos deben tenerse en cuenta, en cuanto a permisos y reconocimientos, para futuros usos.

Al margen de las políticas y mandatos, y de las percepciones del personal investigador, la otra cara de la realidad es que existe una tendencia global de aumento de datos generados en investigaciones en el campo de la salud, como los que se relacionan con la genética, que es exponencial y frente a la que es necesario diseñar estrategias. Según Birney et al., (2017), para el año 2030, 83 millones de genomas de enfermedades raras serán secuenciados, así como casi 250 millones para el diagnóstico de cáncer. En este sentido, Knoppers y Joly (2018), hacen una revisión de qué supone en términos de infraestructuras, políticas y consideraciones legales la gestión de los datos en salud, concretamente en el campo de la genética, y su uso compartido. Principalmente, recalcan el hecho de que si no hay políticas comunes y los datos genómicos no están vinculados a la práctica médica cotidiana mediante sistemas coordinados e interoperables, difícilmente se pueden mejorar las decisiones clínicas. Basados en la afirmación de Deverka et al., (2017) de que el intercambio de datos en salud está formado por un ecosistema con distintas partes interesadas, Knoppers y Joly revisan dos iniciativas relacionadas con los datos compartidos en salud: el International Cancer Genome Consortium (ICGC) y la Global Alliance for Genomics and Health (GA4GH). En ambos casos, se destaca que sus iniciativas parten de los propios científicos y de sus necesidades. En el primero, la ICGC, vigente desde el año 2008, nació con el objetivo de generar y compartir, con unas restricciones mínimas, catálogos de anomalías genéticas en varios tipos de cáncer, por ello pone a disposición en su página web un repositorio de acceso abierto a datos de genética en cáncer. Posteriormente, en el año 2010, el consorcio crea el Data Access Compliance Office (DACO), otro repositorio de datos, pero con unas características diferentes que se explicarán más detalladamente en el siguiente apartado. En el caso de la GA4GH, fundado en el 2013, su objetivo principal es acelerar el progreso en el campo de la genética y de la salud humana mediante el desarrollo de estándares y políticas para uso responsable de los datos de genética y salud en general. En esta línea, pretenden que se creen una serie de normativas legales y éticas que promuevan la interoperabilidad y el consenso. Por ello, la infraestructura del GA4GH aboga principalmente por aspectos como la colaboración

en el uso compartido de datos para la mejora de la salud humana, el respeto por las preferencias de privacidad de los participantes, la innovación para desarrollar un ecosistema que acelere el progreso o la inclusividad en la participación y la construcción de conocimientos entre los principales actores.

En general, estas dos iniciativas, así como las percepciones del personal investigador y las diferentes políticas entre países, sirven para la ilustrar la complejidad de la gestión de los datos en salud, pero también los esfuerzos por hallar soluciones diversas ante las demandas existentes. En las conclusiones de su trabajo, Knoppers y Joly refieren, como principales desafíos, la especial necesidad de protección de este tipo de datos, a veces altamente sensibles, y la necesidad de consensos y protocolos comunes entre países e instituciones. Sin embargo, también defienden que iniciativas como las descritas en el párrafo anterior son ejemplos de éxito y que, en última instancia, la filosofía subyacente es que las bases de datos, de genética en este caso, son un bien global común.

3. Infraestructuras para compartir datos: los repositorios de datos abiertos en Ciencias de la Salud

Además de los mandatos, las políticas y las opiniones, a medida que la filosofía del uso compartido de datos se iba estableciendo, se hacía más evidente la necesidad de una infraestructura que lo sostenga, como ya se adelantaba en el apartado anterior con los ejemplos del ICGC y el GA4GH. Según la Real Academia Española (Real Academia Española, 2019), infraestructura es el *“conjunto de elementos o servicios que se consideran necesarios para la creación y funcionamiento de una organización cualquiera”*. En el caso de las infraestructuras para compartir datos, serían aquellas que proveen de los mecanismos necesarios para que esta práctica pueda llevarse a cabo, es decir, serían los medios para poder pasar de la teoría a la práctica.

Como señalan Mannheimer et al., (2019) en su estudio del año 2019, técnicamente, existen diversas opciones para que el personal investigador pueda compartir sus datos. Refieren que hasta hace aproximadamente 15 años, el personal investigador disponía

de dos sistemas, que eran los más conocidos y comunes. El primero, incluir una anotación en la propia publicación indicando que los datos de la misma están disponibles “bajo petición” expresa al autor de la publicación. Aunque esta modalidad ofrece la oportunidad de pedir los datos, no garantiza la obtención de los mismos, por lo que es una medida muy variable que, además, no garantiza la preservación de los datos. El segundo sistema, compartir los archivos de forma que éstos sean descargables desde una web, bien personal o perteneciente a alguna revista o editorial. Sin embargo, esta opción también presenta inconvenientes relacionados con la preservación, ya que los links se pueden romper, o con la no existencia de opciones de búsqueda sistemática de estos materiales.

No obstante, y de nuevo gracias al desarrollo de la tecnología y de Internet, surge una opción que supera los obstáculos de las dos anteriores: los repositorios de datos, que son, actualmente, la infraestructura para compartir datos más comúnmente aceptada (Hernández-Pérez y García-Moreno, 2013). Las ventajas de los repositorios de datos son, sobre todo, los mecanismos que ofrecen para la preservación a largo plazo, la atención que prestan a los metadatos, las herramientas para favorecer la citación a los conjuntos de datos y la búsqueda sistematizada que se puede hacer de los mismos (Mannheimer et al., 2019).

Según Kim y Burns (2016), los repositorios de datos son la respuesta “natural” a las demandas o peticiones de diversos sectores, como las instituciones que financian proyectos nacionales o internacionales, o las exigencias de las revistas y las editoriales, al personal investigador para que pongan a disposición de la comunidad científica los datos que sustentan sus trabajos. En su gran mayoría, estos repositorios están pensados para ser libres y gratuitos, permitiendo su acceso y su uso para establecer una especie de intercambio en el que quien hoy deposita, mañana puede buscar y descargar datos de otros y viceversa.

Actualmente, hay una importante cantidad de repositorios que se fueron creando como respuesta a las necesidades que iban surgiendo. Financiados y mantenidos por diversos organismos, intentan cubrir huecos para que investigadores de cualquier ámbito puedan hacer uso de ellos. Por este motivo, en el momento de escoger un repositorio, es preciso tener en cuenta cuáles son sus características y qué servicios

ofrecen. Por ejemplo, es importante saber si son específicos de un área en concreto (repositorios temáticos), si son de tipo generalista y sirven para cualquier área (repositorios multidisciplinares), o si pertenecen a una institución en concreto y sirven para que los miembros de esa institución hagan uso de ellos (repositorios institucionales) (Pampel et al., 2013). Además, es importante conocer otros detalles como la capacidad de almacenamiento que ofrecen, si permiten establecer un periodo de embargo, o si ofrecen la posibilidad de poner un DOI al dataset, entre otros (Mannheimer et al., 2019). A continuación, se describe cada uno de los subtipos de repositorios existentes (temáticos, multidisciplinares e institucionales) mediante ejemplos focalizados en el área de Ciencias de la Salud.

- Repositorios temáticos

Los repositorios temáticos son aquellos en los que su contenido está pensado para una temática concreta. Entrarían dentro de este grupo tanto un repositorio de una disciplina en concreto, como los que abarcan un grupo de disciplinas de una misma área temática. También se incluirían dentro de este grupo los que se centran en una tipología concreta de datos, como podrían ser los repositorios de datos cualitativos, o los repositorios de imágenes obtenidas mediante la técnica de neuroimagen. Es decir, son los que se centran en un área concreta y limitan, de alguna forma, la entrada de datos que no corresponden a la misma. Por este motivo, este tipo de repositorios suelen ser de los que mejor aceptación tienen (Abadal, 2012).

A continuación, se pone dos ejemplos de repositorios temáticos de datos en el ámbito de Ciencias de la Salud que son diferentes entre sí: Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>), creado y mantenido por el National Center for Biotechnology Information (NCBI) y el Data Access Compliance Office (DACO) (<https://icgc.org/daco>) creado y mantenido por el ICGC.

El primero, GEO, se define como de acceso público y su finalidad es almacenar datos de microarrays, de secuenciación de nueva generación (NGS) y otros tipos de datos genómicos funcionales de alto rendimiento. Además, provee a los usuarios de herramientas de ayuda tanto para hacer consultas como para descargar experimentos y curar perfiles de expresión génica, así como bibliografía que pueda servir de apoyo.

Por otra parte, también permite a los autores establecer el grado de apertura que tienen los datos, así como período de embargo para los mismos (Barrett et al., 2013).

En la captura de pantalla (Figura 1), se visualiza la página de inicio de GEO, donde se detallan las opciones para obtener información del funcionamiento el recurso, las herramientas que posee, la opción de navegar por el contenido disponible y la explicación de cómo se pueden subir y consultar los datos. Además, la caja de búsqueda de la derecha permite buscar un conjunto de datos en concreto si se conoce previamente el número de referencia. En la Figura 2, se observa un ejemplo del depósito de un dataset de un artículo que ya ha sido publicado y se buscó a través del número de referencia. Los datos han sido compartidos en el formato libre .csv, lo que significa que se trata de un archivo interoperable, donde también se observa cómo están descritos los metadatos que explican su contenido.

Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.

Keyword or GEO Accession

Getting Started	Tools	Browse Content
Overview	Search for Studies at GEO DataSets	Repository Browser
FAQ	Search for Gene Expression at GEO Profiles	DataSets: 4348
About GEO DataSets	Search GEO Documentation	Series: 117956
About GEO Profiles	Analyze a Study with GEO2R	Platforms: 20090
About GEO2R Analysis	Studies with Genome Data Viewer Tracks	Samples: 3215895
How to Construct a Query	Programmatic Access	
How to Download Data	FTP Site	

Figura 1. Captura de pantalla de la página principal de GEO

NCBI > GEO > [Accession Display](#) [?] Not logged in | [Login](#) [?]

Scope: Format: Amount: GEO accession:

Series GSE137988 [Query DataSets for GSE137988](#)

Status Public on Oct 09, 2019
 Title Inhibition of the Dead Box RNA Helicase 3 prevents HIV-1 Tat and cocaine-induced neurotoxicity by targeting microglia activation
 Organism [Rattus norvegicus](#)
 Experiment type Expression profiling by high throughput sequencing
 Summary We have investigated the protective effect of DDX3 inhibitor RK33 on toxic damage done by combined TAT and cocaine.

Overall design Primary neuronal cultures were treated with HIV-protein TAT and cocaine following RK33 treatment. Each treatment condition was done in biological triplicate.

Contributor(s) [Aksenova M](#), [Sybrandt J](#), [Cui B](#), [Sikirzhytski V](#), [Ji H](#), [Odhiambo D](#), [Lucius MD](#), [Turner JR](#), [Broude E](#), [Peña E](#), [Lizarraga S](#), [Zhu J](#), [Safro I](#), [Wyatt MD](#), [Shtutman M](#)

Citation missing *Has this study been published? Please [login](#) to update or [notify](#) GEO.*

NIH grant(s)

Grant ID	Grant title	Affiliation	Name
R21	Impact of Dead Box RNA Helicase 3	UNIVERSITY OF SOUTH CAROLINA	Michael Shtutman
DA047936	signaling on HIV-1 Tat- and cocaine-induced neurotoxicity		

Submission date Sep 25, 2019
 Last update date Oct 09, 2019
 Contact name HAO JI
 E-mail(s) jj5@email.sc.edu
 Phone 8037774689
 Organization name University of South Carolina
 Department College of Pharmacy
 Street address 715 SUMTER ST CLS RM 109
 City COLUMBIA
 State/province SC
 ZIP/Postal code 29208
 Country USA

Platforms (1) [GPL22396](#) Illumina HiSeq 4000 ([Rattus norvegicus](#))

Samples (24) [GSM4096269](#) TAT_Rep1
[GSM4096270](#) TAT_Rep2
[GSM4096271](#) TAT_Rep3

Relations
 BioProject [PRJNA574107](#)
 SRA [SRP223249](#)

Download family	Format
SOFT formatted family file(s)	SOFT [?]
MINiML formatted family file(s)	MINiML [?]
Series Matrix File(s)	TXT [?]

Supplementary file	Size	Download	File type/resource
GSE137988_20180621_Rat.cpm_AVG_glmFit_.csv.gz	4.1 Mb	(ftp) (http)	CSV
GSE137988_20190214_RatNeuro_cpm_AVG_glmFit_.csv.gz	3.4 Mb	(ftp) (http)	CSV

[SRA Run Selector](#) [?]
 Raw data are available in SRA
 Processed data are available on Series record

Figura 2. Captura de pantalla de un conjunto de datos depositados en GEO

En el caso de DACO, se trata de un repositorio con otras posibilidades. Como se adelantaba en el párrafo anterior, su área temática está centrada en el campo de la genética, pero aplicada al cáncer. Al contrario que en el caso de GEO, DACO no ofrece acceso público, ya que solicita que los usuarios se identifiquen para poder acceder a los datos. La razón por la que han decidido hacer esto, según refieren, es debido a las dificultades de protección de datos sensibles que se manejan en algunas de estas investigaciones. Sin embargo, a cambio de ese acceso restringido se comprometen a ofrecer una centralización de los datos, junto con el desarrollo de una solicitud de acceso que sea rápida y sencilla, que evite cláusulas muy legales o técnicas, lo que les permite mayor velocidad en las respuestas. Según sus propias estimaciones, DACO procesa las nuevas peticiones en aproximadamente una semana, mientras que otros recursos más descentralizados tardan meses. La experiencia del International Cancer Genome Consortium y DACO sirve como argumento consistente de que el acceso controlado se puede considerar un sustituto aceptable del acceso completamente abierto a datos genómicos y clínicos más sensibles, solo si el proceso esté tan centralizado y simplificado como lo permitan las normas legales y éticas (International Cancer Genome Consortium, 2019; B. M. Knoppers y Joly, 2018).

En la Figura 3 se puede observar la pantalla de inicio del DACO, donde se encuentran una serie de explicaciones de cómo crear una cuenta y cómo solicitar el acceso a la base de datos.

The screenshot shows the Data Access Compliance Office (DACO) website. At the top, there is a navigation bar with the ICGC logo, a 'Data Portal' link (Get Cancer Data), a 'Data Access Compliance Office' link (Apply for Access to Controlled Data), and a 'Contact Us' button. On the right side of the navigation bar, there are links for 'Log In' and 'Create an Account'.

The main header area includes the ICGC logo and the text 'International Cancer Genome Consortium'. To the right of the logo is the 'Data Access Compliance Office' title and a search bar with the placeholder text 'Enter keywords' and a 'Search' button.

Below the header, there are navigation tabs: 'DACO Home', 'Access Forms', 'Research Practices', 'International Data Access Committee', and 'DACO Approved Projects'.

The central content area is titled 'Data Access Compliance Office' and features a three-step process:

- 1. Create your account**: Includes a 'Register' link to receive an account and a 'Learn How' button.
- 2. Fill your application**: Includes a 'Your application' link that tells users who they are and a 'Learn How' button.
- 3. Submit to DACO**: Includes a 'Send your application' link and a 'Learn How' button.

Below this process, there is a section titled 'The Data Access Compliance Office (DACO): Handles requests from scientists for access to controlled data from the International Cancer Genome Consortium (ICGC)'. A 'Show' link is visible to the right of this text.

A 'Please note' box contains the following text: 'Please note: The Data Access Compliance Office is now accepting applications for access to Controlled Data (see controlled access data types on the [ICGC Policy](#) page). There is now Controlled Access Data at the [ICGC data portal](#) (Data Coordination Center, DCC).'

At the bottom of the page, there is a button that says 'Login | Create your ICGC Account to Start Your Application' and a small icon of a person with a lock.

Figura 3. Captura de pantalla del Data Access Compliance Office

A modo de conclusión, destacar que estos repositorios se utilizan en un campo donde el uso compartido de datos está demostrando ser más efectivo: la genética. Una de las pruebas de que es una práctica efectiva es que es frecuente, tal y como destacan Knoppers, Harris, Budin- Ljøsne y Dove (2014) cuando refieren que “the culture of sharing and the development of policies to enable research collaboration are clearly pervasive in genomics research”. Por otra parte, Lachmann y Ma’ayan (2019) destacan dos de los motivos para el auge del uso compartido de datos en esta área. Por un lado, destaca la aparición de nuevas y mejoradas tecnologías de perfil transcriptómico, como la secuenciación de ARN (secuencia de ARN); por otro, el creciente mandato de las revistas de depositar datos transcriptómicos como requerimiento para publicar, con el objetivo de permitir la reproductibilidad y la reutilización de los datos.

- Repositorios multidisciplinares

Los repositorios multidisciplinares se consideran aquellos que están diseñados para dar cabida a todas o a la mayoría de las disciplinas, y para aceptar múltiples tipos de materiales. Para ejemplificar esta tipología, se describirán los siguientes repositorios: Zenodo, Dryad y Figshare.

- Zenodo

Al contrario de lo que sucede con GEO y DACO, Zenodo es un repositorio abierto para todos los tipos de investigaciones y de disciplinas, postulándose como “la solución” cuando no se dispone de un repositorio temático ni institucional. Es un repositorio alojado en el CERN (Organización Europea para la Investigación Nuclear) y financiado por la UE, que sirve de referencia para ser utilizado por los beneficiarios de los proyectos H2020 cuando quieran o deban depositar sus datos, por lo que también que es el repositorio recomendado Open Aire⁴. Aunque los datos son el núcleo de Zenodo, por lo que pueden subirse sin necesidad de ir acompañados de otros documentos como la publicación original, también admiten otros archivos, como los propios artículos, actas de congresos, proyectos, etc. Otra posibilidad de Zenodo es la de subir varias versiones de un trabajo, que quedan reflejadas como un historial que permite a los propietarios de los datos seguir la evolución de su trabajo, y a quienes los consultan, ver cómo evolucionan los proyectos. En la Figura 4 se presenta la pantalla de inicio de Zenodo, donde se presentan los últimos archivos subidos y una opción para realizar búsquedas. Además, también ofrece la posibilidad de ver las estadísticas de uso de tipo alométrico. En la Figura 5, se observa un ejemplo de un dataset almacenado en Zenodo en formato .xlsx, junto con sus metadatos y su doi (Zenodo, 2019).

⁴ OpenAIRE es una infraestructura digital y un mecanismo de soporte para la identificación, depósito, acceso y monitorización de proyectos financiados en el marco de la UE. Todos los artículos publicados que resulten de proyectos financiados por la UE están accesibles a partir del portal de OpenAire.

The screenshot shows the Zenodo homepage with a blue header containing the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. On the right side of the header are 'Log in' and 'Sign up' buttons.

The main content area is titled 'Recent uploads' and features three entries:

- August 8, 2019 (v2) Dataset Open Access**: 'Analysis of correlation-based biomolecular networks from different omics data by fitting stochastic block models'. Authors: Baum, Katharina; Rajapakse, Jagath C.; Azuaje, Francisco. Uploaded on August 8, 2019. 2 more version(s) exist for this record.
- August 6, 2019 (v0.1) Software Open Access**: 'OpenScienceMOOC/Module-5-Open-Research-Software-and-Open-Source: 3.1'. Authors: Jon Tennant, Julien Colomb, Lisa Matthias, Simon Worthington, Florian Kohrt, Irubio, Tania Allard, Philipp Zumstein, Daniel S. Katz, Alexander Morley, Tobias Steiner, Stephan Druskat, Zoran Pandovski, Arfon Smith, Gabriele Orlandi, Rutger Vos, José Raúl Canay Pazos, Paul Griffiths, Nithiya Streethran, Hollie Marshall, Luke W Johnston, Luis Camacho, Konrad Förstner, Heidi Seibold, EricDWilkey, Encarnación Martínez Álvarez, Brendan Palmer, Alessandro Sarretta, Alberto Marrochino, Abigail Cabunoc Mayes. Released now that the content is live on the open source platform, Elademy. Uploaded on August 6, 2019. 3 more version(s) exist for this record.
- March 15, 2019 (v1.1.0) Dataset Open Access**: 'Global whole-rock geochemical database compilation'. Authors: Gard, Matthew Geoffrey; Hasterok, Derrick; Halpin, Jacqueline. Dataset associated with the submitted publication - Global whole-rock geochemical database compilation. Uploaded on August 4, 2019. 1 more version(s) exist for this record.

On the right side, there are three promotional banners:

- Zenodo now supports usage statistics!** with a line graph icon and a link to read more.
- Using GitHub?** with the GitHub logo and a link to log in with a GitHub account.
- Zenodo in a nutshell** with a list of features:
 - Research. Shared.** — all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
 - Citeable. Discoverable.** — uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
 - Communities** — create and curate your own community for a workshop, project, department, journal, into which you can accept or reject uploads. Your own complete digital repository!
 - Funding** — identify grants, integrated in reporting lines for research funded by the European Commission via OpenAIRE.
 - Flexible licensing** — because not everything is under Creative Commons.
 - Safe** — your research output is stored safely for the future in the same cloud infrastructure as CERN's own LHC research data.

Figura 7. Captura de pantalla de la página principal del repositorio Zenodo.

zenodo Search Upload Communities Log in Sign up

December 18, 2018 Dataset Open Access

Use and sharing of raw data in the Journal Citation Reports' Emergency Medicine Category: Metrics and Journals including supplementary material classification sorted by quartile of the JCR emergency medicine category.

Sixto-Costoya, A; Aleixandre-Benavent, R; Lucas-Dominguez, L; Vidal-Infer, A

Raw data belonged to the study of use and sharing of raw research data in the Journal Citation Reports' Emergency Medicine Category.

Files (19.3 kB)

Name	Size	Download
SM features.xlsx	19.3 kB	Download

md5:5c92bf5f2dc089b9f9c8a8da649e4e08

Citations 0

Show only: Literature (0) Dataset (0) Software (0) Unknown (0) Citations to this version

No citations.

14 views 4 downloads See more details...

Indexed in OpenAIRE

Publication date: December 18, 2018

DOI: 10.5281/zenodo.2384871

Keyword(s): data sharing emergency medicine supplementary material raw data open science

License (for files): Creative Commons Attribution 4.0 International

Versions

Version	Date
Version 1 10.5281/zenodo.2384871	Dec 18, 2018

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.2384870. This DOI represents all versions, and will always resolve to the latest one. Read more.

Share

Cite as

Sixto-Costoya, A, Aleixandre-Benavent, R, Lucas-Dominguez, L & Vidal-Infer, A. (2018). Use and sharing of raw data in the Journal Citation Reports' Emergency Medicine Category: Metrics and Journals including supplementary material classification sorted by quartile of the JCR emergency medicine category. (Version 1) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.2384871>

Export

BibTeX CSL DataCite Dublin Core JSON JSON-LD MARCXML Mendeley

Figura 8. Captura de pantalla de un registro que incluye un conjunto de datos depositados en el repositorio Zenodo

- Figshare

Se trata de otro repositorio abierto que acoge datos de varias disciplinas, que está sustentado por la Digital Science-MacMillan Publishers Company. Su público objetivo

son el personal investigador de cualquier disciplina, e igual que Zenodo, también otorga un doi a los archivos subidos y estadísticas de uso de tipo almétrico. El tipo de archivos que acepta va desde los propios artículos, a imágenes, vídeos, datasets, pósters, etc (Figshare, 2019). Según la autora Jane Cho (2019), Figshare es el repositorio multidisciplinar que más datos contiene. En la Figura 6, se presenta una imagen de la pantalla de inicio de Figshare, que tiene un aspecto similar a una red social. Una de las utilidades de Figshare es que permite filtrar por materia. En la Figura 7 se pone un ejemplo de una imagen relacionada con una investigación en el área de la epigenética.

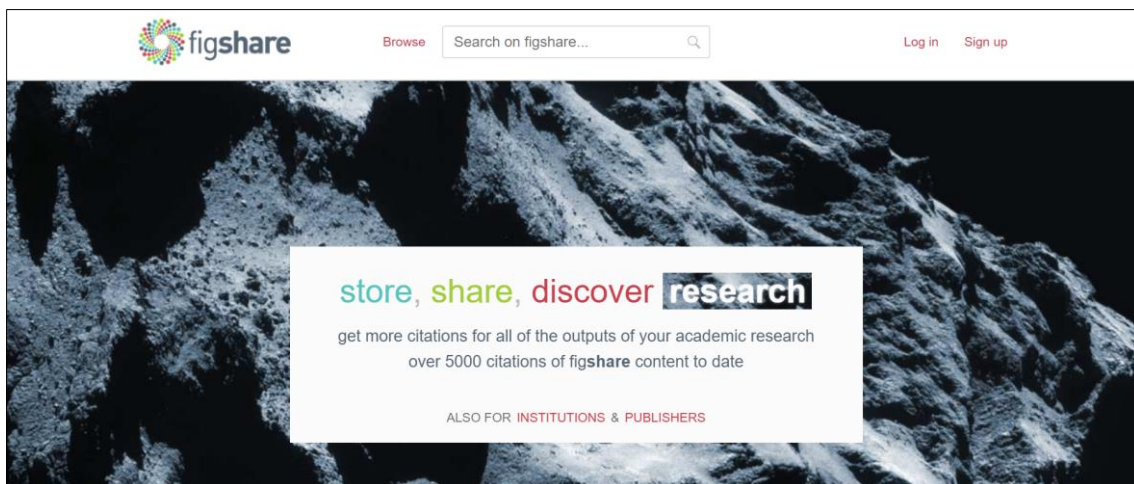


Figura 6. Captura de pantalla de la página de inicio del repositorio Figshare

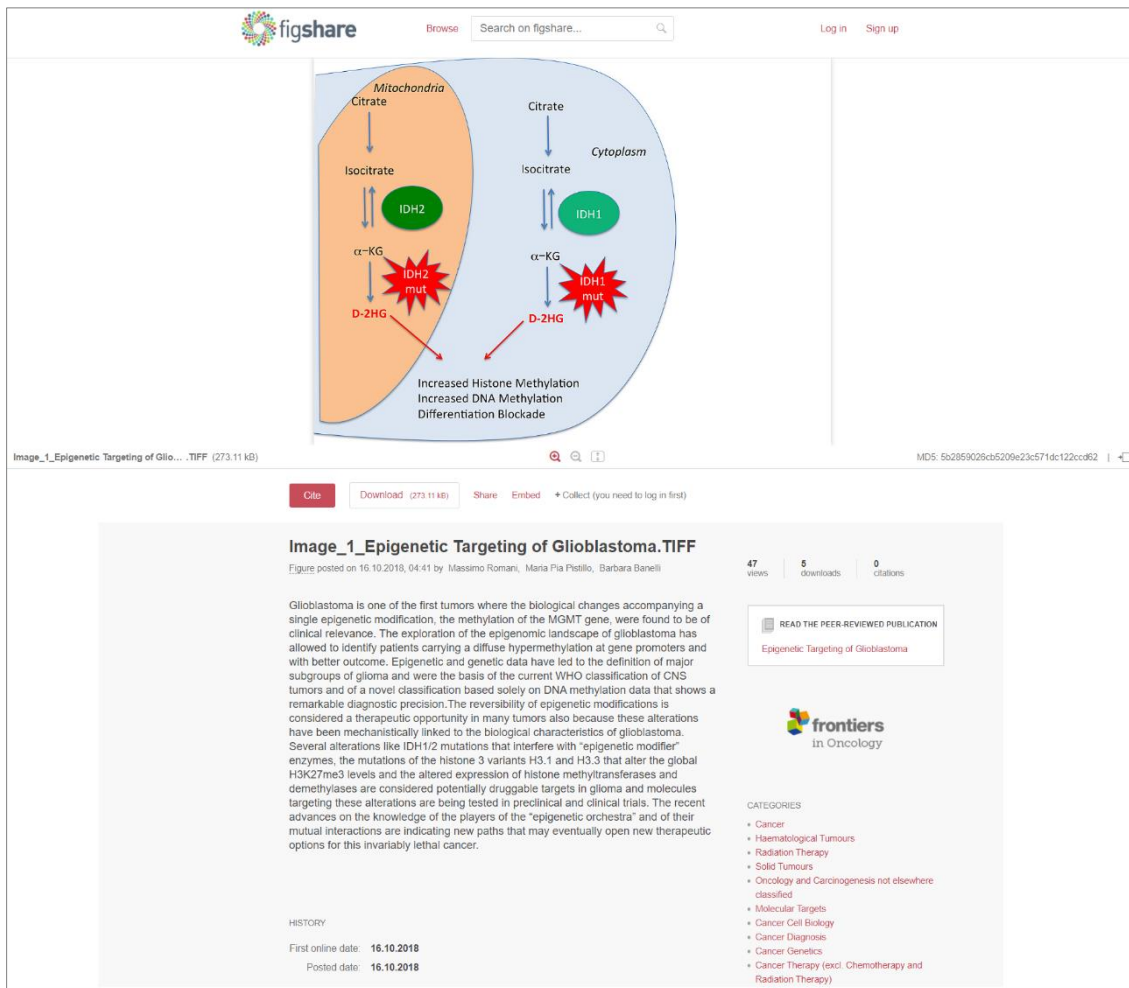


Figura 7. Captura de pantalla de una imagen compartida en el repositorio Figshare

- Dryad

Dryad es un repositorio multidisciplinar pero especializado en disciplinas fundamentalmente de Ciencias de la Salud y Ciencias Biológicas. Está registrada como una ONG y utiliza un software de código abierto para funcionar. Una de las particularidades de Dryad es que todos los conjuntos de datos depositados necesariamente deben estar ligados a una publicación, al contrario de lo que sucede en Zenodo y Figshare. Este punto hace que Dryad sea un recurso que permite al usuario crearse un contexto más global al poder acceder a la publicación derivada de los datos brutos (Dryad, 2019). En la Figura 8, se observa la página principal de Dryad, en la que se encuentra información acerca del recurso y opciones para subir datos, o

para buscarlos. En la Figura 9 se presenta un ejemplo de búsqueda en Dryad de un dataset relacionado con una investigación sobre la eficacia de un fármaco.

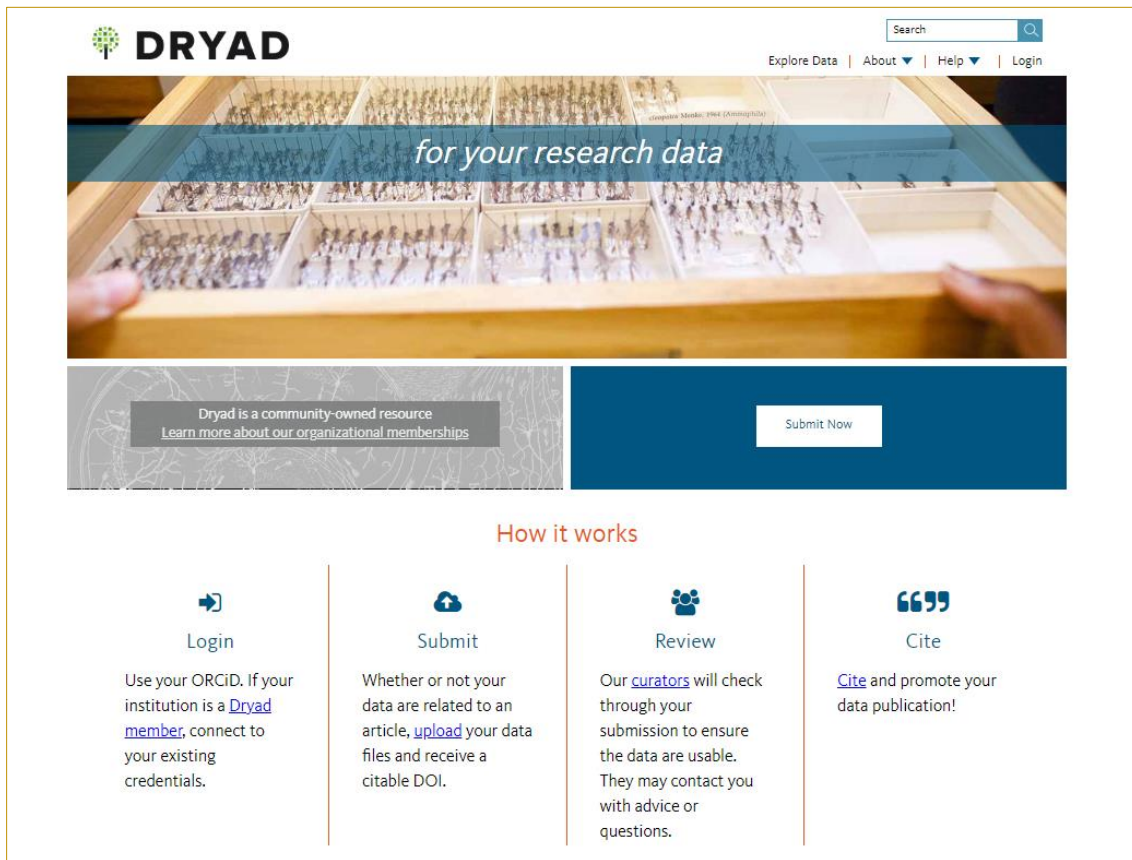



Figura 8. Captura de pantalla de la página principal del repositorio Dryad



Explore Data | About ▼ | Help ▼ | Login

Data from: Onset of clinical and MRI efficacy of ocrelizumab in relapsing multiple sclerosis

Barkhof, Frederik, VU University Medical Center
Kappos, Ludwig, University College London
Wolinsky, Jerry S., University Hospital of Basel
Li, David K. B., Unknown
Bar-Or, Amit, Unknown
Hartung, Hans-Peter, Unknown
Belachew, Shibeshih, Unknown
Han, Jian, Unknown
Julian, Laura, Unknown
Sauter, Annette, Unknown
Napieralski, Julie, Unknown
Koendgen, Harold, Unknown
Hauser, Stephen L., Unknown
elisa.glaze-santonen@articulatescience.com

Publication date: September 20, 2019
 Publisher: Dryad
<https://doi.org/10.5061/dryad.3jd86nj>

Citation

Barkhof, Frederik et al. (2019), Data from: Onset of clinical and MRI efficacy of ocrelizumab in relapsing multiple sclerosis, v2, Dryad, Dataset, <https://doi.org/10.5061/dryad.3jd86nj>

Abstract

Objective: To assess the onset of ocrelizumab efficacy on brain magnetic resonance imaging (MRI) measures of disease activity in the Phase II study in relapsing-remitting multiple sclerosis (RRMS), and relapse rate in the pooled Phase III studies in relapsing multiple sclerosis (RMS). **Methods:** Brain MRI activity was determined in the Phase II trial at monthly intervals in patients with RRMS receiving placebo, ocrelizumab (600 mg), or intramuscular interferon (IFN) β -1a (30 μ g). Annualized relapse rate (ARR; over various epochs) and time to first relapse were analyzed in the pooled population of the Phase III OPERA I and OPERA II trials in patients with RMS receiving ocrelizumab (600 mg) or subcutaneous IFN β -1a (44 μ g). **Results:** In patients with RRMS, ocrelizumab reduced the number of new T1 gadolinium-enhancing lesions by Week 4 vs placebo ($p=0.042$) and by Week 8 vs intramuscular IFN β -1a ($p<0.001$). Ocrelizumab also reduced the number of new or enlarging T2 lesions appearing between Weeks 4 and 8 vs both placebo and IFN β -1a (both $p<0.001$). In patients with RMS, ocrelizumab significantly reduced ARR ($p=0.005$), and the probability of time to first protocol-defined relapse ($p=0.014$) vs subcutaneous IFN β -1a within the first 8 weeks. **Conclusion:** Epoch analysis of MRI-measured lesion activity in the Phase II study and relapse rate in the Phase III studies consistently revealed a rapid suppression of acute MRI and clinical disease activity following treatment initiation with ocrelizumab in patients with RRMS and RMS, respectively. **Classification of evidence:** This study provides Class II evidence that for patients with RRMS and RMS, ocrelizumab suppressed MRI activity within 4 weeks and clinical disease activity within 8 weeks.

Usage Notes

Figure e-1
Phase II study of ocrelizumab in patients with RRMS
Sup fig 1_v1b.jpg

Figure e-2
Phase III OPERA I and OPERA II studies of ocrelizumab in patients with RMS
Sup fig 2_v1b.jpg


Figure e-3
The effect of treatment with ocrelizumab on B-cell counts in the pooled OPERA population
Sup fig 3_v1b.jpg


Figure e-4
The number of new T2 lesions in the Phase II population
Sup fig 4_v1c.jpg

Figure e-5
The number of enlarging T2 lesions in the Phase II population
Sup fig 5_v1c.jpg

References

This dataset is supplement to <https://doi.org/10.1212/WNL.00000000000008189>


 Download dataset ~ 359 kB


 Download Data Publication (PDF)


Data Files

> September 20, 2019

Metrics

 27 views

 7 downloads

 [1 citations](#)

Keywords

[Multiple Sclerosis](#)


[Clinical Neurology](#)

[MRI](#)




[Clinical Trials](#)

License

This work is licensed under a [CCO 1.0 Universal \(CCO 1.0\) Public Domain Dedication](#) license.



[Privacy Policy](#) | [Accessibility Policy](#) | [Terms of Service](#)

 Follow us on Twitter
  Check out our Blog
  Subscribe to our mailing list

Copyright (c) 2019 Dryad

Figura 9. Captura de pantalla del resultado de una búsqueda en el repositorio Dryad

- Repositorios institucionales

En este último grupo se encuentran los repositorios de datos creados por instituciones (universidades, centros de investigación, entre otros) con el fin de que sirvan para almacenar los datos generados por el personal investigador que pertenece a las mismas. Como ejemplo de este tipo de repositorios está Roderic, repositorio de acceso abierto de la Universitat de València, que incluye a los datos entre los materiales para los que permite su depósito. En la Figura 10 se observa la página de inicio del repositorio, que cuenta con una explicación del funcionamiento del recurso y con una opción de búsqueda. En la Figura 11, se presenta un ejemplo un dataset relacionado con una investigación sobre VIH.



Figura 10. Captura de pantalla de la página de inicio de Roderic

The screenshot shows the Roderic repository interface. At the top, there is a navigation bar with 'Valencià Castellano', 'FAQ', 'TESIS', 'El meu RODERIC', and 'Contacto'. Below this is a search bar with the text 'Buscar en RODERIC' and a search button. The main content area displays the title of the dataset: 'Dataset: Sexual reproductive phenology of three charophyte species from two Mediterranean brackish shallow ponds'. Below the title, there are social media sharing icons (Facebook, Twitter, LinkedIn, etc.) and a download button labeled 'Descargar: (47.98Kb)'. The authors are listed as 'Calero Cervera, Sara; Rodrigo Alacreu, María Antonia'. The document is dated 'Este documento es un/a , creado/a en: 2017'. There is also a link to the document's DOI: 'http://hdl.handle.net/10550/61341'. The page includes a sidebar with navigation options like 'Comunidades y colecciones', 'Títulos', 'Autores', 'Temas', and 'Tipo de documento'.

Figura 11. Captura de pantalla de la búsqueda de un conjunto de datos en Roderic

4. Las revistas científicas del ámbito de las Ciencias de la Salud y el uso compartido de datos

En el contexto del uso compartido de datos, el papel que juegan las revistas y las editoriales es primordial. Esto se debe fundamentalmente a dos razones. La primera, y quizás la más obvia, es que las revistas son el vehículo de comunicación científica más importante actualmente, por lo que las posiciones que adopten en relación al uso compartido de datos van a influir, de un modo u otro, en el éxito o fracaso de estas iniciativas (González Alcaide y Ferri, 2014). La segunda razón es que hay más probabilidades de que el personal investigador comparta datos si existe o bien requerimientos o bien incentivos. Según estudios previos, los requerimientos que pongan las revistas con respecto a la publicación de los datos tienen un enorme potencial a la hora de cambiar las perspectivas del personal investigador (Federer et al., 2018; Giofrè, Cumming, Fresc, Boedker, y Tressoldi, 2017).

En el contexto actual, con respecto a la publicación de los datos de investigación y a las revistas científicas, el auge del movimiento por el uso compartido de datos a partir de

la década de los 90 y 2000 provoca que algunas revistas y editoriales hayan tomado nota y actuado en consecuencia. En el trabajo de Alsheikh-Ali et al., (2011), sobre políticas relacionadas con la disponibilidad de los datos de investigación en revistas de alto impacto, se encuentra que de 50 revistas estudiadas, 44 tienen una declaración en sus instrucciones para los autores relacionada con la disponibilidad pública y el intercambio de datos. Aunque las características de estas declaraciones varían entre las revistas, que van desde la sugerencia hasta el requerimiento explícito, en general se observa que es una opción a tener en cuenta en las políticas de las revistas. En una línea similar, Thelwall y Kousha (2017), hacen hincapié en que el método más efectivo para compartir datos es a través del mandato de revistas y de entidades financiadoras. Además, agregan que el “espacio natural” en el que deben estar detalladas las políticas de data sharing es en las “author guidelines” que todas las revistas científicas ponen a disposición de los posibles autores. En disciplinas donde el data sharing es una práctica frecuente, como el campo de la genética, la existencia de una política de datos compartidos en las revistas también es común, como se observa en el estudio de Piwowar y Chapman (2008), cuyos resultados muestran que de 70 políticas de revistas donde se ha observado que hay links al repositorio GEO, 42 tienen políticas sobre data sharing.

A nivel de editoriales, algunas de las más relevantes y que acogen a un mayor número de publicaciones de impacto también se han sumado a añadir políticas sobre los datos. Entre estas editoriales, la primera que decidió sumarse a esta iniciativa fue PLOS en el año 2014, seguida de otras como Springer Nature, Elsevier o Science. Las políticas que establecen estas editoriales para justificar la publicación de los datos se basan en facilitar la validación y replicación de los resultados de investigación, promocionar la reutilización de los datos y apoyar las citas a los datos para asegurar que el personal investigador recibe crédito cuando sus datos son reutilizados (Federer et al., 2018). Además de estas editoriales “grandes”, otras instituciones como la German Association for Medical Education (GMA) también han apostado por este tipo de políticas en su revista *GMS Journal for Medical Education*, en una iniciativa llevada a cabo en colaboración con el repositorio Dryad (Arning, 2015).

Como detallan Federer et al., (2018), a pesar de que la mayoría de las políticas sobre publicación de datos son similares a grandes rasgos, la diferencia está en cómo de estricta sea la política, ya que mientras algunas editoriales o revistas alientan estas prácticas, otras la exigen como requisito. Un ejemplo ilustrativo son las diferencias entre las políticas de las editoriales PLOS y Elsevier. En el caso de PLOS, y tomando como ejemplo la revista biomédica *PLOS One* (PLOS One, 2019) editada por esta editorial, se establece que:

When submitting a manuscript, authors must provide a Data Availability Statement describing compliance with PLOS' data policy. If the article is accepted for publication, the Data Availability Statement will be published as part of the article.

Lo que significa que se trata de un requerimiento, es decir, un paso más que hay que cumplir si se desea publicar en esta revista. Al margen de este requerimiento, PLOS reconoce que existen determinados casos en los que los autores no pueden, por motivos éticos o legales, hacer públicos los datos que subyacen a las publicaciones, En ese caso, deben explicar y justificar sus razones en el “Data Availability”, un apartado que se debe cubrir en el momento del envío del artículo en el que se describe cómo son compartidos los datos, o en el caso de que no se puedan compartir, explicar el porqué.

Por otra parte, en el caso de las revistas de la editorial Elsevier, en el primer punto de sus políticas sobre la disponibilidad de datos se establece lo siguiente (Elsevier, 2019):

Encourage and support researchers to share research data where appropriate and at the earliest opportunity, for example by enhancing our submission processes to make this easier.

De esta frase se desprende que Elsevier apoya esta práctica e incluso muestra la intencionalidad de facilitarla, pero, al contrario que PLOS, se trata de una promoción y no de una obligación.

Otro ejemplo, en este caso de una revista que alienta, pero no obliga es la *British Medical Journal* (BMJ), en la que la política establecida en relación a los datos de investigación se basa en promover, pero sin detallar requerimientos específicos. A pesar de esta política, según McDonald et al., (2017), se trata de una revista en la que

se obtiene que, en el caso de los estudios observacionales publicados entre 2015 y 2017, un 63% incluyen una declaración de que los datos no serán compartidos. La posición de esta revista es interesante ya que pone hincapié en la utilidad de la investigación que publica para mejorar la práctica clínica, por lo que su posible público objetivo serán profesionales que estén en estrecho contacto con pacientes, profesionales que, según Tenopir et al., (2011) son de los más reticentes a la hora de compartir datos. Un aspecto relevante que muestra esta revista es que, en todo caso, los autores deberán incluir un “data sharing statement” (British Medical Journal, s.f.):

We require a data sharing statement for all research papers. For papers that do not report a trial, we do not require that the authors agree to share the data, just that they say whether they will.

Esta declaración obligatoria, semejante a la de PLOS, tiene la ventaja de que permite saber por qué no se decide no compartir los datos, que deberá ir acompañado de una explicación.

Con respecto a esto, Federer et al., (2018) refieren que las palabras “disponibilidad de los datos” pueden interpretarse de varias maneras e implicar distintos niveles de compromiso. En este sentido, existen tres formas según las cuales puede encararse una política de datos compartidos en revistas y editoriales: bajo petición al autor/es, como material suplementario a la publicación y depositarlos en un repositorio de datos. A continuación, se explicará con detalle cada una de estas tres modalidades.

5. Modalidades que utilizan las revistas para compartir datos

5.1 Bajo petición expresa al autor/es del artículo

La primera modalidad, que no implica ni obligación ni exigencia, es la que le permite al autor poner una indicación de que los datos pueden ser solicitados al autor o autores. En esta modalidad los autores de la publicación dan la opción, mediante una frase más o menos de este estilo: “*The datasets generated for this study are available on request to the corresponding author*”, de que otros autores se pongan en contacto y soliciten

los datos. Para ello, el autor o autores deben facilitar una forma de contacto, que debe ser preferiblemente un correo electrónico de la institución o departamento correspondiente, evitando las direcciones personales (Comisión Económica para América Latina y el Caribe, 2019). Se trata de una forma limitada de ofrecer disponibilidad de los datos, ya que el autor original puede aceptar o rechazar la peticiones sin justificar su decisión. Además, es una modalidad en la que se pasan por alto los mecanismos para controlar que los datos estén correctamente preservados y puedan estar disponibles a largo plazo. Por otra parte, autores como Wicherts et al., (2006), remarcan otra desventaja de esta modalidad y es que, según refieren después de un análisis realizado en un grupo de revistas editadas por la American Psychological Association, descubren que del total de peticiones que se han hecho a los autores solicitando los datos, un 73% rechazaron compartirlos, a pesar de que las peticiones habían sido debidamente justificadas (descripciones detalladas de los propósitos del estudio, aprobación del comité ético y aseguramiento firmado de que no se compartirán los datos sin consentimiento).

5.2 Material suplementario

El segundo nivel, que implica un grado más apertura de los datos, es el que ofrece la posibilidad de que éstos vayan adjuntos como material suplementario al artículo. Estos materiales cumplen la función de ayudar a decidir a los editores y revisores sobre la publicación del artículo, proporcionar información extra para lectores que quieran profundizar sobre la temática, o utilizar los datos bien para replicar la investigación, o para realizar un análisis secundario. Según Price et al., (2018), en el caso de que una revista opte por permitir este tipo de materiales, existen diferentes enfoques para que los autores adjunten el material, que varían según el énfasis que las revistas pongan en la publicación de este material. Por ejemplo, las revistas pueden optar por indicar si los revisores deben revisar también el material suplementario, o si, por el contrario, es un material que no se revisa. Por otra parte, las revistas pueden decidir qué uso quieren que los autores hagan del material suplementario, si se desea que sirva para replicar resultados, o para ofrecer información “extra” de los resultados de la investigación.

Esta variedad de enfoques obliga a los autores, revisores y lectores a otorgar distintos grados de priorización a la hora de decidir si incluirlo, leerlo o utilizarlo, dependiendo de cuáles sean sus intereses. En relación con esto último, en los resultados del estudio de Price et al., cuyo objetivo es explorar el valor, el uso y el rol que juega este tipo de material en las revistas biomédicas desde la perspectiva de autores, lectores y revisores, se observa precisamente que dependiendo del uso que se le quiera dar a los datos, la perspectiva cambia. Por ejemplo, si lo que se desea es ampliar el contenido del artículo a modo informativo, se prefiere el material suplementario esté formado por tablas o figuras; sin embargo, éstas no sirven cuando lo que se desea es reutilizar los datos o replicar el estudio. Además, según este estudio, es muy poco frecuente encontrar datos brutos como material suplementario ya que un 80% de los revisores manifiestan encontrarlos muy raramente. Esto concuerda con los resultados de los estudios de Aleixandre-Benavent et al., (2018), Vidal-Infer et al., (2018, 2019) y Sixto-Costoya et al., (2020, 2021) en los que se analizan diferentes disciplinas de Ciencias de la Salud y su relación con el uso compartido de datos a través del material suplementario.

En relación con estos estudios de Aleixandre-Benavent, Vidal-Infer y Sixto-Costoya et al., se explican de manera más detallada en el apartado de Método, ya que se realizaron en el mismo marco contextual de esta tesis: el proyecto coordinado “Modelos para publicar, consumir y medir la reutilización de datos de investigación: más allá de las fronteras institucionales y geográficas”, financiado por Ministerio de Ciencia e Innovación (Plan Estatal de Investigación Científica y Técnica y de Innovación 2013-2016. Programa Estatal de I+D+I Orientada a los Retos de la Sociedad).

Para finalizar este apartado sobre el material suplementario, y volviendo al trabajo de Price et al., estos autores concluyen que, aunque adjuntar material suplementario tiene aspectos positivos y son una posibilidad real para compartir datos, presenta problemas en aspectos relacionados con la replicabilidad de los estudios y con la preservación de los datos a largo plazo. A este respecto, otros autores como Hofner, Schmid y Edler (2016) argumentan que son los repositorios reconocidos y con probada solvencia la vía más eficaz para depositar y compartir este tipo de material, especialmente el que contiene datos brutos.

5.3 El depósito en repositorios de datos

La tercera vía, considerada la más eficaz para depositar datos, es el depósito en repositorios de datos (Hofner et al., 2016). Esta modalidad, basada en que la revista obligue o recomiende subir los datos a repositorios como los que se han mencionado en el apartado 3, es la que presenta más posibilidades en cuanto a la apertura, acceso y preservación a medio y largo plazo. Además, mediante esta vía se ofrece la posibilidad de que los conjuntos de datos sean tratados como unidades independientes de la publicación, que puede venir dada por el doi que muchos repositorios (Zenodo, Dryad, Dataverse...) otorgan a los datasets una vez depositados. En el caso de los repositorios biomédicos, Pampel et al., (2013) refieren que, a nivel de disciplinas, existe un gran número de repositorios, mencionando la existencia de 1.512 infraestructuras donde se pueden depositar datos biomédicos. Sobre las políticas de las revistas y de las editoriales a las que pertenecen, pueden variar sobre todo dos cuestiones. La primera, es si se trata de una obligación o de una recomendación. La segunda, el tipo de instrucciones que se detallen, por ejemplo, si se da libertad al autor para que se escoja el repositorio o si la revista ya tiene una lista cerrada de repositorios para elegir. Por ejemplo, en el caso de la política de la revista *PLOS One*, se observa que se recomienda encarecidamente la opción del depósito en repositorios mediante la siguiente frase “Deposition within data repository (strongly recommended)”, seguido de una serie de instrucciones sobre cómo debe hacerse; sin embargo, se ha demostrado que ésta es la opción que menos utilizan los autores que publican en esta revista (Federer et al., 2018; *PLOS One*, 2019). Por otra parte, otras revistas, como es el caso de *American Journal of Political Science*, establecen una política según la cual se obliga a los autores que faciliten los datos que sostienen sus resultados como condición para publicar en esta revista, principalmente con el fin de permitir la replicabilidad (*American Journal of Political Science*, s.f.).

6. El personal investigador de Ciencias de la Salud y el uso compartido de datos

Con el objetivo de conocer las opiniones y posturas del personal investigador sobre el uso compartido de datos se han elaborado una serie de estudios a nivel internacional, sobre todo basados en cuestionarios y posterior análisis de los mismos. Aunque ya se ha comentado que hay referencias al uso compartido de datos entre investigadores desde antes de la popularización del Open Science, que incluye a los datos en su concepto de “open”, el interés por las posiciones del personal investigador sobre este tema se incrementa. La mayor parte de estos estudios están fundamentados en que indagar acerca de los hábitos, las motivaciones, las actitudes o las amenazas que el personal investigador percibe sobre esta práctica es fundamental para conocer el alcance y posibilidades de la misma.

6.1 Perspectiva cuantitativa

En primer lugar, el marco del proyecto que sustenta esta tesis, se ha llevado a cabo un estudio para averiguar cuáles son las prácticas del personal investigador español en el uso compartido de datos (Aleixandre-Benavent et al., 2015; Rafael Aleixandre-Benavent, Vidal-Infer, Alonso-Arroyo, Peset, y Ferrer-Sapena, 2020). Para ello, se elaboró un cuestionario de 40 preguntas que tiene como objetivo identificar los hábitos y experiencias del personal investigador españoles en relación con la gestión e intercambios de datos bruto, así como sus motivaciones para compartir, las circunstancias que dificultan esta práctica y el soporte técnico y las infraestructuras de las que disponen. La muestra estuvo formada por 1.053 investigadores de distintas áreas, clasificadas en Ciencias de la Salud, Ciencias Sociales, Artes y Humanidades y Física y Tecnología. Con respecto al personal investigador de Ciencias de la Salud, en los resultados de este estudio se halla que, en general, los conocimientos y hábitos sobre infraestructuras para compartir datos son escasos y medios técnicos para el depósito y la conservación son escasos. Sobre los miedos, destacan especialmente la pérdida de autoría, el que otros investigadores malinterpreten sus datos, aspectos

relacionados con la propiedad intelectual y la pérdida del liderazgo en la investigación. Por otro lado, la opción más escogida también por científicos de Ciencias de la Salud con respecto a si sus proyectos cuentan con políticas de datos y por qué, la mayoría responde que, si la tienen, es por obligación de las bases del proyecto. Sin embargo, se desataca que el personal investigador de Ciencias de la Salud es el que mayoritariamente manifiesta aportar datos brutos como material suplementario. Finalmente, mencionar la paradoja que se da en todas las áreas analizadas de que, aunque la mayoría de los encuestados manifiestan estar interesados en consultar y utilizar datos de otros (especialmente el personal investigador más joven), esto no se corresponde con su propia disposición a la hora de compartir datos.

Con respecto a otros estudios de características semejantes, se encuentran similitudes con respecto al área de Ciencias de la Salud y sus investigadores. Si se toman los estudios de Tenopir et al., (2011, 2015), interesantes por su amplitud (la muestra se extiende a investigadores de varios países y de varias disciplinas) y por su transversalidad (el estudio del 2015 utiliza los mismos parámetros que el estudio del 2011 y compara los resultados), se manifiestan parecidos en aspectos como la percepción del esfuerzo que cuesta generar datos para luego tener que prepararlos para compartirlos, o el riesgo de perder delantera o el control sobre los mismos. Sobre este último punto, el control de los datos, el personal investigador de las áreas biomédicas es el que manifiesta mayor reticencia a “dejar ir” los datos y perderles la pista, en contraste con investigadores de áreas como las Ciencias Medioambientales y la Ecología, más dispuestos a compartir sin la preocupación del control posterior una vez se comparten.

Por otra parte, también se han llevado a cabo otros estudios más focalizados en comportamientos y percepciones desde un punto de vista psicológico, como el de Kim (2017), que utiliza una combinación del marco teórico que integra la teoría institucional y la teoría del comportamiento planificado. Este estudio tiene como objetivo examinar cómo las presiones institucionales, las motivaciones individuales y los recursos disponibles influyen y afectan a los comportamientos del personal investigador sobre los datos compartidos, incluyendo el hacer los datos accesibles a través de repositorios de datos, adjuntarlos como material suplementario, o

proveerlos mediante la petición entre autores. Los datos se obtuvieron mediante un cuestionario que cumplimentaron investigadores docentes y posdoctorales en disciplinas STEM (Science, Technology, Engineering and Mathematics) en EEUU. Los resultados mostraron como diferentes presiones percibidas (agencias financiadoras, revistas, algunas normativas, etc.) influyen en la práctica de compartir datos y también en la vía mediante la que se comparten. Además, muestran como las percepciones de beneficio o de amenaza para la carrera investigadora también influyen en el comportamiento del personal investigador. Ambos aspectos, la percepción de las presiones y el sentimiento de beneficio o riesgo, se relacionan con las tres vías (repositorios, material suplementario o comunicaciones entre colegas) para compartir datos. Los resultados de estas relaciones ofrecen perspectivas muy interesantes. En primer lugar, con respecto a la presión por parte de las agencias financiadoras, se halla una relación significativa en cuanto al data sharing entre colegas, pero no para el material suplementario o el depósito en repositorios. Sin embargo, la presión de las revistas sí es significativa para el depósito en repositorios o los datos como material suplementario. En cuanto a la percepción de beneficio, el personal investigador solo manifiesta sentirlo cuando adjunta material suplementario, ya que se percibe como una posibilidad de recibir más reconocimiento y e incluso citas, ya que esos datos siempre quedan ligados a la publicación. Sobre la sensación de riesgo, solo sucede con el intercambio entre colegas, ya que no se trata de una vía “oficial” y reconocida, por lo que puede acarrear menores garantías de reconocimiento de la autoría. En último lugar, en esta investigación se detecta que, en general, el personal investigador percibe de manera significativa como un esfuerzo extra el compartir datos en cualquiera de las tres modalidades.

Otro estudio destacado, también basado en un cuestionario, es el que realiza la editorial Wiley, donde se pregunta a investigadores de varios países y disciplinas sobre sus práctica y percepciones del uso compartido de datos (Ferguson, 2014). Algunos de los resultados interesantes que aporta esta investigación están relacionados con las motivaciones para compartir datos y las razones que les producen dudas al personal investigador. En el primer caso, de las 13 motivaciones que exponen en el cuestionario, detectan como primera en preferencia el que compartir datos es una práctica estándar

en su disciplina, seguido del incremento de visibilidad para su trabajo, la creencia de que se hace por un bien público y el requerimiento por parte de las revistas. En cuanto a las dudas o razones para no compartir, también entre 13 opciones de respuesta, las más recurrente fueron: preocupaciones relacionadas con la propiedad intelectual, seguido de que no es un requisito por parte de la institución/entidad financiadora, la preocupación por que te adelanten en la investigación y, en cuarto lugar, la intranquilidad por una interpretación errónea de los datos.

6.2 Perspectiva cualitativa

En el caso de los estudios cualitativos, se ha encontrado evidencia de trabajos que abordan perspectivas sobre el uso compartido de datos en Ciencias de la Salud siguiendo esta metodología. Principalmente, se centra en aspectos específicos de colectivos muy concretos.

Sobre todo, uno de los problemas que más ha motivado la elaboración de estos estudios ha sido la preocupación por la protección de datos de los pacientes. En este caso, el data sharing está focalizado en los datos que se generan relacionados con la salud de los pacientes. Estos datos pueden generarse y almacenarse en instituciones como hospitales, centros de salud, institutos de investigación o la Academia. Algunos de estos estudios exploran directamente la perspectiva de los pacientes sobre el compartir sus datos clínicos, sobre todo para fines de investigación, pero también para industria. Estos estudios buscan fundamentalmente explorar la posibilidad de que los pacientes consientan que sus datos se compartan, hasta qué punto están dispuestos, y cuáles son sus condiciones, aunque difieren considerablemente en el contexto. De esta manera, los estudios van desde la protección de los datos de salud de colectivos vulnerables o de países con menos recursos (Cheah et al., 2018; Garrison et al., 2019); la comparativa de las actitudes de los pacientes y las implicaciones éticas con respecto a compartir datos ligados a la industria, ya sea para monitorizar la salud (pulseras de actividad, smartwatches, etc.) o empresas que ofrecen servicios de pruebas genéticas (Haeusermann et al., 2018; Ostherr et al., 2017); o bien estudios vinculados al

consentimiento de los pacientes de que sus datos se utilicen para fines de investigación, como la reutilización de los datos de sus muestras clínicas (Broes, Verbaanderd, Casteels, y Lacombe, 2020).

Por otra parte, también sobre la perspectiva del consentimiento de los pacientes, pero sin preguntarles a ellos sino a los profesionales de la salud, hay estudios que indagan sobre las implicaciones éticas, los peligros, las ventajas y las responsabilidades que implica el que los datos de los pacientes se compartan. En este caso, los estudios profundizan en la opinión de los profesionales sanitarios que tratan día a día con datos de los pacientes. Exploran las percepciones sobre aspectos como las particularidades de compartir datos en formatos concretos, como son los registros médicos electrónicos, o con tecnologías específicas (Grando, Murcko, Kaufman, Whitfield, y Maupin, 2020; Morain, Kass, y Grossmann, 2017; Quinn et al., 2019); las complejidades a las que hacen frente los profesionales de instituciones de salud de países con menos recursos (Anane-sarpong, Wangmo, Leonie, Tanner, y Simone, 2018); o cómo se abordan los conflictos en especialidades concretas, como salud mental, a los que se enfrentan los profesionales entre compartir datos que son relevantes y la reticencia de los pacientes a esta práctica (Grando et al., 2020).

Por otra parte, también hay estudios que recogen una muestra con perfiles más variados, que normalmente incluye lo que denominan “stakeholders” (partes interesadas), y que suele hacer referencia a investigadores, profesionales sanitarios, figuras vinculadas a la industria o altos cargos de la administración (directores de hospitales, figuras relevantes en los departamentos de salud, etc.). En estos estudios, se intenta formar una idea más global de la práctica de compartir datos que va un poco más allá de la protección de datos de los pacientes, como era en los casos comentados en los párrafos anteriores. En estos estudios, aunque también hacen referencia a la confidencialidad de los datos, se abordan aspectos más amplios sobre las tensiones y equilibrios que se generan en torno a la práctica de compartir datos, haciendo balance entre las ventajas y las desventajas. En diferentes combinaciones y con distintos protagonismos, tienen en cuenta los puntos de vista científicos, éticos, legales, políticos, empresariales (sobre todo relacionados con la industria

farmacéutica) y de la sociedad en general (Broes, Lacombe, Verlinden, y Huys, 2017; Colledge, Persson, Elger, y Shaw, 2014; Saunders, Wilhelm, Lee, y Merkhofer, 2014).

Objetivos

- Objetivo general

Indagar acerca del uso compartido de datos brutos entre investigadores e investigadoras del ámbito de Ciencias de la Salud desde una perspectiva cualitativa.

- Objetivos específicos

Objetivo específico 1.

Averiguar qué conocimientos tiene el personal investigador de Ciencias de la Salud sobre el uso compartido de datos (fundamentalmente, dónde se publican los datos, cómo se publican, cuál es el papel de las revistas y cuál es el papel de los repositorios).

Objetivo específico 2.

Averiguar las creencias que tiene el personal investigador de Ciencias de la Salud sobre el uso compartido de datos (ventajas, inconvenientes, miedos y recelos).

Objetivo específico 3.

Indagar acerca de las actitudes que muestra el personal investigador de Ciencias de la Salud sobre el uso compartido de datos (sobre todo, si están dispuestos a compartir, si ya comparten, si podrían condiciones a la hora de compartir, cuáles serían las condiciones).

Objetivo específico 4.

Identificar si hay diferencias en cuanto a conocimientos, creencias y actitudes teniendo en cuenta la jerarquía y los años de experiencia.

Objetivo específico 5.

Examinar si hay diferencias entre diferentes tipos de investigación (básica y aplicada).

Método

1. Origen de este estudio y contextualización de la metodología cualitativa

1.1 Origen del estudio y trabajos previos en el marco del proyecto DATASEA

1.1.1 Origen del estudio

Esta tesis nace dentro del proyecto “Modelos para publicar, consumir y medir la reutilización de datos de investigación: más allá de las fronteras institucionales y geográficas”, cuyo acrónimo es “DATASEA”. Se trata de un proyecto competitivo financiado en el marco del Programa Estatal de I+D+i Orientada a los Retos de la Sociedad dependiente del Ministerio de Economía y Competitividad. Es un proyecto coordinado integrado por investigadores del Valencia International Campus of Excellence (UPV, UV-CSIC). La duración del proyecto ha sido del año 2015 al 2019.

El objetivo general de este proyecto es describir qué modelos hay actualmente de publicación y consumo de datasets (conjuntos de datos), de modo que se pueda saber cómo explorar los métodos adecuados para medir el uso de los datos de investigación. La idea es que DATASEA aporte soluciones en consonancia con las acciones que se están llevando a cabo en este ámbito en el entorno internacional para conocer cómo se descubren y usan los datos. Dentro de este proyecto, se han realizado diversas actividades relacionadas con la investigación, formación y difusión que supone distintas aproximaciones a la temática abordada (el uso compartido de datos de investigación). Entre estas actividades, algunas han influido directamente en la gestación de esta tesis, ya que han servido para evidenciar la necesidad de la misma a la vez que le han dotado de un contexto. A continuación, se explica en qué han consistido estas actividades y de qué manera han influido en el presente estudio.

1.1.2 Estudios complementarios en el marco del proyecto DATASEA

En este apartado se explican los estudios que se han llevado a cabo en el proyecto DATASEA y que sirvieron para dar sentido a la presente tesis. Este “dar sentido” se debe ser entendido desde dos perspectivas. Por una parte, estos estudios han servido para dar contexto a la tesis, un marco de trabajo acotado que ofrece comprensión y amplitud de miras sobre la temática tratada. Por otra parte, evidenciaron cuáles eran las necesidades de información de las que carecía el proyecto para entender desde todas las perspectivas el uso compartido de datos entre investigadores. Cabe mencionar que estos estudios han tenido distintos recorridos con respecto a la tesis. Es decir, no necesariamente se han llevado a cabo antes de la tesis, sino que han podido empezarse antes y continuarse un tiempo una vez la tesis ya había sido empezada, se pudieron haber empezado de manera paralela a la tesis, o se realizaron una vez la tesis ya había sido comenzada.

- **Cuestionario**

Una de las actividades que se ha llevado a cabo dentro del proyecto es la elaboración de una encuesta para identificar los hábitos y experiencias del personal investigador español en relación con la gestión e intercambios de datos brutos, así como sus motivaciones para compartir, las circunstancias que dificultan esta práctica y el soporte técnico y las infraestructuras de las que disponen (Aleixandre-Benavent et al., 2015, 2020). Este estudio se inició y acabó antes que la tesis y fue uno de los principales gérmenes para la elaboración de la misma, ya que comenzó a abordar la temática de compartir datos entre el personal investigador español desde una perspectiva cuantitativa.

La muestra de esta encuesta estuvo formada por 1.063 investigadores de distintas áreas, clasificadas en Ciencias de la Salud, Ciencias Sociales, Artes y Humanidades y Física y Tecnología.

La metodología seguida consistió en la elaboración de un cuestionario de 40 preguntas distribuidas en tres bloques: 1) datos personales (sexo, edad, área, posición); 2)

creación y reutilización de los datos (existencia de planes de gestión de datos, dispositivo de almacenamiento y copia de seguridad, la voluntad de utilizar los datos de otros investigadores, qué datos externos utilizan, si les han pedido datos otros investigadores y problemas o prejuicios para compartir.); y 3) preservación de los datos (prácticas de conservación de datos, sistema de almacenamiento de datos, dispositivos utilizados para su preservación, planes de almacenamiento y preservación y ayudas para llevar a cabo la gestión de los datos).

Con respecto al personal investigador de Ciencias de la Salud, en los resultados de este estudio se halla que, en general, tanto los conocimientos y hábitos sobre infraestructuras para compartir datos como los medios técnicos para el depósito y la conservación son escasos. Sobre los miedos, destacaron especialmente la pérdida de autoría, el que otros investigadores malinterpreten sus datos, aspectos relacionados con la propiedad intelectual y la pérdida del liderazgo en la investigación. Por otro lado, la opción más escogida también por científicos de Ciencias de la Salud con respecto a si sus proyectos cuentan con políticas de datos y por qué, la mayoría responde que, si la tienen, es por obligación de las bases del proyecto. Sin embargo, se desataca que el personal investigador de Ciencias de la Salud es el que mayoritariamente manifiesta aportar datos brutos como material suplementario. Finalmente, mencionar la paradoja que se da en todas las áreas analizadas de que, aunque la mayoría de los encuestados manifiestan estar interesados en consultar y utilizar datos de otros (especialmente las personas más jóvenes), esto no se corresponde con su propia disposición a la hora de compartir datos.

Estos resultados ponen de manifiesto aspectos relacionados con los conocimientos y hábitos de compartir datos y supone una aproximación en términos cuantitativos a las motivaciones del personal investigador en general, y del de Ciencias de la Salud en particular, para llevar a cabo esta práctica.

- **Material suplementario**

En este caso, se ha tratado de una serie de trabajos que fueron diseñados para abordar el uso que el personal investigador de Ciencias de la Salud hacía del material suplementario y la relación que éste podía tener con la práctica de compartir datos.

Se trata de varios estudios que siguen una misma metodología, pero que abordan diferentes áreas dentro de las Ciencias de la Salud. Dichos estudios fueron planificados y empezados antes de que comenzara la tesis y siguieron desarrollándose de manera paralela a la misma.

La finalidad de estos trabajos es averiguar qué se comparte como material suplementario en revistas de varias áreas de Ciencias de Salud y, más concretamente, saber cuánto de ese material podría considerarse datos brutos. Además, algunos de estos estudios incluían también un análisis de las políticas editoriales de las revistas con respecto a compartir datos brutos a través del material suplementario y otras características relacionadas con el Open Science en general. Las áreas de Ciencias de la Salud estudiadas corresponden a 5 categorías temáticas del *Journal Citation Report* (JCR), que son: *Substance Abuse*, *Emergency Medicine*, *Reproductive Biology*, *Cell y Tissue Engineering* y *Dentistry, Oral Surgery y Medicine*. Las categorías fueron escogidas por la variedad que representaban en cuanto a tipo de investigación (básica, clínica), así como por la multidisciplinariedad que tienen dentro de las Ciencias de la Salud las disciplinas que conforman estas áreas (medicina, enfermería, psicología, biología...). Por otro lado, también se tuvo en cuenta la afinidad del grupo de investigación con estas áreas, debido a colaboraciones y proyectos con otros grupos pertenecientes a las mismas, mediante los que se había acumulado experiencia y background.

Sobre el contenido concreto de los artículos y sus principales resultados, desde el año 2014 hasta el año 2021 se han publicado 6 artículos originales y se ha examinado el material suplementario de los artículos de alrededor de 200 revistas de estas 5 categorías indexadas en JCR (Aleixandre-Benavent et al., 2018, 2014; Sixto-Costoya et al., 2020, 2021; Vidal-Infer et al., 2019, 2018). Se ha analizado el material suplementario de los artículos que estaban disponibles en PubMed Central, el repositorio de acceso abierto de la base de datos Medline. Para averiguar si los

artículos tenían material suplementario, se llevaron a cabo dos vías. En primer lugar, para los 5 primeros estudios realizados (Cell y Tissue Engineering, Substance Abuse, Emergency Medicine y Dentistry, Oral Surgery y Medicine) se diseñó una ecuación de búsqueda que permitiese recuperar los documentos que contuviesen algún tipo de material suplementario. Sin embargo, para el último estudio, que hacía referencia al área de Reproductive Biology, no ha sido necesario recurrir a esta ecuación. Esto se debió a que, entretanto, PubMed Central había creado un nuevo filtro dentro del grupo “Article attributes”, denominado “Associated data”, que permite limitar la búsqueda a los documentos que tengan algún tipo de material asociado. La adhesión este filtro también ofrece una medida de cómo la temática de compartir información extra a la que ya está contenida en las publicaciones ha ido cobrando importancia desde que se comenzó esta serie de artículos (publicado en el 2014) hasta el último (publicado en el 2021).

En un segundo paso, una vez obtenidos los artículos con material suplementario, se averiguó qué tipo de material era según el formato y el contenido para clasificarlo y poder saber cuál correspondía con las características de datos brutos (material no procesado y reutilizable para nuevos análisis). En primer lugar, se descubrió que el hecho en sí de adjuntar material suplementario era de menos del 50% en todas las categorías, pero representando una notable diferencia entre Cell y Tissue Engineering (un 43% contenían material suplementario) y el resto de las categorías (todas por debajo del 20%). En cuanto al contenido, de manera general en todas las categorías, los archivos más frecuentemente compartidos eran de texto (por ejemplo, .pdf y .doc/docx) y éstos contenían mayormente figuras o tablas complementarias u otros materiales ya procesados que no “cabían” en el manuscrito principal. Sin embargo, en cuanto a datos brutos, fundamentalmente identificados como hojas en formato tabular (.csv, .xls, .xlsx) o en ocasiones archivos en formatos de imagen o vídeo, las diferencias entre categorías resultaron más acusadas. De nuevo los artículos de las revistas de la categoría Cell y Tissue Engineering son los que más archivos con datos brutos compartían, proseguido por los artículos de la categoría Reproductive Biology, Substance Abuse, Emergency Medicine y Dentistry, Oral Surgery y Medicine. Por otra parte, también de manera general en las 5 categorías, los artículos que más

frecuentemente compartían archivos con datos brutos eran los pertenecían a revistas clasificadas en los cuartiles 1 y 2 de JCR.

El estudio de estas 5 áreas de las Ciencias de la Salud han sido las que se han tomado como referencia en esta tesis para realizar las entrevistas semiestructuradas. El personal investigador se fue seleccionando partiendo de la relación de su investigación con alguna de estas áreas.

Por otro lado, la serie de resultados obtenidos mostraron varios aspectos interesantes sobre las categorías analizadas que sirvieron como inspiración para el objetivo de esta tesis. En primer lugar, que existen tendencias que son similares en las 5 áreas (menos de un 50% de artículos con material suplementario, los que tienen son en su mayoría datos ya procesados, y los que tienen datos brutos están predominante en revistas clasificadas en cuartiles 1 y 2 de JCR), y tendencias diferenciadas (cuanto más relacionada esté la categoría con la parte clínica, menos datos brutos se comparte y cuánto más lo esté con la investigación básica más se comparte). En el apartado de 2.1 “El universo y la muestra” se define con más profundidad la influencia de estos trabajos en la presente tesis.

- **ORCID como recurso para estudiar la práctica de compartir datos**

En otra aproximación para conocer el alcance del uso compartido de datos entre investigadores, se realizó un estudio a través del recurso ORCID (Sixto-Costoya et al., 2021). Este trabajo consistió en una aproximación metodológica acerca de cómo ORCID puede ser útil para explorar los datasets que el personal investigador comparte a través de esta plataforma. La realización de este estudio se llevó a cabo durante una estancia de investigación en el Center for Science and Technology Studies (CWTS) en la University of Leiden (Países Bajos).

La decisión de utilizar ORCID se basó en la utilidad que este recurso tiene a la hora de identificar tanto a los autores y autoras como los trabajos de los mismos. Aunque, como infraestructura, ORCID fue creada con el fin de resolver el problema de ambigüedades en los nombres de los autores y autoras, a medida que la herramienta va desarrollándose, cada vez añade más servicios relacionados con las actividades

académicas del personal investigador. Una de las actividades que ORCID actualmente recoge es la publicación de diferentes productos de investigación, entre ellos los datasets. Teniendo en cuenta esta funcionalidad, el objetivo del estudio fue utilizar esta identificación que ORCID realiza de los datasets que el personal investigador comparte en la plataforma para estudiar tanto las características de los datasets como de las personas que los producen.

Para ello, se obtuvieron todos los perfiles de ORCID que reportasen al menos un dataset en su lista de trabajos, “works”, junto con la información de los investigadores que producían esos datasets. Los datos fueron obtenidos a través del archivo que ORCID comparte de manera pública en el repositorio de datos Figshare. Los datos fueron posteriormente organizados y analizados en una base de datos SQL alojada en el CWTS.

Los objetivos concretos de este estudio fueron:

- Cuantificar el número de datasets reportados por los investigadores en ORCID.
- Caracterizarlos según el repositorio de datos donde hayan sido depositados y la fuente desde donde ORCID ha obtenido la información sobre este depósito. Para los repositorios, se tienen en cuenta los repositorios generalistas Zendodo, Figshare y Dataverse, dada la posibilidad de localizar los datasets depositados en estos repositorios a través de su doi.
- Estudiar el impacto del data sharing teniendo en cuenta el área de la ciencia a la que pertenece el personal investigador (mediante el doi de sus artículos en ORCID con la Web of Science), los países donde realizan su trabajo (información que se obtiene de la breve descripción biográfica que aparece al principio de cada perfil de ORCID), y la proximidad o lejanía que existe desde la publicación de su tesis doctoral hasta la publicación de su primer dataset.

Teniendo en cuenta las dimensiones estudiadas, los resultados señalaron que DataCite es con mucha diferencia la fuente desde la que ORCID obtiene más información sobre los datasets. DataCite tiene la posibilidad de obtener los datos desde varios recursos, entre ellos los tres repositorios estudiados: Zenodo, Dataverse y Figshare. De los tres repositorios, Figshare es el más utilizado. Sobre el área de la ciencia a la que el

personal investigador pertenece, se ha encontrado que la más común es Ciencias Naturales, seguida muy de cerca por las Ciencias de la Vida, entre ambas acumulan el 85,6% de las publicaciones. En cuanto a los países, los anglosajones como Estados Unidos, Canadá, Australia y Reino Unido son los que ha aparecido con más frecuencia en los perfiles biográficos en ORCID que tenían como mínimo un dataset. Finalmente, con respecto a la edad académica (medida según la lejanía de publicación de su primer dataset después de finalizar la tesis), se encuentra que cuánto recientemente hayan publicado sus tesis, más pronto publicarán su primer dataset, lo que podría apuntar a una mayor asimilación de la práctica en las generaciones más nuevas de investigadores e investigadoras.

Como se ha dicho al principio de este apartado, estos estudios han servido para acompañar y dar sentido y contexto a esta tesis, pero también para evidenciar las necesidades que existían de información. Todos los estudios explicados ofrecieron perspectivas muy útiles y mayormente cuantitativas para abordar la comprensión del fenómeno de compartir datos entre investigadores e investigadoras de Ciencias de la Salud. Sin embargo, casi inevitablemente los resultados obtenidos en estos estudios conducen a la siguiente pregunta: más allá de los números, ¿por qué las personas entrevistadas han respondido a las cuestiones de una determinada manera y no de otra?, a pesar de que se ha evidenciado en ORCID que las Ciencias de la Vida son de las que más comparten, ¿por qué los investigadores de ciencia básica comparten más que quienes tienen investigaciones más aplicadas? Estos “por qué” son frecuentes y se surgen en todos los estudios que se han ido realizando en el marco del proyecto DATASEA. Por ejemplo, en el caso de la encuesta, con sus resultados se puede saber que existe poco conocimiento sobre las infraestructuras para compartir datos o a cuántos investigadores les preocupa que a la hora de compartir datos en un servicio externo se pueda hacer una mala interpretación de los mismos. Pero no se puede profundizar acerca del por qué tienen poco conocimiento acerca de las infraestructuras, o por qué han llegado a tener determinadas reticencias. Es decir, con la encuesta y sus resultados se ha obtenido una vista aérea que permite cuantificar aspectos relacionados con la gestión que lleva a cabo el personal investigador de sus datos de investigación, pero no se puede acceder (por las propias características del

método cuantitativo) a una aproximación que analice en profundidad por qué el personal investigador escoge unas opciones del cuestionario y no otras. Lo mismo sucede con el estudio del material suplementario y también de ORCID. Se pueden obtener visiones generales que aportan información muy valiosa, pero surge la necesidad de más profundidad.

Debido a ese vacío, a esa necesidad, se decide poner en marcha una metodología que permita colocar al investigador en el centro, con toda su subjetividad como persona que forma parte de un grupo, la comunidad científica, y también como individuo que forma parte de la sociedad, e indagar acerca de sus percepciones sobre el uso compartido de datos, profundizando en sus experiencias vitales como profesional. Esa aproximación permite, en vez de obtener una vista área, focalizar en el aspecto concreto de lo que el investigador piensa y siente sobre el uso compartido, es decir, permite, poniendo el símil de una cámara de fotos, acercar mucho más la lente. Para ello, se ha determinado que la metodología cualitativa es la más acertada, ya que es la única capaz de responder estos interrogantes planteados. A continuación, se contextualiza la metodología cualitativa, para posteriormente proceder a explicar el uso de esta metodología en el presente trabajo.

1.2 Contextualización de la metodología cualitativa

1.2.1 Definición

La metodología cualitativa es una de las dos metodologías de investigación que tradicionalmente se han utilizado en ciencia. Cuando se buscan definiciones que expliquen qué es o en qué consiste la metodología cualitativa en la literatura científica, prácticamente siempre se utiliza su contraste con la cuantitativa (Ruiz Olabuénaga, 2012; Valles, 2000). Por ello, lo primero que se debe mencionar sobre la metodología cualitativa para su comprensión es que, al contrario que la metodología cuantitativa, ésta se centra en aspectos no susceptibles de cuantificación.

De esta manera, mientras la investigación cuantitativa se focaliza en los datos numéricos, los datos en la investigación cualitativa no se presentan en forma de números. El tipo de datos cualitativos se caracterizan por ser capaces de captar las creencias subyacentes de las personas entrevistadas, sus opiniones o sus motivaciones, así como describir sus comportamientos y experiencias vitales. En este sentido, la investigación cualitativa genera una comprensión detallada y en profundidad de fenómenos o experiencias que permiten responder a cuestiones que comienzan por “cómo”, “por qué” o “qué” con respuestas más allá del “sí” o “no”. Se suele utilizar para explorar experiencias, percepciones o preferencias en una determinada población, o entender mejor ciertos comportamientos y cuáles son los factores que influyen en ellos. Aunque la investigación cualitativa típicamente no es generalizable a toda la población, puede ser muy útil para diseñar y focalizar determinadas actuaciones en un contexto particular con una población específica (Tuckerman, Kaufman, y Danchin, 2020).

1.2.2 Investigación cualitativa y cuantitativa: diferencias y sinergias

Dado que, como se menciona en el apartado anterior, se suele mencionar la metodología cuantitativa para comprender mejor el objetivo y las delimitaciones de la cualitativa, en este punto se profundiza un poco más sobre las principales diferencias entre ambas metodologías.

En cuanto a los planteamientos y las teorías que sustentan cada una, normalmente la metodología cuantitativa se asocia con el paradigma positivista mientras que el cualitativo se relaciona con el humanista. A este respecto, el paradigma positivista defiende una investigación que se concentra en la descripción y la explicación, donde los focos de estudio son más generales, centrados en la estructura, deductivos (de lo general a lo específico) y causales. Por otro lado, el humanista opta por objetos de estudio inductivos (de lo específico a lo general), más centrados en la interpretación y en la comprensión de lo humano (Plummer, 1983). En este sentido, la investigación cualitativa acentúa el aspecto holístico y dinámico de la experiencia humana e intenta

comprenderla dentro de un todo para entender las experiencias subjetivas desde el punto de vista de las personas que la viven. Por ello, este paradigma no se centra en la predicción y control de un fenómeno (estudio de variables), sino en la descripción sistemática de la experiencia cotidiana (estudio de las personas) (Vivar, Arantzamendi, López-Dicastillo, y Gordo Luis, 2010).

En la investigación sobre fenómenos sociales, el uso de los métodos y técnicas cuantitativas y cualitativas (por separado o combinadas) es común a lo largo de la historia. Sin embargo, no siempre ha existido una convivencia pacífica entre ambas metodologías. Una de las causas más destacadas esta fricción es el hecho de que ha existido un cierto monopolio de lo cuantitativo sobre lo cualitativo en la investigación y un cruce de críticas mutuas entre investigadores que optan por una o la otra metodología. En este debate centrado en la polémica cantidad-cualidad se manifiesta, desde la posición cuantitativa, que la metodología cualitativa flaquea en aspectos relacionados con la objetividad y la neutralidad, así como en la falta de precisión y replicabilidad. Por su parte, los defensores del método cualitativo achacan al cuantitativo aspectos como que los métodos matemáticos, cada vez más sofisticados, generan una percepción de lejanía de la realidad social, así como la que la complejidad relacional de sus análisis dificulta la orientación práctica a la toma de decisiones o que existe una predominancia de estudios “ex post facto” (validación de la hipótesis cuando el fenómeno ya ha ocurrido), por encima de los que sugieren claves de realidades sociales de cara al futuro (Ruiz Olabuénaga, 2012; Van Maanen, 1983).

Sin embargo, a pesar de estas fricciones, lo cierto es que a partir de los años 80 se asiste a una situación de mayor complementariedad y acercamiento de ambas posiciones. En este sentido, las dos perspectivas han llevado a cabo un proceso de reconocimiento mutuo y convergencia (Valles, 2000). A este respecto, destacan las perspectivas conciliadoras como la de Cook y Reinhardt (1979), que defienden que los que ven el debate en términos de un contraste entre paradigmas suelen proporcionar toda una lista de atributos que implican que cada método está inevitablemente unido a un paradigma y que estos son una especie de compartimentos estanco. Es decir, que como el personal investigador que aboga por uno u otro método ven el mundo de distintas formas, ha de emplear sí o sí distintos métodos de investigación y que,

además, supone que los paradigmas cualitativo y cuantitativo son tan rígidos que solo se puede elegir uno de ellos. En opinión de estos autores, es un error promover la incompatibilidad de estos dos métodos y resaltan que existen beneficios del empleo conjunto de ambos.

Dado que el foco de investigación de esta tesis y la necesidad de la misma parte de otros trabajos de tipo cuantitativo (Aleixandre-Benavent et al., 2014; Aleixandre-Benavent et al., 2015, 2018, 2020; Sixto-Costoya et al., 2020; Vidal-Infer et al., 2018, 2019), esta última perspectiva conciliadora es la que se ha tenido en cuenta para la elaboración de este estudio.

1.2.3 Breve recorrido histórico de la metodología cualitativa y aproximación a su multidisciplinariedad

Atendiendo a la literatura existente, no hay un relato consensuado sobre el origen de la perspectiva metodológica cualitativa. Sin embargo, algunos autores han hecho el trabajo de intentar sintetizar una línea cronológica que aporte luz sobre su recorrido histórico, centrada especialmente en los campos de la Sociología y la Antropología del siglo XX. Poniendo como punto de partida los inicios del siglo XX, se sitúa como comienzo el momento en que algunos etnógrafos pertenecientes a la denominada etnografía clásica, empiezan a desarrollar un trabajo de campo sistemático, que consistía en permanecer un determinado periodo de tiempo recolectando datos primarios que permitieran una comprensión detallada del conjunto de sujetos estudiados y sus relaciones cotidianas en su medio naturales. También en los inicios del siglo XX, destaca el trabajo cualitativo de la Escuela de Chicago con las historias de vida, que sentaron las bases de la metodología interpretativa. En una etapa posterior, que se sitúa a mediados del siglo XX, se inscribe el llamado “periodo modernista” también llamado “la edad dorada del análisis cualitativo riguroso”, donde tuvo lugar un gran esfuerzo de formalización y sistematización, que pusieron énfasis en la validez interna y externa en los modelos interaccionista y constructivista. Entre mediados y finales del siglo XX se darían dos momentos interconectados, donde el personal

investigador de Ciencias Sociales busca nuevas teorías y métodos de análisis con una gran variedad de enfoques novedosos. Se trata de un periodo dominado por perspectivas más amplias y plurales (neomarxismo, etnometodología, deconstruccionismo...), pero también por abogar por un modo de investigación tenga en cuenta cuestiones hasta ahora descuidadas como el género, la raza y la clase social del propio investigador. La última etapa, que bebe de todas las anteriores, comienza a finales de los 90 sigue vigente en la actualidad y destaca por la gran opcionalidad de paradigmas, así como de estrategias de investigación y de métodos. Se trata de un momento que se diferencia por el redescubrimiento de las maneras de investigar cualitativas, donde una de las conclusiones es que este tipo de investigación no puede enfocarse desde una perspectiva positivista y neutra, como es el caso de la metodología cuantitativa (Valles, 2000).

Sin embargo, además de la Antropología y Sociología, también hay autores que sitúan los inicios de las técnicas cualitativas en Europa a finales del siglo XIX con autores como Frédéric Le Play, que dedicaron buena parte de su carrera a estudiar las condiciones de vida de la clase obrera en distintos países de Europa. Un detalle diferenciador importante sobre los trabajos de Le Play es que no tenían el objetivo principal de generar teoría (como más tarde sucedería con los estudios antropológicos o sociológicos), sino que tenían una finalidad práctica inmediata de resolución de problemas. Le Play, de origen ingeniero de minas, se dedicó a investigar a las familias como unidad que él consideraba fundamental para el bienestar de un Estado, es decir, partía de la base de que la situación en la que se encontraran las familias (salud, seguridad, trabajo...) impactaría directamente en el conjunto de un país. Por ello, mediante la observación directa, las entrevistas en profundidad semiestructuradas y un trabajo exhaustivo de sistematización y codificación, se dedicó a estudiar las familias obreras y su modo de vida, con el fin de elevar posteriormente sus descubrimientos a instancias políticas y económicas y conseguir un cambio social que trajera paz y prosperidad (Forni, Frey, y Quaranta, 2012).

En la misma línea de Le Play, cabe destacar que otras disciplinas no necesariamente ligadas al ámbito de la Ciencias Sociales y Humanidades, han reparado en que la metodología cualitativa podía dar respuesta a algunas de sus necesidades e

interrogantes. De esta manera, existen numerosos ejemplos recientes provenientes de campos tan diversos como la biomedicina y las ingenierías que usan los métodos cualitativos para fines prácticos de resolución de necesidades. Los objetivos de estos estudios suelen girar en torno a querer conocer algún tipo de trasfondo al que los datos cuantitativos no pueden acceder. Por ejemplo, el conocimiento del personal sanitario sobre las percepciones y experiencias de los pacientes sobre algún tema, la utilidad de averiguar qué piensan los profesionales de la ingeniería de determinados procesos para mejorar el funcionamiento, o, en el caso de profesionales de la documentación, saber por qué sus usuarios tienen unas determinadas necesidades de información y no otras (Hunt y Bakker, 2018; Sandelowski y Leeman, 2012; Szajnfarber y Gralla, 2017)

Haciendo un paralelismo con el trabajo de Le Play, esta tesis se propuso también utilizar los métodos cualitativos no solo para generar teoría, sino para hacer propuestas que permitan una resolución de problemas / necesidades en el tema que ocupa a esta investigación. De esta manera, del mismo modo que Le Play focalizó sus estudios en la familia como unidad esencial para el buen funcionamiento de una sociedad, esta tesis pone el foco en el personal investigador, en este caso concreto de Ciencias de la Salud, como unidad fundamental para el funcionamiento y “bienestar” de la ciencia. Comprendiendo al personal investigador de manera profunda, tal y como permiten las técnicas cualitativas, se pretender inducir sobre qué funciona y qué no en la práctica de compartir datos, cual es el alcance de la misma según los parámetros políticos y económicos actuales (financiación de ciencia, precariedad del personal investigador, presión por publicar más cantidad que calidad...) y de qué manera se podrían hacer propuestas para que esta práctica se ajuste realmente al día a día del personal investigador.

2. Métodos y técnicas cualitativas utilizadas en este estudio

En este apartado se describen las características del método utilizado en este estudio: universo y muestra, la recogida de información y cómo se ha realizado el análisis de los datos.

2.1 El universo y la muestra

2.1.1 El universo

El universo de este estudio son el personal investigador de Ciencias de la Salud.

Concretamente, el personal investigador de Ciencias de la Salud relacionados con las siguientes áreas: *Substance Abuse, Emergency Medicine, Reproductive Biology, Cell & Tissue Engineering y Dentistry, Oral Surgery & Medicine*.

La denominación de estas áreas es la utilizada por la herramienta Journal Citation Reports incluida en la plataforma Web of Science (WOS) de la empresa Clarivate Analytics. La clasificación distribuye en distintas categorías temáticas las diferentes revistas indexadas en JCR por áreas de conocimiento. Aunque la denominación en inglés es la que se utilizó de manera predominante en este trabajo, ya que es la oficial, a continuación, se presenta la traducción al español de cada una de ellas, así como una breve descripción basada en la que ofrece JCR para estas áreas (Clarivate Analytics, 2021):

1. Substance Abuse: Abuso de Sustancias

Esta categoría tiene la particularidad de acoger revistas indexadas tanto en el Social Sciences Citation Index como del Sciences Citation Index, debido a su carácter multidisciplinar. Por la parte social, abarca recursos relacionados principalmente con los problemas sociales y psicológicos de la adicción, la educación sobre el abuso de sustancias y el tratamiento de los dependientes químicos; así como los problemas concernientes a la parte biomédica del abuso de sustancias.

2. Emergency Medicine: Medicina de Emergencias, Medicina de Urgencias

La cobertura de esta categoría abarca recursos sobre la ciencia, la educación y la práctica clínica de la medicina de urgencias. Además, la cobertura abarca la amplitud de la especialidad en trauma, pediatría, toxicología, prevención y control de lesiones, reanimación y servicios médicos de emergencia. Cabe mencionar que, actualmente, no existe en España la especialidad de Medicina de Urgencias. Por lo tanto, se debió recurrir a profesionales de otras especialidades que ocupasen frecuentemente puestos destinados a esta área de intervención.

3. Reproductive Biology: Biología Reproductiva

En este caso, la categoría incluye recursos que tienen que ver con la reproducción en humanos, animales y plantas. Esta categoría abarca desde la biología molecular de la reproducción hasta la nutrición reproductiva, la inmunología y la toxicología. La reproducción de animales y plantas no se ha tenido en cuenta en este estudio.

4. Cell & Tissue Engineering: Ingeniería celular y de tejidos

Esta categoría comprende los recursos que aplican la tecnología a la alteración o preparación química, mecánica y eléctrica de células y tejidos. Los recursos de esta categoría abarcan una serie de tecnologías que incluyen: la regeneración por ingeniería del tejido nervioso, adiposo o endotelial; el desarrollo de varios tipos de células madre por ingeniería; y los aspectos biomecánicos de los tejidos/células por ingeniería. Cabe destacar que esta categoría no incluye recursos sobre intervenciones clínicas.

5. Dentistry, Oral Surgery & Medicine: Odontología, Cirugía Oral y Medicina

Esta categoría cubre recursos sobre la anatomía, fisiología, bioquímica y patología de los dientes y la cavidad oral. Además, incluye recursos específicos sobre la enfermedad periodontal, los implantes dentales, la cirugía oral y maxilofacial, la patología oral y la cirugía oral, así como la odontología comunitaria, de salud pública y pediátrica.

Para este estudio, se ha tomado como punto de partida que el personal investigador estuvieran relacionados con estas cinco áreas de Ciencia de la Salud por las razones que se detallan a continuación:

- Variedad en los tipos de investigación (básicas y aplicadas) tanto entre las distintas áreas (por ejemplo, emergencias y odontología, más aplicadas y células madre, más básica) como dentro de las mismas (en biología reproductiva se puede encontrar en proporción semejante tanto básica como aplicada).
- La multidisciplinariedad. En el caso de Adicciones, pertenece tanto a Ciencias de la Salud como Ciencias Sociales, por lo que incluye a una gama amplia de profesionales.
- La proximidad con otras investigaciones y proyectos del grupo. Existe un expertise en estas áreas obtenido en trabajos y colaboraciones que permiten partir de un conocimiento sobre las mismas.
- Algunos de estos trabajos y colaboraciones se enmarcan dentro de la temática del uso compartido de datos, pero desde otras perspectivas. Por ello, suponen un buen punto de partida para localizar posibles candidatos a los que entrevistar.

Por otro lado, se ha decidido que residan en España (aunque no necesariamente deben ser españoles) por la facilidad de acceso para poder conseguir, en el mayor número de entrevistas posible, la presencialidad. Ésta se ha considerado importante para poder captar de manera directa todos los matices de comunicación verbal y no verbal.

Además, se ha estimado necesaria la variedad en cuanto a años de experiencia. Esto fue así ya que se considera que dependiendo de los años realizando el trabajo de investigación, las vivencias y la perspectiva podrían cambiar. Teniendo en cuenta que más años de experiencia normalmente se relacionan también con más jerarquía en cuanto al poder de toma de decisiones, el factor años de experiencia es más relevante si cabe.

Por último, se ha partido de la base de que en Ciencias de la Salud (igual que sucede en otras áreas) investigar no significa necesariamente ser investigador. Es decir, que

existe la posibilidad de que las personas que llevan a cabo investigaciones no sean exclusivamente profesionales a los que se les remunera únicamente por investigar. De este modo, es frecuente que el personal sanitario lleve a cabo investigaciones, a veces promovidas conducidas en el mismo centro de trabajo (hospitales, centros de salud), y otras mediante convenios y/o colaboraciones con otras instituciones como universidades o centros de investigación. Además, se ha considerado que existen investigadores que solo se dedican a investigar y otros que lo combinan con docencia, como suele ser el caso del personal docente e investigador de las universidades. Por otro lado, se ha tenido en cuenta que la investigación a veces es financiada con dinero público, otras con dinero privado, y otras con una combinación de ambos. Del mismo modo, los profesionales pueden estar contratados por instituciones públicas y privadas. Todas estas combinaciones y realidades se tuvieron en cuenta a la hora de conformar la muestra.

Con respecto a la práctica de compartir datos, se ha partido de la posibilidad de que todas las personas entrevistadas eran potencialmente conocedores de la práctica y bien podían haberla llevado a cabo, solo conocerla, o ambas. No obstante, se considera la necesidad de aportar una definición formal de qué son los datos de investigación y a qué nos referimos cuando hablamos de la práctica de compartirlos.

2.1.2 La muestra

Para este trabajo se ha escogido un tipo de muestreo intencional. Este tipo de muestreo no especifica de antemano el número mínimo de unidades que se van a seleccionar, ya que se asume que el número podrá verse alterado a lo largo de la investigación. De esta manera, se abre la posibilidad de que, por un lado, puedan seleccionarse unidades no previstas inicialmente con intención de mejorar la calidad de la información y, por otro lado, pueda interrumpirse la selección de más unidades cuando se entienda que se ha llegado a un punto de saturación. Esta saturación se consigue cuando el investigador (que está recogiendo la información al mismo tiempo que analizándola) considera que los nuevos datos comienzan a ser repetitivos y dejan

de aportar información novedosa. Por lo tanto, el criterio de confianza que se ha elegido para garantizar la representatividad de la muestra es el criterio intencional.

Siguiendo esta premisa, se han seleccionado 23 unidades de muestra (investigadores) no al azar ni siguiendo un cálculo o ley de probabilidades, sino por el método opinático. Según el método opinático, se seleccionaron los integrantes que van a componer la muestra con el siguiente criterio estratégico: las personas (investigadores) que por su conocimiento de la situación a investigar son las más idóneas de la población a estudiar y las personas que entran en contacto a través de sujetos entrevistados previamente (técnica snowball o de bola de nieve). La muestra obtenida se ha orientado a la selección de aquellas unidades y dimensiones que mejor garanticen: la cantidad (saturación) y la calidad (riqueza de la información) (Ruiz Olabuénaga, 2012).

Los participantes fueron escogidos teniendo en cuenta la variedad de responsabilidades, incluyendo diferentes grados de responsabilidad en cargos de gestión y liderazgo en proyectos de investigación, y los años de experiencia en investigación. Con respecto a esto último, los años de experiencia, se realiza una división entre júnior y sénior, donde júnior sería el personal investigador con menos de 10 años de experiencia y sénior los que tienen 10 o más. Asimismo, se han tenido en cuenta tanto instituciones públicas (universidades, hospitales) como privadas (instituciones y empresas líderes en investigación). Se ha procurado obtener una variedad equilibrada entre perfiles con un historial de investigación tanto básica (investigación sin fines prácticos inmediatos, sino con el propósito de incrementar el conocimiento en determinadas materias, que tiene en el laboratorio uno de sus espacios de mayor desarrollo) como aplicada (investigación que tiene el propósito de generar conocimiento con aplicación directa en la sociedad o en el sector productivo). Además, se ha procurado un equilibrio en el sexo de los participantes.

Tras una primera oleada de entrevistas, se identificaron encuestados adicionales mediante muestreo de bola de nieve. En la Tablas 1 se pueden observar el total de las entrevistas teniendo en cuenta las características de las personas entrevistadas según las variables: 1) tipo de investigación, 2) lugar de trabajo, 3) años de experiencia y 4) área de investigación. En la Tabla 2 se presenta la identificación de cada entrevistado

con una etiqueta. Estas etiquetas serán las utilizadas en los resultados para referirse a cada entrevistado.

VARIABLE	CATEGORÍA	N
Tipo de investigación	Básica	12
	Aplicada	10
	Ambas (sin predominio)	1
Lugar de trabajo	Universidad	7
	Hospital/Centro de Salud	4
	Empresa/ Fundación privada	8
	Centro de Investigación público	4
Años de experiencia	Sénior (+ de 10)	12
	Júnior (- de 10)	11
Sexo	Hombre	11
	Mujer	12
Área de investigación	Adicciones/ Salud Mental	9
	Odontología	2
	Medicina Regenerativa	5
	Biología Reproductiva	4
	Emergencias	3

Tabla 1. Características de las personas entrevistadas (N=23)

ETIQUETA	SEXO	CARGO	AREA	EXPERIENCIA
E001	F	Profesora titular de Universidad pública	Investigación básica en áreas relacionadas con las adicciones	Más de 10 años de experiencia
E002	F	Investigadora posdoctoral en Universidad pública	Investigación básica en áreas relacionadas con las adicciones	Menos de 10 años de experiencia
E003	F	Investigadora sénior y Jefe de laboratorio en Fundación privada	Investigación básica en áreas relacionadas con las adicciones	Más de 10 años de experiencia
E004	M	Médico especialista en Salud Mental en Hospital público	Investigación aplicada en áreas relacionadas con las adicciones y la salud mental	Más de 10 años de experiencia

ETIQUETA	SEXO	CARGO	AREA	EXPERIENCIA
E005	F	Investigadora posdoctoral en Universidad pública	Investigación básica en áreas relacionadas con las adicciones	Menos de 10 años de experiencia
E006	F	Investigadora sénior en Fundación privada	Investigación básica en áreas relacionadas con la medicina regenerativa	Más de 10 años de experiencia
E007	F	Investigadora sénior en Instituto de Investigación público	Investigación aplicada en áreas relacionadas con las adicciones	Más de 10 años de experiencia
E008	M	Profesor Ayudante doctor de Universidad pública	Investigación aplicada en áreas relacionadas con la odontología	Menos de 10 años de experiencia
E009	F	Investigadora predoctoral en Instituto de Investigación público	Investigación básica en áreas relacionadas con la medicina regenerativa	Menos de 10 años de experiencia
E010	M	Investigador predoctoral en Instituto de Investigación público	Investigación básica en áreas relacionadas con la medicina regenerativa	Menos de 10 años de experiencia
E011	F	Investigadora sénior y Jefe de laboratorio en Fundación privada	Investigación básica en áreas relacionadas con la medicina regenerativa	Más de 10 años de experiencia
E012	M	Investigador sénior y Jefe de laboratorio en Fundación privada	Investigación básica en áreas relacionadas con la medicina regenerativa	Más de 10 años de experiencia
E013	F	Investigadora posdoctoral en Instituto de Investigación público	Investigación aplicada en áreas relacionadas con las adicciones y salud mental	Menos de 10 años de experiencia
E014	M	Investigador posdoctoral en Fundación privada	Investigación básica en áreas relacionadas con las adicciones	Menos de 10 años de experiencia
E015	F	Investigadora posdoctoral en Instituto de Investigación público	Investigación aplicada en áreas relacionadas con las adicciones	Menos de 10 años de experiencia
E016	M	Catedrático de Universidad pública	Investigación aplicada en áreas relacionadas con la odontología	Más de 10 años de experiencia

ETIQUETA	SEXO	CARGO	AREA	EXPERIENCIA
E017	M	Director científico de empresa privada	Investigación aplicada y básica en áreas relacionadas con la biología reproductiva	Más de 10 años de experiencia
E018	M	Investigador sénior en empresa privada	Investigación aplicada y básica en áreas relacionadas con la biología reproductiva	Más de 10 años de experiencia
E019	F	Investigadora posdoctoral en Universidad pública	Investigación aplicada en áreas relacionadas con la biología reproductiva	Menos de 10 años de experiencia
E020	F	Investigadora posdoctoral en empresa privada	Investigación básica en áreas relacionadas con la biología reproductiva	Menos de 10 años de experiencia
E021	M	Médico de Urgencias en Hospital público	Investigación aplicada en áreas relacionadas con las emergencias médicas	Más de 10 años de experiencia
E022	M	Médico Internista en Hospital público	Investigación aplicada en áreas relacionadas con las emergencias médicas	Más de 10 años de experiencia
E023	M	Médico Interno Residente Internista en Hospital público	Investigación aplicada en áreas relacionadas con las emergencias médicas	Menos de 10 años de experiencia

Tabla 2. Descripción de cada entrevistado e identificación con una etiqueta

2.2 Recogida de información

2.2.1 La entrevista en profundidad

Para la recogida de información se ha utilizado la técnica de la entrevista en profundidad, mediante la cual se lleva a cabo una conversación profesional con una o más personas para un estudio de investigación. El objetivo de esta técnica es conseguir que los individuos entrevistados transmitan al investigador su definición personal de la situación. De esta manera, la entrevista se convierte en un relato narrado por la

persona desde su punto de vista y, por ese motivo, el momento de la entrevista es también el contexto en el que se lleva a cabo ese relato, ya que se crea la situación social para que esa narración pueda tener lugar. En este contexto creado, el papel del entrevistador es el de facilitar y transmitir la entrevista. Es una situación en la que el entrevistado es preguntado sobre una serie de temas diseñados para obtener información que será interpretada mediante un esquema previamente definido. Se diferencia del proceso conversacional en el punto en que tiene una intencionalidad científica, ya que va dirigida a la obtención de datos para una investigación y las personas entrevistadas desarrollan papeles orientados en función del diseño de investigación establecido (Bell, 1993; Fernández Esquinas, 1998; Ibáñez, 1986).

Dentro de la entrevista en profundidad, se estimó la pertinencia de que tuviese una orientación de entrevista de tipo focalizada, ya que su carácter de especificidad se ajusta a los objetivos de la tesis. El rasgo más destacado de la entrevista focalizada es la concreción, mediante la que se busca un núcleo o foco de interés. Por este motivo, va dirigido a un tipo de entrevistado concreto, caracterizado por formar parte de ese núcleo de interés (Fiske, Kendall, y Merton, 2002; Ruíz Olabuénaga, 2012). A continuación, se relatan las características que definen la entrevista focalizada y su relación con la tesis:

- La persona entrevistada es un sujeto de quien se sabe que ha podido tener contacto con la práctica de compartir datos de investigación.
- La entrevistadora conoce de antemano la situación y la ha analizado sistemáticamente. Esto se ha logrado a través de la revisión bibliográfica y la experiencia previa del proyecto Datasea.
- La entrevista se concentra en la experiencia subjetiva de los sujetos escogidos con el objetivo de lograr su propia definición de la situación.
- La entrevista está abierta a la posibilidad de que las respuestas o los temas imprevistos puedan dar pie a nuevas hipótesis e interpretaciones de la experiencia.

Siguiendo los criterios de la entrevista en profundidad focalizada, el conocimiento de antemano de la situación, el análisis previo del estado de la cuestión y los perfiles de entrevistados que se necesitan, el siguiente paso consiste en el desarrollo de un guion de entrevista (guion en el Anexo 2).

2.2.2 Las preguntas y respuestas

Las preguntas han sido abiertas, sin categorías de respuestas preestablecidas. Fueron grabadas conforme a un sistema de codificación flexible y abierto a cambios en todo momento. Se ha buscado que la entrevista adopte un talante de conversación, no directivo, acercándose al mundo simbólico del entrevistado de forma escalonada y gradual, comenzando por cuestiones amplias que se van acotando, aproximándose al núcleo de la experiencia poco a poco.

Las preguntas se van encadenado por temas, de manera que, sin perder el espíritu de conversación en la que la persona entrevistada se pueda sentir cómodo, se va orientando de manera que se pueda ir tocando todas las temáticas que realmente interesan para el objeto de la investigación. A medida que la conversación progresa, el interés se irá desplazando de los conocimientos y las descripciones que el entrevistado pueda hacer para focalizarse más en los significados, los sentimientos expresados y las interpretaciones.

En hecho de que las preguntas abiertas formen parte de la estrategia predominante no significa que el entrevistador no necesite acudir a las preguntas cerradas para comprobar y verificar determinadas afirmaciones y datos, o para confirmar que alguna cuestión se ha comprendido correctamente. El optar por preguntas abiertas o cerradas, emitir alguna opinión u optar por el silencio está guiado por el criterio de obtener la máxima cantidad de riqueza informativa en cada momento de la entrevista (Ruíz Olabuénaga, 2012).

2.2 Dimensiones de análisis

Las dimensiones de análisis de este estudio se inspiraron en los resultados de las investigaciones cuantitativas previas del proyecto DATASEA, especialmente las encuestas (Aleixandre-Benavent et al., 2015, 2020).

En base a esto, se han escogido como dimensiones de análisis tres conceptos que han sido ampliamente utilizados en la literatura para abordar temas relacionados con el estudio de diferentes aspectos de la cultura: los conocimientos, las actitudes y las creencias. Concretamente, en el ámbito de las Ciencias de la Salud estas dimensiones fueron utilizadas para llevar a cabo estudios sobre diversas temáticas cuando el objetivo era realizar aproximaciones a distintas realidades de una manera más específica. Por ejemplo, se han hallado estudios relacionados con temáticas de cáncer (Holroyd, Taylor-Piliae, & Twinn, 2003; Opoku, Benwell, & Yarney, 2012), VIH o VPH (Fernandez et al., 2009; Petro-Nustas, Kulwicki, & Zumout, 2002), o temas de salud pública como los aspectos nocivos del tabaco (Ferrante et al., 2013). Estos estudios, que pueden estar centrados tanto en la perspectiva de los pacientes como de los profesionales de la salud, buscan ofrecer una perspectiva cercana a la realidad subjetiva de las personas, ya que tienen en común ir dirigidos a grupos poblacionales muy específicos e indagar sobre sus particularidades. Por este motivo, son útiles para obtener visiones más holísticas, en este caso, de temáticas relacionadas con la salud.

A continuación, se explica con detalle en qué consiste cada una de las tres dimensiones y cómo han sido adaptadas para abordaje del estudio de los conocimientos, actitudes y creencias de los investigadores en Ciencias de la Salud sobre el uso compartido de datos.

- Los conocimientos y experiencias del personal investigador sobre el uso compartido de datos.

Esta dimensión pretende abordar qué conocimientos y experiencias ha tenido el personal investigador con la práctica de compartir datos, lo que también implica indagar acerca de las destrezas de los mismos sobre esta práctica.

A este respecto, se han relacionado los conceptos de conocimientos y experiencias, ya que se parte de la base de que ambos están relacionados. En su trabajo, Díez Patricio (2017) comenta que “en su origen, saber tiene, pues, relación con lo que se sabe por experiencia” y de esta idea se parte la

conceptualizar esta dimensión a la que se le denomina “conocimientos y experiencias” que el personal investigador tiene sobre el uso compartido de datos. Se entiende que lo que sabe el personal investigador sobre la práctica de compartir datos lo sabe en la medida que han tenido algún tipo de experiencia o contacto con la misma. Cabe señalar también que la carencia o ausencia de conocimientos y experiencias también es información en sí misma y que ha servido igualmente para responder a los objetivos de este trabajo.

El *modus operandi* para realizar una aproximación a los conocimientos y creencias es escalonado y gradual. En primer lugar, se indaga acerca de sus prácticas investigadoras en general, para ofrecer la oportunidad a las personas entrevistadas de explicar a qué se dedican y cómo llevan a cabo su trabajo, a la vez que se crea un ambiente relajado donde las personas entrevistadas hablen sobre temas que controlan y no se sientan incómodos o intimidados. Dentro de este tema se pretende que el personal investigador explique qué tipo de investigaciones llevan a cabo y qué tipo de datos suelen manejar, para poco a poco acercarse a la práctica de compartir datos. Una vez llegado al tema los datos compartidos, se hace una aproximación desde varios ángulos para averiguar si efectivamente han compartido alguna vez, lo hacen frecuentemente, o no lo han hecho nunca. Estos ángulos incluyen hablar sobre las revistas en las que han publicado, los proyectos en los que han participado y la relación con las instituciones donde trabajan. Esta primera dimensión tiene el doble objetivo de crear un ambiente distendido, a la vez que se indaga sobre las experiencias que ha tenido el personal investigador. Ambos elementos tienen una función tanto para la entrevistadora como para la persona entrevistada. Para la primera, es la información que necesita para saber en punto en el que se encuentra la persona entrevistada en cuanto a las prácticas de compartir datos, que es importante para continuar la entrevista semiestructurada. Para la persona entrevistada, sirve para “aterrizar” de alguna manera en la entrevista y hacerse una idea del tema que se está abordando.

Como se ha comentado, “conocer” o “saber” casi siempre implica un conocimiento directo, un conocimiento que exige una proximidad con la cosa conocida (en este caso, el uso compartido de datos), una experiencia o un contacto. De esta manera, en el conocimiento del objeto existe por fuerza siempre una huella, mayor o menor, del sujeto. Lo que, dicho de otro modo, significa que todo saber tiene un componente de creencia, de subjetividad. En este sentido, no solo el conocimiento o experiencias guían y determinan las acciones, sino que también lo hacen las creencias que el sujeto tenga sobre las realidades que lo rodean. Los conocimientos y las creencias constituyen entonces una especie de mapas de la realidad en la que se desenvuelven los sujetos (Diez Patricio, 2017).

- Las creencias del personal investigador sobre el uso compartido de datos.

Tal y como se mencionaba en la dimensión sobre conocimientos y experiencias, las creencias son otro elemento importante en la vida de las personas que guían de alguna manera su manera de ser y estar en el mundo, ya sea de manera individual como colectiva.

Como concepto, el de “creencia” ha tenido un largo recorrido en la ciencia y ha sido clave para explicar y entender las bases de por qué las personas hacen lo que hacen. De acuerdo con Beck (1976), las creencias son estructuras cognitivas que se desarrollan a partir de experiencias del individuo, así como de factores ambientales, culturales y biológicos. Las creencias pueden entenderse entonces como un marco de referencia o conjunto de reglas que determinan la forma de ser de las personas en el mundo, el modo en que evalúan tanto las situaciones, como a los demás y a sí mismos, así como la forma en que unas personas interactúan con otras. Estas estructuras, que generalmente actúan sin que las personas sean conscientes de ellas, condicionan la manera en que se piensa y se actúa y les da un sentido tanto para las personas mismas como para quien las observa y las estudia desde fuera (Calvete & Cardeñoso, 2001).

Es importante tener en cuenta que las creencias actúan como una suerte de verdades subjetivas, es decir, algo que las personas sienten como cierto, y que no se debe confundir con la verdad objetiva, que es aquella que de alguna manera ha sido probada y confirmada. En definitiva, las creencias son aquello con lo que las personas cuentan a la hora de actuar de una forma u otra, sin pensar en ello, y automáticamente se comportan teniéndolas en cuenta. Por este motivo, el concepto de creencia es considerado como un constructo útil para entender las motivaciones de la conducta (Diez Patricio, 2017).

Por ello, en esta dimensión se pretende profundizar en las creencias e ideas preconcebidas que tienen el personal investigador sobre la práctica de compartir datos. Se trata de una dimensión muy subjetiva, como no podía ser de otro modo, donde se abordan temas especialmente sensibles, tales como los miedos, las preocupaciones, las ventajas percibidas, las dificultades, los puntos a favor y en contra de la práctica de compartir datos. En primer lugar, se pone al investigador en el centro como protagonista para que hable sobre sus sensaciones en la práctica diaria de su trabajo, pero también se le invita a que se distancie y explore sus sensaciones y percepciones sobre la ciencia en su conjunto. Mientras que la dimensión anterior se buscaba que los relatos se correspondiesen con realidades tangibles del tipo “he hecho esto”, “tengo experiencia en aquello”, o “esto no me lo han pedido nunca, pero he hecho esto otro”; en las creencias se busca que las personas den rienda suelta a lo que “creen” en el sentido de buscar las bases subyacentes a sus conductas, que a veces no son tan evidentes, sino que son más latentes, inconscientes o incluso contradictorias.

Por último, es necesario mencionar que las creencias y los conocimientos no constituyen los únicos factores que condicionan la conducta. Por su relación con el tema que ocupa esta tesis, es preciso destacar el concepto de “actitud”. En este sentido, las creencias podrían ser un predictor de cuáles van a ser actitudes hacia un determinado objeto, constituyendo una especie de expectativa de actuación. Por ello, se podría decir que toda actitud viene

ocasionada por un objeto/situación/referente (en este caso el uso compartido de datos), que a su vez es susceptible a una serie de creencias. Las actitudes adoptadas ocurren, entonces, porque hay unas determinadas creencias detrás y éstas son, a su vez, las que determinan en cierta medida la actitud.

- Las actitudes del personal investigador sobre el uso compartido de datos.

Desde la psicología social, se ha entendido históricamente que el comportamiento del ser humano viene determinado por la actitud. A medida que se evoluciona en la investigación sobre este concepto, se diversifica en la manera de conceptualizarse, así como en las diferentes aproximaciones y sobre todo entre las actitudes y su relación con el comportamiento. Estas diversificaciones incluyen ciertas controversias sobre el papel que pueden tener las actitudes en los comportamientos y la capacidad que pueden tener las primeras de predecir los segundos, pero no hay duda de que, de un modo u otro, estos dos constructos son relevantes en la realidad social de los individuos. Entre las diferentes aproximaciones que han ido surgiendo al respecto de esto, el enfoque de acción razonada ha logrado un cierto consenso. Según este enfoque, se asume que el comportamiento de las personas se deriva de sus creencias, actitudes e intenciones (Ajzen & Fishbein, 2005; Bohner & Dickel, 2011). Si se atiende a la definición contemporánea de actitud, se entiende que ésta es un constructo complejo y multidimensional, y que estaría compuesto por componentes cognitivos, afectivos y conativos (Rokeach, 1970).

Dentro de estas distintas formas de entender las actitudes y su relación con los comportamientos, se ha entendido para este trabajo como apropiada aquella que entiende que, para poder predecir el comportamiento de los individuos sobre un determinado tema, deben de abordarse las actitudes intentado concretizar lo máximo posible (Ajzen & Fishbein, 2005). Es decir, no se trata de indagar acerca de las actitudes en general hacia la práctica de compartir datos,

sino de trata de aterrizar a la realidad e intentar plantear situaciones concretas para explorar las actitudes que mantendrían ante ellas.

Con esta dimensión se aborda, entonces, lo que el personal investigador expresa que estaría dispuestos a hacer y el cómo lo haría planteando situaciones concretas. Todo ello teniendo en cuenta, como se ha mencionado, que tanto los conocimientos como las creencias actúan como trasfondo de guías para las actitudes. Esta dimensión tiene, además, mucho que ver con la anterior, ya que la anterior explora lo que creen y esta dimensión, lo que realmente están dispuestos a llevar a la práctica.

En definitiva, la idea de estas tres dimensiones es que se complementen ofreciendo una visión más holística de la relación del personal investigador con la práctica de compartir datos, desde un punto de vista más subjetivo y también más práctico.

2.3 Procedimiento

Se realizaron 21 entrevistas de manera presencial y 2 por videoconferencia debido a circunstancias derivadas de la COVID-19. Las entrevistas han tenido una media de entre 30 y 45 minutos de duración. Se utilizó una guía de entrevista semiestructurada con preguntas abiertas que fue elaborada y consensuada por el equipo de investigación en cuyas líneas se integra este trabajo. Los temas incluidos en la guía se inspiran en las preguntas del cuestionario sobre las prácticas del personal investigador españoles explicado anteriormente (R Aleixandre-Benavent et al., 2015, 2020) y en base a la literatura sobre el tema (ver guía en el anexo 1). Las cuestiones giraron en torno a las tres dimensiones escogidas: 1) los conocimientos/experiencias del personal investigador sobre aspectos relacionados con el data sharing (cuál es su nivel de conocimiento sobre la práctica en sí, experiencias de requerimiento de compartir datos en proyectos o revistas, experiencias formativas y facilidades de la institución a la que pertenecen, conocimiento sobre infraestructuras para compartir datos,

incentivos que se les haya ofrecido, entre otros); 2) las creencias (su valoración sobre la importancia de la práctica, las ventajas y desventajas que perciben en su día a día, consideración de los incentivos, percepción de barreras, miedos, entre otros); y 3) las actitudes (en su caso concreto, grado de disposición a compartir datos, exigencias o condiciones que se pondrían, disposición a compartir datos de otros, cómo les gustaría que se presentasen los datos para ser fácilmente reutilizables). Además, la guía estaba pensada para introducir otros temas emergentes interrelacionados si la dinámica de la entrevista lo requiere. Se hicieron entrevistas hasta hallar el punto de saturación, esto es, cuando las respuestas comenzaron a repetirse y resultar redundantes. Todas las entrevistas se hicieron en español.

Antes de comenzar la entrevista se les pidió a las personas entrevistadas que ofrecieran su consentimiento informado conforme aceptaban ser entrevistados garantizando la no identificación. Las entrevistas fueron recogidas con grabadora.

2.4 Análisis

Las entrevistas fueron transcritas manualmente en documentos en formato .doc y analizadas con el software de investigación cualitativa Atlas.ti (Scientific Software Development GmbH, 2021). Se eliminaron de las transcripciones todos los datos que pudiesen identificar a la persona entrevistada. Una vez en Atlas.ti, se llevó a cabo un enfoque de análisis temático, en el que las entrevistas se codificaron en base a los temas incluidos en las tres dimensiones de análisis (conocimientos/experiencias, creencias y actitudes) y a los temas emergentes, así como a las características relacionadas con el tipo de investigación, el área a la que pertenecen, los años de experiencia y cargo que ocupan.

Las unidades de análisis para la codificación, que se refiere a la parte del contenido que es la base para las decisiones que se tomen durante el desarrollo de los códigos, fueron las frases, párrafos y, en ocasiones, la respuesta completa a una cuestión de la entrevista. De esta manera, se codificaron aquellos fragmentos que, tras la transcripción y lectura, se considerase que tuvieran relación con las dimensiones de

análisis. En este sentido, se han tomado decisiones de codificación basadas en varios elementos del contenido, incluida la longitud, la complejidad, los significados manifiestos y los significados latentes, donde se tuvieron en cuenta tanto el tono como las expresiones de las personas entrevistadas (Milne y Adler, 1999).

Posteriormente a la codificación, se utilizó una de las herramientas de Atlas.ti para elaborar redes. Cada entrevistada cuenta con tres redes por cada dimensión de análisis (conocimientos/experiencias, creencias y actitudes) donde se recogen las frases textuales consideradas más relevantes de acuerdo con las temáticas. En el Anexo 2 se puede observar tanto las redes como las características de las mismas.

Para la redacción de los resultados, se tuvo en cuenta las categorías temáticas (conocimientos/experiencias, creencias y actitudes), los ámbitos de investigación (adiciones, biología reproductiva, emergencias, odontología y medicina regenerativa) y los años de experiencia investigadora. Con respecto a esto último, los años de experiencia, se realiza una división entre júnior y sénior, donde júnior serían el personal investigador con menos de 10 años de experiencia y sénior los que tienen 10 o más. Aunque se ha recogido como información sociodemográfica de los y las participantes, el sexo no se ha considerado como variable para este estudio. Cabe destacar que en la presentación de los resultados se evitan los conteos exactos de frecuencia, pero se usan los siguientes términos cuando un número considerable de entrevistados se detiene en un tema o significado: "la mayoría", cuando las tres cuartas partes de los participantes informan un significado; "Frecuentemente", "muchos" o "a menudo" para más de la mitad de participantes; y "algunos" u "otros" para menos de la mitad. Se corrigieron errores gramaticales no significativos en las citas para facilitar la legibilidad y la comprensión. Para mantener el anonimato, se clasificaron a las personas entrevistadas utilizando el número de entrevista, institución de afiliación y tipo de investigación.

Resultados

La estructura de los resultados está elaborada teniendo en cuenta las dimensiones utilizadas (conocimientos/experiencias, creencias y actitudes). Las áreas de investigación a las que pertenecen las personas entrevistadas, que fueron obtenidas a partir de los cinco estudios sobre el uso del material suplementario llevados a cabo en el marco del proyecto en el que se inscribe esta tesis, han servido para escoger a personas investigadoras de distintas disciplinas de las Ciencias de la Salud. Estas áreas son *Substance Abuse, Emergency Medicine, Reproductive Biology, Cell & Tissue Engineering y Dentistry, Oral Surgery & Medicine*. Por sus distintas características, estas áreas son muy útiles para abarcar a personal investigador con perfiles variados, diferentes tipos de investigación, distintas titulaciones y procedentes de diversas disciplinas. A medida que se realizaban las entrevistas, el tipo de investigación que las personas llevaban a cabo emergió como variable más relevante, junto con los años de experiencias. Es decir, las áreas de investigación a las que las personas entrevistadas pertenecían no eran tan importantes como el tipo de investigación llevado a cabo dentro de las mismas. Así, en cuanto a la práctica de compartir datos, investigadores con líneas más básicas eran más parecidos entre sí, aunque pertenecieran a áreas diferentes, que un básico y un clínico dentro de una misma área. Esta realidad fue especialmente evidente en el área de *Substance Abuse*, que tenía una combinación de personas entrevistadas clínicas y básicas, cuyas coincidencias en los conocimientos, creencias y actitudes sobre compartir datos no iban más allá de llevar a cabo investigaciones en el ámbito de las adicciones. Por otro lado, personal investigador de áreas como *Emergency Medicine y Dentistry, Oral Surgery & Medicine*, eminentemente clínicas, es decir, aplicadas, contaron con numerosos aspectos en común. Esto mismo sucedió con las personas entrevistadas de disciplinas afines a *Reproductive Biology y Cell & Tissue Engineering*, mayormente con perfiles básicos, que tenían más en común entre sí que con cualquiera de las dos anteriores (*Emergency Medicine y Dentistry, Oral Surgery & Medicine*).

Por ello, en la explicación de los resultados que viene a continuación se tienen en cuenta, dentro de las dimensiones de análisis, aspectos relacionados con los tipos de investigación (básica y aplicada) y los años de experiencia investigadora.

1. Conocimientos y experiencias

Los conocimientos y experiencias se han abordado desde tres bloques de temáticas: las revistas y sus modalidades para compartir datos, los datos compartidos en proyectos de investigación y la formación y apoyo de la institución sobre la gestión y el manejo de los datos de investigación. A continuación, se analiza cada uno de estos bloques por separado.

1.1. Experiencias relacionadas con las revistas y las modalidades de éstas para compartir datos

Cuando se ha abordado el tema de las revistas y de las demandas de éstas para compartir los datos brutos, las diferencias entre el personal investigador de líneas básicas y aplicadas ha sido notable. Mientras que el personal de básica afirmaba, en mayor o menor medida, haber tenido algún contacto con revistas que tenían este requerimiento (bien obligatorio, bien opcional), el personal con un volumen mayor de investigaciones aplicadas raramente había tenido siquiera contacto con este tipo de peticiones.

Dado que son el personal investigador con líneas predominantemente básicas quienes han tenido contacto con este tipo de situaciones, se comienza el análisis con este perfil.

En primer lugar, como se ha mencionado, todo el personal investigador de básica entrevistado tuvo algún contacto con peticiones de compartir los datos por parte de las revistas. Pero la frecuencia, las modalidades y esfuerzo percibido varían bastante no solo entre tipos de investigación, sino también entre investigadores junior y sénior.

Con respecto a la frecuencia, mientras que hay investigadores que refieren hacerlo de manera rutinaria, otros solo lo han hecho alguna vez de manera puntual. Entre los que lo hacen de manera rutinaria, lo vinculan a una petición obligatoria que hacen las revistas, una condición que hay que cumplir y que asumen como parte del trabajo:

“Las revistas lo solicitan claro, y lo tienes que hacer. Porque antes de publicarse los revisores lo evalúan, hay algunos que se ponen a ver si lo has hecho bien o no, y otros que se fían. Pero tú tienes que hacerlo, tú tienes que poner tus datos crudos y ya es su responsabilidad mirarlo o no” (E020)

Además, otros opinaban que es una petición normal cuando se quiere publicar en revistas con impacto alto, así como cuando lo que se está publicando es algo novedoso:

“pero desde luego, los que estamos acostumbrados a publicar arriba no te planeas eso. Tú te planteas hacer las cosas bien, y luego publicarlas. Tú te planteas hacerlo en NIH.gov, en ClinicalTrials, en seguir los paths que hay, en cumplir los comités éticos. Y cuando tienes tus datos, los publicas ahí arriba y que todo el mundo los vea. Otra cosa es que yo no la concibo” (E017)

“De hecho, muchas veces para publicar en según qué revistas, si son así de índice de impacto alto, incluso te pueden pedir los resultados de los “western blot” y eso es lo más crudo que hay” (E014)

“Te lo obliga la misma revista. Las revistas cuando es una cosa original, una cosa nueva, una cosa creativa, pues tienes que compartir las bases de datos” (E003)

Sobre las modalidades, también se detecta un patrón entre el personal investigador que publica datos frecuentemente y quien lo hace de manera puntual. Quien publica frecuentemente, quien tiene integrada la práctica, menciona los repositorios de datos de manera espontánea. El repositorio al que más hacen referencia es Gen Expression Omnibus (GEO), aunque también mencionan otros como ArrayExpress, Cancer Gen

Atlas y Sequence Read Archive. Mientras, los que publican de manera puntual relatan experiencias más relacionadas con subir los datos como material suplementario:

“Lo que hice fue un Excel suplementario en el que cada hoja era una prueba. Es un poco organizar lo que tienes, pero eso es algo que no sé, no tardas más de un par de horas” (E002)

“De hecho, lo que hago a veces es que copio la matriz del SPSS y la pego en un Excel, y eso es lo que subo, porque esos son los datos. Con un repositorio nunca he interactuado, no tengo ni idea de si es complicado, ni como se hace o cual elegir” (E001)

Además de las opciones de repositorios y material suplementario, también existe la opción de “bajo demanda”, explicada en la introducción. Esta opción también salió a relucir en una de las entrevistas:

“Me pedían, o subirlo a un repositorio, que te daban opción de varios, o bajo demanda, entonces si me dan la opción de “bajo demanda”, yo lo doy bajo demanda y así no tengo que trabajar en ese momento, lo dejo para más adelante” (E001)

La última afirmación de la persona entrevistada E001 de “así no tengo que trabajar en este momento” sirve para enlazar con otro aspecto, que es el del esfuerzo percibido. En este caso, las diferencias más evidentes se daban entre el personal investigador con menos responsabilidad en la jerarquía (normalmente, júnior) y los que más (sénior). En el caso de los júnior, solían referir que son ellos los encargados de lidiar con la tarea de curar y subir los datos, bien a repositorios o como material suplementario:

“Entonces te encuentras con que tú eres la responsable del trabajo, de subirlo, pero no tienes ni idea de bioinformática, no sabes manejar los sripts y no sabes hacer eso...pues a lo mejor te cuesta un poco más, y tienes que buscar ayuda de alguien que sí sepa. Porque no es tan trivial, o al menos, yo no lo he visto tan trivial.” (E020)

“Sí, autodidacta. Porque aparte los jefes nos dicen “informaros de como se hace esto y hacerlo”, ellos mismos no saben tampoco.”

(E010)

Todos estaban de acuerdo en que se trata de un esfuerzo extra que se añade al resto de tareas propias de la investigación, sin embargo, algunos lo percibían como una tarea engorrosa, mientras que otros la aceptaban con más naturalidad:

“Hombre, fue un trabajazo... O sea, le dediqué muchísimas horas. Porque claro tú vas evaluando, pero tú no vas guardando cada imagen con cada cálculo. Tú directamente guardas el Excel. Entonces guardas el dato y vas organizando esa información. De repente, descárgate todas las imágenes de nuevo, vuelve a calcular, porque tú tienes que hacer esa representación sobre la imagen” (E005)

“Para mí no (supuso una dificultad añadida), cuando yo vi que ese era un requisito, a mí no me supuso más que una tarde de hacerlos en inglés, por ejemplo, ponerle los nombres digamos “formales” que corresponden con los que están en el artículo, haciendo un poco la correspondencia con el paper en cambiar los nombres de las variables” (E002)

“Entonces todo entonces siempre supone un estrés añadido al trabajo adicional. O la ayuda de los servicios informáticos, del personal que bueno es super amable y te ayuda, pero no es algo que sea ni muy intuitivo, ni muy sencillo ni muy rápido. Es algo que va totalmente en contra de lo que normalmente esperamos de todo este tipo de experimentos, porque cuesta más de lo que puedas tú analizar a nivel individual o personal.” (E006)

Sobre la mención a los informáticos que hace la última referencia, las personas entrevistadas, sobre todo investigadores seniors que trabajan en instituciones privadas con una apuesta muy fuerte por la investigación, hicieron alusión a la presencia de bioinformáticos en sus grupos y en la importancia de éstos debido a la dificultad que supone el manejo de técnicas como la secuenciación masiva.

“Esto los bioinformáticos me lo hicieron, porque yo no sabía dónde se subía, ellos subieron toda la secuenciación masiva” (E003)

“Realmente, lo suben ellos, en nuestro caso los datos al repositorio lo suben los bioinformáticos responsables del artículo publicado y ellos simplemente nos dan el link para que nosotros lo pongamos en el artículo, comprobemos que está bien, y luego se manda a los revisores” (E018)

En el caso del personal investigador de perfiles más aplicados, como se ha mencionado, normalmente no había tenido contacto con este tipo de peticiones por parte de las revistas, ni para publicar datos brutos como material suplementario, ni para depositarlos en un repositorio. Solamente hubo un entrevistado que manifestó que los datos de uno de sus artículos habían sido compartidos como material suplementario. En la mayoría de los casos no sabían de su existencia, pensaban que era una práctica que no se aplicaba en su campo u opinaban que era muy reciente:

“Al nivel de la investigación nuestra, no sabemos ni si las revistas lo permiten, o lo fomentan. En el proceso de envío no nos lo piden, ni nos lo sugieren, así que hacemos como toda la vida, vamos” (E008)

“No, y principalmente yo creo que es porque son (las revista) del ámbito de la antropología, y en este sector no se suele compartir...” (E019)

“No, no. No he publicado por ahí, y no sabía que existía esa opción, ese requisito” (E013)

“No, no tengo constancia. A mí personalmente no me ha pasado. Por lo que me cuentas, parece algo muy reciente, no debe ser algo que se haga desde hace mucho tiempo” (E022)

“En el artículo del registro de COVID sí que está la información bruta de la base de datos. La verdad es que es un trabajo estadístico muy duro porque son muchos miles de pacientes. Está en el propio artículo, como archivo suplementario en formato PDF” (E023)

Sin embargo, cuando a las personas entrevistadas con investigación mayormente aplicada se les introducía el tema de los repositorios, se halló un patrón de respuesta que era común en muchos de ellos. Aunque ninguno había tenido contacto con repositorios de datos propiamente dichos (descritos en el apartado 3 de la introducción “Infraestructuras para compartir datos: los repositorios de datos abiertos en Ciencias de la Salud”), relacionaban el concepto con otro tipo de infraestructuras digitales que alojan bases de datos de acceso restringido, ligadas a proyectos concretos o pertenecientes consorcios privados donde los datos se comparten entre los miembros con permisos de acceso. A este respecto, surgieron varias veces nombres como la “Red de Trastornos Adictivos”, la “Red de Toxicovigilancia”, el “GRIDSAN” o el registro de la Sociedad Española de Medicina Interna. Tal y como ellos explicaban, son redes, consorcios o plataformas formadas por investigadores con temas de investigación comunes donde se suben datos brutos correspondientes a proyectos o estudios compartidos. Aunque estos recursos explicados tenían características diferentes y se organizaban de distintas formas, todos tenían dos denominadores comunes: el acceso era restringido con usuario y contraseña y, aún teniendo usuario y contraseña, el acceso a la consulta y descarga de la información estaba limitada a los datos que ese investigador había introducido, no a toda la base de datos.

“Yo puedo buscar la información de mi centro. Puedo ver los datos que he metido yo de mi hospital o los que ha metido algún compañero mío. Para tener acceso a todo el registro de datos tendría que presentar alguna propuesta de artículo o de estudio más o menos serio. Tendría que hablar con el coordinador de proyecto, con el IP, y pedirlos. Con un motivo de peso podría utilizar todos los datos” (E023)

Por otra parte, con respecto a investigadores con un perfil más clínico, también se relacionaba el uso compartido de datos clínicos con estudios multicéntricos (en el caso del personal investigador que trabajan en hospitales y otros departamentos de salud) y con los metaanálisis y los estudios randomizados:

“Sí ha habido proyectos colaborativos, en CIBERSAM hemos participado en primeros episodios psicóticos, y luego en rehabilitación funcional del trastorno bipolar, y ahí no ha habido problemas en compartir datos, porque eran multicéntricos. Esa es la clave. Son estudios multicéntricos en los que la muestra se recluta de varios centros, entonces claro, hay que unirla, hay que reunirla”
(E004)

“Es que un ómnibus es muy fácil decir qué es qué. Pero en un estudio clínico puede pasar que las indicaciones sean distintas, que las pacientes sean distintas. Por lo tanto, un repositorio público como tal de esto no hay, pero yo creo que los equivalentes son los randomizados y los metaanálisis” (E017)

1.2 Experiencias relacionadas con los proyectos de investigación

Cuando se preguntó por su participación en proyectos en los que existiera la recomendación, obligación o cualquier otro tipo de alusión a compartir los datos de investigación, la respuesta fue unánime en todas las personas entrevistadas: ninguna recordaba una referencia de este tipo en los proyectos en los que participaban o habían participado. Al profundizar un poco más en el alcance de los proyectos en los que participaban, era frecuente que las personas entrevistadas refirieran participar tanto en proyectos nacionales como internacionales. En el caso de los internacionales, surgieron referencias a la obligatoriedad de algunos proyectos financiados de publicar los resultados en revistas de acceso abierto, pero no recordaban ninguna mención a los datos. Cabe matizar que, entre algunos de las personas entrevistadas juniors, surgió espontáneamente la aclaración de que ellos no eran quienes tenían el control o poder de decisión sobre qué se incluía o no en los proyectos, y de que no siempre estaban al tanto de cuáles eran las directrices de los mismos. A pesar de esta aclaración, manifestaban que, probablemente, de existir el requisito de compartir datos hubieran tenido alguna constancia.

1.3 Formación y apoyo de la institución

Con respecto a la formación recibida sobre datos brutos, su uso compartido en investigación y todas las técnicas que involucran su gestión y difusión, se halló un patrón de respuestas bastante similar entre todas las personas entrevistadas: en general, no recordaban haber recibido formación específica sobre este tema y sus alusiones eran (en el caso de los que hubieran tenido contacto con esta práctica alguna vez) haberla aprendido sobre la marcha:

“Pues no, esto es que tú te vas a publicar el paper, y entonces es cuando te lees las “guidelines” para publicarlo. Entonces te dicen que si quieres publicar datos de secuenciación masiva los tienes que subir a este repositorio o, bueno, te da una lista de repositorios a los que los puedes subir y tú eliges en función del tipo de datos que tú tengas. Entonces te metes, ves un poco como funcionan, y así. Pero nadie te lo dice, es un poco lo que te vas encontrando” E020

“Si la universidad nos enseñara cómo funciona la universidad ya sería suficiente. Te vas enterando a medida que te vas encontrando con los obstáculos” E008

“No, desde que estoy aquí no. Sí que nos han ofrecido cursos de análisis, pero de reposición de datos y todo eso, no” E012

Dentro de esta negativa general, existía algún matiz a la hora de responder a esta cuestión. Estos matices hacían referencia a puntualizaciones sobre la existencia de esta formación en sus instituciones, aunque para otros públicos, o la posibilidad de que la institución haría formación si fuera necesario. Normalmente, el denominador común de estas respuestas era que las personas entrevistadas pertenecían a instituciones solo dedicadas a la investigación:

“Ellos (la institución) nos dicen de qué hacemos los cursos, y cuando cambia la ley te hacen los cursos, para que veas como tienes que

almacenar los datos...todo, sí. Ellos son los que te cuentan cómo hay que hacerlo” E007

“A ver, los bioinformáticos sí que hacen muchos cursos de metaanálisis, tenemos un departamento muy fuerte de bioinformática. Y ellos hacen cursos, pero son cursos muy avanzados. No son para gente que trabaje en investigación básica en bancada de normal, sino para alguien que ya tenga unos conocimientos previos y ellos sí que te enseñan todo eso, pero yo sí que he ido a un curso y me quedé por el camino, no me enteré demasiado. Lo que sí que hacen es ofrecernos curso de análisis de datos y todo eso, pero no tanto con las bases de datos. De hecho, es una posible sugerencia que yo podría hacer, que ellos nos preguntan ¿en qué queréis formaros? Y nunca caemos en eso” E006

“El departamento de medicina interna ha puesto en marcha un curso sobre investigación clínica y tratamiento de datos, pero aún no lo he empezado, así todavía no sé... Pero bueno, parece que sí que hay algo más de movimiento ahora que hace un par de años. De todos modos, en general en el hospital la formación en investigación es bastante pobre, más que nada porque no tienes tiempo.” (E023)

2. Creencias

Las creencias se han abordado desde tres bloques de temáticas: 1) las ventajas, 2) las desventajas y los miedos y 3) las reflexiones sobre otros aspectos de la práctica de compartir datos. A continuación, se analiza cada uno de estos bloques por separado.

2.1. Sobre las ventajas

El primer punto que es importante aclarar sobre las ventajas, cuyo patrón se repite también en los miedos, es que las personas entrevistadas que investigan en áreas básicas (y que sí tuvieron contacto con la práctica de compartir datos) tratan estos temas basándose en sus experiencias, mientras que el personal investigador de áreas aplicadas se mueven más en el terreno de la especulación, ya que únicamente habían hecho intercambio de datos entre personas conocidas mayormente en contextos de colaboración.

Comenzando con el personal investigador de básica, cuando se les pregunta por ventajas que observan en la práctica de compartir datos, las respuestas son muy semejantes, sin distinciones de patrones entre júnior y senior.

Destaca especialmente la alusión al aumento de la transparencia y la reproductibilidad de los trabajos cuando se comparten los datos brutos, ya que permiten ver la base del estudio y no solo la información procesada. Relacionan esta garantía de transparencia y reproductibilidad que ofrece el mostrar los datos a una forma de elevar la calidad y la credibilidad tanto de los trabajos como de las investigaciones:

“Me va a dar muchísima pereza, es antipático, pero yo creo que va a subir la calidad y credibilidad de las publicaciones” E001

“Porque claro, cuando te van a evaluar a ti como investigador, tienen que tener acceso a esos datos para comprobar que realmente no estás engañando ni estás haciendo nada fraudulento” E020

“Creo que es imprescindible que compartan los datos. Primero porque la ciencia tiene que ser reproducible, si no tienes los datos tal y cómo han sido usados para desarrollar cada una de las publicaciones, no puedes reproducir los resultados de ninguna manera, aunque quieras. Entonces, yo creo que se tiene que compartir todo lo que se ha utilizado para generar los resultados”

E010

Otra ventaja mencionada fue la posibilidad que ofrece compartir los datos de que otras personas puedan ver en ellos aspectos interesantes que los investigador iniciales no detectaron, o que descartaron por no responder a sus objetivos. En general, existía la conciencia de que no siempre se les saca el máximo provecho a los datos y consideraban la importancia de exprimirlos tanto como sea posible, sobre todo cuando la financiación proviene de fondos públicos:

“Y también creo que habría que compartir los datos porque a veces a otro investigador de otra disciplina con esos datos se le pueden ocurrir cosas, o sea, variables diferentes o modos diferentes de analizar algo, que yo creo que no estás explotando un dato” E002

“Yo creo que es interesante que haya alguien ajeno que pueda ver esos datos, porque a la larga tú estás investigando, y aunque intentas hacer algún tipo de ciego, es muy complicado” E005

“Una vez tu publicas el artículo, tus datos son públicos. Tienen que ser públicos. O sea, tú estás trabajando realmente para la comunidad científica, tú realmente no trabajas para ti. Tú cuando encuentras todo eso es para darlo al conocimiento común, habrá cosas que las utilicemos nosotros, pero habrá otras muchas que nosotros no trabajamos en reproducción o en enfermedad, que otros muchos científicos se pueden beneficiar” E018

Además, se menciona la ventaja de poder sumar muestras cuando los parámetros de la investigación sean compatibles. Relacionado con esto, también se menciona que, si los datos estuvieran en abierto, se podría ver qué y cómo están haciendo

investigadores de áreas cercanas a las suyas, para que la ciencia sea más abierta, menos parcelada y se agilizaran más los procesos. En esta línea, también se hizo referencia a la posibilidad de que, a partir de ver los datos de otros o que otros puedan ver los suyos, surjan colaboraciones:

“Para mí es importante, claro, porque cuantos más datos tengas, mejor vas a entender cómo funciona el proceso biológico. En otros campos me imagino que será igual, cuantos más datos tengas mejor, para comprender los procesos” E020

“A ver yo siempre creo que es mejor subirlo, aunque no te obliguen. Porque igual hay un colega que lee tu paper y dice, ostras, pues voy a coger sus datos y voy a compararlos con los míos y vamos a ver si podemos colaborar” E011

“Claro, y sería interesante, porque así de esa manera también agilizaría mucho el tener acceso a otras cosas. Entonces si tú tienes acceso a los resultados de otros puedes conocer mejor lo que se está haciendo, yo muchas veces no sé ni siquiera en qué trabaja el laboratorio de enfrente. Hay mucho secretismo al respecto y no creo que sea favorable, porque si compartes más cosas, vas a tener más ideas luego” E006

En cuanto al personal investigador con líneas aplicadas, como se mencionaba al principio del párrafo, la gran mayoría no había tenido experiencia de compartir datos ni en la modalidad de material suplementario ni de repositorios de datos, únicamente la habían llevado a cabo peer to peer en contextos de colaboración donde el personal investigador se conocía. A pesar de estos, las ventajas que resaltaban de esta práctica fueron similares a las de los básicos y también entre los júnior y los sénior.

De hecho, la principal diferencia con el personal investigador de básica fue el resaltar la posibilidad de aumentar muestras que ofrece la práctica de compartir datos. Muchas de las personas entrevistadas mencionaban que en sus disciplinas no es fácil conseguir sujetos, por lo que la suma de los mismos es especialmente relevante.

“Interesante sobre todo porque yo siempre parto de que lo que hacemos es de casuísticas bajas, sumaría y sería bueno” E022

“Que vaya más rápido la investigación. Especialmente en el área donde yo trabajo, que es salud mental, que es muy difícil conseguir pacientes” E013

El resto de las cuestiones percibidas como ventajas son similares a las que resaltó el personal investigador básico, especialmente la relacionada con la transparencia. En este sentido, se mencionaban aspectos como que compartir datos podría enriquecer el feedback en las revisiones por pares, ya que los revisores podrían chequear con más fidelidad los resultados. Además, se destacan aspectos como el aumento de la rapidez en las investigaciones y ahorro de dinero y de esfuerzo, así como la posibilidad de crear nuevas colaboraciones y como canal para llevar a cabo buenas ideas.

“Pues yo creo que sí deberían compartirse. Porque claro, es una forma de buscar una garantía de que el fraude científico sea menor. Aunque lo datos brutos pueden ser igualmente fraudulentos, pero bueno, es otro mecanismo más por el cual si alguien se pone a rascar tu artículo tiene una base para hacerlo” E023

2.2. Sobre las desventajas y los miedos

Se ha considerado tratar el tema de las desventajas y los miedos en un mismo apartado, ya que en el transcurso de las entrevistas se detectó que, bajo el concepto de desventaja, las personas entrevistadas también estaban haciendo referencia a miedos percibidos.

En este caso, sí se ha detectado un matiz de diferencia entre investigadores júnior y sénior en relación a jerarquía profesional. En el caso de los júnior, se destaca como desventaja la dependencia del Investigador Principal (IP) a la hora de compartir los datos. Esto significa que, independientemente de que tengan el propósito o no lo tengan de compartir, tendrán que seguir el dictado de sus superiores. Esto se aplica

tanto si quieren compartir como si no, es decir, que, si no desean hacerlo, pero les obligan, refieren no tener mucho margen de elección. En este sentido, también se nombra la presión y la responsabilidad que tienen los IP, ya que son ellos los que consiguen financiación para nuevos proyectos, lo que puede hacerlos más conservadores a la hora de liberar los datos:

“Entre los técnicos, los predoc, los posdoc, la cosa va mucho más fluida. La gente que está pidiendo el dinero, los IP que consiguen el dinero, que consiguen los proyectos, lo entiendo más porque igual, si cuentas algo pues otra gente puede utilizar eso para pedir un proyecto que tú habías pensado pedirlo al año siguiente” E012

“Si tú trabajas en un grupo y hay un cabeza por encima de ti, si su opción no es compartir datos, tú ya puedes querer, que nunca compartirás datos. Entonces, me refiero un poco a eso, si alguien te apoya, pues es mucho más fácil ¿no?” E005

“De hecho, los IP suelen proteger bastante esa información. En el caso del COVID, nosotros les pedimos que nos dieran nuestra propia información, en la base del COVID, y hasta que no salió la publicación, no nos pasaron ni siquiera la información de nuestros propios pacientes” E023

Otra desventaja que constituye un patrón entre las personas entrevistadas es la sensación de propiedad de los datos. A este respecto, es frecuente que manifiesten que la percepción de propiedad de los datos es algo común entre investigadores. Cabe resaltar que cuando abordan este tema, las personas entrevistadas hablan en tercera persona del plural, como si se tratara de algo que sucede al resto, pero sin incluirse ellos.

“Uf, pero yo creo que ahí lo que hay que vencer es la resistencia del investigador casi innata a soltar algo que es suyo” E001

“Desde mi punto de vista, yo creo que en ciencia siempre está el punto, aunque compartamos mucho, siempre está el punto de

propiedad. El “esos son mis datos, y si tú quieres tener ese tipo de información, hazlo tú también” E014

Más relacionado con el plano de los miedos, se detecta un discurso de temor a que “te roben los datos”, es decir, a que otros utilicen sus datos para tomarles delantera en la investigación. Sobre esto, recalcan la alta competitividad en la ciencia y la cuestión de la financiación como detonante para activar el afán proteccionista. No obstante, es interesante destacar que este discurso era común entre investigadores cuya práctica profesional estaba alejada de la práctica clínica, es decir, de la aplicación directa sobre pacientes.

“Yo no defiando mis datos con dientes, porque creo que no es algo positivo. Pero la gente que defiende su trabajo para tenerlo bajo llave, es diferente, porque hay personas que igual trabajan de cara a patentes, entonces no puedes compartir los datos y hay mucha gente que lleva ese tipo de orientación. Entonces yo sí que estoy a favor de compartir los datos, pero ¿y si alguien me los roba y adelanta?” E006

“A ver, la única desventaja que veo es que, si tú sacas un protocolo que te ha costado un montón de tiempo y si, una vez se valida, se lo pasas a los demás, pues bueno, tira que te va. Pero sí que es verdad que, si a ti te ha llevado un trabajazo y es algo así puntero y tal, que de repente tú tengas que enviar toda esa información a alguien, pues igual, no sé...” E005

“Entonces, si tu publicas todo tu registro, tu base de datos de pacientes en el paper, el segundo (un posible segundo paper) lo puedes publicar tú, o lo puede publicar cualquier otra persona. Entonces en ese aspecto, para el investigador regalar la información que ha publicado puede ser desventajoso. (...) porque hay una presión por publicar, porque si no públicas no tienes contratos para los siguientes años, o no te dan la misma financiación que te han dado porque son unos proyectos privados, o de centros mixtos, pues hay más presión.” E023

También en el terreno de los miedos, se detectó el temor a la crítica. En este sentido, se hacía referencia a que, al enseñar los datos, se exponían a que otros te pudieran acusar de haber realizado un mal trabajo, conllevando un desprestigio sobre el mismo. En relación con esta idea, varios entrevistados hicieron hincapié en la tendencia en la comunidad científica a ocultar o a no publicar datos negativos, también por miedo a la crítica.

Otro miedo que fue común entre las personas entrevistadas es a la mala interpretación que podían hacer de sus datos otras personas. En ocasiones, este miedo iba acompañado de la inseguridad que genera el compartir y que otros investigadores hagan interpretaciones erróneas o inexactas de los datos y que eso pueda generar debates o incongruencias.

“Estás viendo la cocina de los otros grupos y claro, están trabajando con unas técnicas que tú no conoces tú tampoco tienes criterio para decir si estos datos han tratado bien o nos han tratado bien, eso también hay que verlo, porque cuando tú no entiendes, tú no conoces no tienes la experiencia de la técnica de esos datos brutos, tú puedes sacar conclusiones totalmente equivocadas” E001

Sin embargo, es importante destacar que hubo algunos entrevistados que restaron importancia a este aspecto, resaltando que el problema de las malas interpretaciones no es de quien crea los datos sino de quien los reinterpreta.

“El uso que haga la gente de tus datos puede ser bueno o malo, pero la base es buena y es transparente, si alguien los usa y lo hace mal, tú siempre puedes ir después y decir “oiga mire, usted lo ha hecho esto de forma errónea basándose en una base que era correcta”” E017

Por otro lado, muchas de las personas entrevistadas hicieron referencia a la problemática de asegurar la confidencialidad. Sin embargo, en este caso hubo una diferencia notable entre el personal investigador de básica y aplicada. En el caso de los primeros, esta preocupación se mencionaba más para ejemplificar una preocupación que pueden tener otro tipo de investigadores que para referirse a un problema que les concierne a ellos. Esto se debía a que, dada su experiencia (la mayoría habían tenido

contacto con repositorios o con datos como material suplementario), daban por sentado que los responsables de las plataformas correspondientes asegurarían el aspecto de la confidencialidad. De hecho, hubo incluso un entrevistado que aseguró que alegar desconfianza debido a la confidencialidad podría ser una excusa para no compartir.

Sin embargo, el personal investigador de aplicada ponía mucho más énfasis en la confidencialidad no tanto como una desventaja, sino como un factor importante a tener en cuenta. En este sentido, muestran preocupación porque los datos que manejan, en el caso de que tengan que ser compartidos, estén bien protegidos. Otro tema que salió a colación de manera espontánea en varias entrevistas (sin haber preguntado directamente por ello), es el consentimiento de los pacientes a la hora de plantearse el compartir los datos. La mayoría del personal investigador manifestó que no pensaban que los pacientes fueran a plantear ningún problema, siempre que se les explicase la temática, las razones por las que era importante y se respetase la confidencialidad.

Solo hubo una persona que manifestó dudas en este tema, y es importante destacarla porque, según ella, el problema de la confidencialidad y de solicitarle a los pacientes que los datos sean compartidos es más peliagudo cuando se trata de datos cualitativos. En su razonamiento, los datos cualitativos contienen mayor carga de información personal y sensible y los propios sujetos del estudio podrían poner problemas en este sentido. Se podría presentar la situación de que fuera difícil anonimizarlos.

“Yo creo que depende del tipo de datos. Los datos cuantitativos lo veo pertinente y no veo ningún problema. Los datos cualitativos a pesar de que sí, yo me posiciono a favor de la transparencia, tengo dudas de que todo el mundo llegue a opinar de esa manera, y, además, depende del tema. Entre los datos clínicos no está la historia de la persona, que te cuenta la relación con su madre o... es que estamos hablando de la intimidad, la vida íntima de la persona.”

E019

Otro aspecto que salió a colación entre el personal investigador aplicado fue el señalar como desventaja la falta de cultura o tradición de compartir datos en sus áreas de investigación. Fue un patrón generalizado el manifestar que el uso compartido de datos no se llevaba a cabo en sus áreas y que esto influía en la sensación de que era algo que “no iba con ellos”. Además, se manifestó la idea de que mientras no se obligue, mientras no exista la exigencia, bien por proyectos o por revistas, es posible que esta realidad no cambie.

Yo creo que para mí las barreras fundamentales son una falta de costumbre, de cultura de trabajo de hacerlo, que no te lo exigen, y luego este asunto del sentimiento de propiedad de los datos de uno, que se los ha trabajado y que los ha recolectado. E004

Yo supongo que los utilizan sobre todo los básicos y los epidemiólogos. Los clínicos no... yo creo que no. E007

“Entonces igual que nos piden ahora registrar todos los ensayos clínicos desde el 2017, supongo que llegará el momento en que te sugieran o fomenten que compartas datos. Pero bueno, vamos adaptándonos a medida que van llegando las necesidades, como ya tenemos mucho lío...líos adiciones solo cuando nos obligan prácticamente.” E008

Con respecto al comentario que hace el entrevistado E004, otra de las desventajas – miedos percibidos, que coincide con las personas entrevistadas de básica, es la sensación de propiedad de los datos. En este caso, sí se detecta un matiz de diferencia a la hora de tratar este tema entre los júnior y los sénior. Se detecta que los segundos (sénior) manifiestan más preocupación por temas relacionados con gestión de proyectos y la propiedad intelectual, haciendo alusión a que te puedan adelantar en la investigación antes de que se puedan extraer todos los resultados:

“Lo otro, los datos, si es antes de finalizar (el proyecto) lo veo difícil, porque si está financiado, ya tienes que cumplir con unos plazos y unos objetivos. Si por en medio te sale otra persona y te coge tus datos, te rompe en ese sentido.” E022

Sin embargo, las alusiones fueron mayoritariamente dirigidas a percepción de propiedad y del ego del personal investigador hacia los datos. Igual que sucedía con el personal investigador básico, trataban esta cuestión en tercera persona del plural, como algo que le sucede a “el personal investigador” en general y no a ellos en particular.

“Uno, pues por temor a que el resultado de tu trabajo pueda ser aprovechado por otros, entra dentro de los egos del investigador, y luego, porque yo no sé muy bien cómo articularlo... por ejemplo, en el tema de la genética sí que lo veo más claro.” E021

“Cada uno tiene sus justificaciones para poner una barrera, cada uno de ellos. Claramente no te van a decir que es mi ego el que está involucrado. El ego en el sentido de que no se ve como parte de un todo. Así como el planeta es uno solo, la investigación es solo una, es una forma de avanzar la humanidad y tiene estas cuestiones que la atascan a veces.” E013

“Porque yo creo que a nivel nacional (de España) debe haber todavía esa cosa. Esto es mío, estos son mis datos, lo he publicado yo y lo voy a seguir explotando yo, y no te vas a sacar tú el Nature si los he recogido yo. Me parece absurdo, pero creo que puede existir todavía.” E015

Otro aspecto destacado en el caso del personal investigador aplicado, que en este caso difiere del básico, es el de las percepciones de las dificultades técnicas. De manera general, las dificultades o dudas técnicas sobre el propio proceso de compartir datos (relacionados con la infraestructura, los procesos de confidencialidad, la organización de los datos y los metadatos) venían sobre todo del hecho de no haber compartido nunca. Al no haberse visto en la situación (ni con datos como material suplementario, ni en repositorios, ni en proyectos) en muchos casos, les costaba incluso hacerse una idea del proceso y lo veían alejado de su práctica investigadora cotidiana. Según sus percepciones, la infraestructura tendría que estar muy bien organizada y gestionada y debería de garantizar en todo momento la confidencialidad.

Finalmente, algunos investigadores (tanto básicos como aplicados) manifestaron que los mandatos deberían venir desde instancias elevadas, como el Ministerio o Sanidad, o de las propias revistas y editoriales y no depender de la buena voluntad. Pedían, en resumidas cuentas, estrategias uniformes.

“Claro, si al final eso fuera un estándar que se introduce en todas las revistas, de todos los países, y es obligatorio siempre hacerlo, pues sería un buen mecanismo. Pero si se introduce solo en ciertas revistas, pues dejas la puerta abierta a que haya prácticas que no son correctas. Además, yo creo que esas revistas se irían quedando...a no ser que esas revistas sean las de mayor impacto, que esas son siempre las más buscadas, pues se irían quedando un poco relegadas con unas condiciones que no tienen el resto de las revistas” E023

“Pero claro tú eso no lo ves hasta el final, tú no decides qué revista lo vas a enviar hasta el final, igual tú lo mandas a una revista y te dicen que no, e igual tienes que intentarlo con otras. Entonces claro, si una revista lo pide, lo más normal es que lo pidiesen todas, o si nadie lo pide, que no lo pida ninguna. Lo ideal sería que lo pidiesen todas. Bruto, analizado y gráfico, un poco toda la secuencia. Y ya está, y al final todos vemos todo el proceso” E005

“O sea, no tenemos una estructura superior que nos mande. Aquí hay unos cuantos laboratorios de odontología, y cada uno hace lo que le da la gana, no compartimos nada. En el Ministerio de Sanidad todavía se podría hacer más fácil, el dueño de todos los hospitales es el Ministerio de Sanidad. Es más fácil que diga “los datos se van a compartir”” E017

3. Actitudes

En cuanto a las actitudes, se pregunta a las personas entrevistadas fundamentalmente sobre dos aspectos. El primero, sobre las actitudes en sí hacia el hecho de compartir o

no compartir datos, es decir, la disposición hacia la práctica. El segundo aspecto sobre el que se pregunta es el cómo les gustaría que los datos fueran compartidos desde el punto de vista metodológico. Es decir, si se vieran en la situación tanto de compartir sus datos como de utilizar los datos compartidos de otros, cómo creen que los datos deberían ser presentados.

3.1. Disposición a compartir datos

Aunque en la mayoría de los casos las apreciaciones relacionadas con la disposición a compartir surgen de manera espontánea en algún momento de la entrevista, también se pregunta al final de la misma de manera intencionada para abordar este tema en concreto. Se considera importante que sea al final, ya que se contraponen las respuestas con el discurso sobre las creencias. A este respecto se observa que, en la gran mayoría de las entrevistas, sin distinción entre básicos y aplicados o júnior y sénior, las creencias sobre las ventajas de la práctica de compartir datos iban en una dirección y las actitudes, en otra. Es decir, que, aunque pensasen que los datos compartidos reportan grandes ventajas para ciencia e incluso las numerasen y detallasen, cuando aterrizaban a su caso concreto, a una hipotética situación en la que ellos se vieran en la posibilidad de compartir, sus actitudes solían ser recelosas en algunos casos y con varias condiciones en la práctica totalidad de las entrevistas.

En el caso del personal investigador básico, las respuestas siguieron un patrón muy similar. Prácticamente todos mostraban una mayor disposición a compartir después de haber publicado el trabajo correspondiente a esos datos y no antes de la publicación o aceptación final del trabajo (por ejemplo, cuando el artículo está en revisión, pero no aceptado) o durante (el transcurso de la investigación). No obstante, sobre esta base, surgieron algunos matices.

Por un lado, hubo investigadores con una actitud más tajante sobre este tema. En este caso, la tendencia era más proclive a no estar dispuestos a compartir a no ser que, o bien ya hubieran publicado todo lo que tuvieran planificado publicar, o se planteara

una colaboración en la que ambas partes (el que ofrece los datos y el que los solicita) obtengan un beneficio.

“Entonces, hay que reconocer que yo los datos, eso también te lo digo, solo los soltaría una vez que ya está la publicación, si no ni de casualidad.” E001

“Que sea una colaboración. A ver, no es pensar mal, pero... ¿para qué quieres mis datos si no es para colaborar conmigo?” E006

“Si es algo nuevo, yo creo que no compartiría si no obtengo un beneficio científico claro. Si me piden algo que no está publicado, solo compartiría con unas condiciones. Porque son datos que tú has generado, que has sufrido muchísimo, como para dárselo a alguien sin que exista ningún beneficio” E011

“Intentaría colaborar con ellos, y si podemos publicar juntos mejor que mejor, pero no les daría mis datos así con toda la predisposición del mundo. Entablaría una negociación.” E010

Aunque la opción anterior fue la mayoritaria, otros investigadores mostraron una posición más ambigua según la que, a pesar de que también abogaban por compartir después de publicar o por establecer colaboraciones, también se abrían a otras posibilidades. Estas posibilidades incluían opciones como ceder los datos a cambio de que se reconozca el origen de los mismos, solicitar permiso al IP, firmar un acuerdo para proteger la autoría, o poder hacerles seguimiento a los datos.

“Requisitos yo creo que lo estándar, que se reconozca de dónde vienen los datos y ya está. Y si es un grupo que está interesado en establecer alguna colaboración pues genial, pero que no sería un requisito imprescindible. Que se cite de donde vienen los datos y ya está.” E009

“Me colocaría más hacia ningún problema. Pero sí que es verdad que siempre y cuando lo preguntara a mi IP. Yo no tendría ningún problema, mi disponibilidad es total. Pero tendría que preguntarlo

como filtro. Aunque en mí caso no tendría ningún problema con mi IP. Pero otra gente no te podría decir lo mismo.” E013

Cabe mencionar que las personas entrevistadas más habituadas a compartir datos, por ejemplo, en plataformas como GEO, ya dan por sentado que los datos después de la publicación van a estar disponibles, porque es parte de los requisitos que las revistas les piden para publicar. Y que, por tanto, el único dilema es el hacerlo antes y durante o solo después. Esto no sucedía de manera tan evidente con los que comparten de manera esporádica.

“Entonces les diría que sí, que por supuesto. Si estuvieran subidos al GEO les daría el link y si no se los pasaría yo de alguna manera, pero siempre teniendo en cuenta que me van a citar.” E010

En el caso del personal investigador aplicados, aunque la experiencia se solía limitar a compartir entre pares o en redes de colaboración de acceso restringido, las respuestas con respecto a la predisposición para compartir datos son similares a las del personal investigador básicos en algunos aspectos. El primero de ellos sería la tendencia que se detecta a mostrar más disposición a compartir datos después de haber publicado:

“Posteriormente me daría igual, claro, podría estar en abierto. Pero antes lo veo complicado, por el tema de la competencia que hay y todo eso. Lo veo más factible entre grupos de gente que, aunque estén en sitios diferentes, puedan trabajar en red.” E022

Por otra parte, cuando se les pone en el caso de qué harían si alguien les solicitara compartir sus datos antes de haber publicado, muchos se plantean la posibilidad de aceptar, pero siempre con condiciones. Estas condiciones, compartidas por juniors y seniors, giran sobre todo en torno a saber con quién se van a compartir los datos, es decir, información detallada sobre el investigador o investigadores que van a recibir esos datos. Para todos las personas entrevistadas sería muy importante conocer quien o quienes serían los destinatarios de sus datos y poder establecer alguna trazabilidad sobre los mismos. Es decir, saber qué se va a hacer con los datos y poder supervisarlos de alguna manera. Esto significa que las personas entrevistadas de investigación aplicada no les gusta la posibilidad de perder el control sobre las personas que

entrarían, consultarían y descargarían sus datos. Además, también manifiestan que les gustaría conocer el proyecto para que el que se están solicitando los datos y tener la posibilidad de colaborar. En general, asumen que el uso compartido de datos es una práctica que, para que para ellos tenga sentido, debe ser un proceso lo más bilateral posible, lo que significa que no sea dar los datos a cambio de nada, sino que conlleve una negociación para establecer colaboraciones, compartir autorías o recibir algún otro tipo de reconocimiento.

“Estar informada de lo que se hace, saber para qué. Si puedo participar, mejor, porque a mí me gusta hacer cosas nuevas, entonces si están haciendo alguna cosa nueva y van a necesitar mis datos, me gustaría participar en ese proyecto.” E015

“Pues un intercambio. Vale, pues yo te cedo los míos porque a ti te interesan, porque los he recogido de una manera que a ti te interesa explotarlos... pues cuéntame qué tienes tú, porque igual tienes algo que a mí me encantaría explotar de alguna manera. Entonces, no exigiría publicación, pero sí que diría... cuéntame obviamente los resultados, porque me interesa, porque seguro que si yo lo he recogido quiero saber sobre lo que está saliendo de esos pacientes de los que yo he recogido datos, o sea, estar informada.” E016

“Sí, tengo que saber quiénes son estos investigadores, si no los conozco es más complicado. Es mejor si los conozco y hay unas colaboraciones previas, un respecto intelectual científico, en el sentido de que me parecen personas que tienen una trayectoria sólida, que van a contribuir a un proyecto serio que va a salir adelante.” E004

“Pero todo esto siempre y cuando no pueda acceder a esa información cualquiera, y se garantice la confidencialidad de esa información.” E021

“Pues bueno, quizás una publicación... o bueno, si es un grupo de investigación que ya vamos a colaborar con ellos, pues que, si yo comparto información, también la compartan ellos. Pero si se da la situación de que me piden la información porque sí, porque tiene que ser pública y la queremos, pues no... A parte que si quieres hacer una investigación con pacientes tienes que clarificar muy bien el objetivo, que es lo que quieres medir, y tienes que pasar por el comité ético del hospital. Y eso la verdad es que es costoso, es un esfuerzo y esperaría que tuviera algún tipo de contraprestación” E023

Finalmente, muchos coinciden en que necesitarían bien la aprobación por parte de los IP (en el caso de los júnior), o de la propia institución (en el caso de los sénior), debido sobre todo a la propia naturaleza de los datos (por su sensibilidad o protección de los pacientes/informantes) y a los compromisos de confidencialidad previamente firmados.

3.2. Cómo deberían de ser presentados los datos

Cuando se les pregunta a las personas entrevistadas sobre este aspecto, las respuestas fueron en la misma línea tanto en básicos y aplicados, como en júnior y sénior. Prácticamente todos hacían referencia a la necesidad de que los datos estuviesen muy bien ordenados, con explicaciones detalladas, sobre todo del material y método utilizado. La diferencia se halló en el nivel de detalle y especificación, ya que algunos se limitaban a mencionar aspectos generales de cómo les gustaría que fueran presentados los datos, y otros profundizaban en elementos más específicos. De nuevo, los aspectos más concretos o específicos partían de las personas entrevistadas con más experiencia en el uso compartido de datos.

Dentro de estos aspectos más específicos, se destaca lo siguiente:

- a) La mención a que los datos realmente se presenten sin haber realizado ningún análisis sobre ellos:

“Estuvieran lo más crudos posible (los datos), lo menos analizados posible. Sí, porque en bioinformática hay muchos pasos de preprocesado de los datos, de normalización, de limpieza...donde según lo que quieras medir, hacer un procesamiento o haces otro.”

E010

- b) Que haya una correlación clara entre los datos brutos y el resto de contenido:

“Pues me gusta que estén bien anotados, que se sepa qué es qué en cada caso, y como te comentaba antes, a veces es complicado. Muchas veces, incluso teniendo el dataset y teniendo el artículo, es muy complicado determinar qué es qué, porque las co-variables tienen nombres diferentes, o por lo que sea...pues sí, que estuviese todo muy bien anotado y que además se especifique muy bien la información que está disponible en el artículo, que estuviese todo muy bien enlazado para que se pudiese identificar bien qué es qué.”

E009

- c) Y que, además de dicha correlación, los datos, así como la información relativa a ellos (material y método en general, explicación de las variables) estuviera lo más limpia y estandarizada posible:

“Pues que los datos estuvieran en un Excel sencillo, que pusieran, depende, hablando del caso que estamos hablando, pues como lo han hecho. Tú siempre buscas un poco la facilidad para ver cómo son los datos. Pues muy claramente como son las muestras, y muy claramente cuál es el valor de cada una de las variables. Con eso, sería suficiente en el tipo de datos que estamos hablando. Si me lo plantearan así, en un Excel sencillo sería genial porque además que ya he trabajado con este tipo de datos, la información que le podría sacar sería mucho más rápida y sencilla. Yo he visto algunos que te va a costar más tiempo entenderlo que la utilidad que les puedas sacar.”

E014

“Me gustaría que hubiera una descripción de variables, que hubiera una etiqueta que describe bien esa variable, porque si tú no la conoces, a veces le damos nombres a variables que luego viene otra persona externa y no tiene ni idea, no lo entiende. Entonces, para mí, una etiqueta es súper útil que esté porque te clarifica un poco esa pregunta. Que los valores estén dados, que muchas veces me he encontrado bases de datos de compañeros en las que no están los valores dados.” E015

“Si hubiera una forma de estandarizar eso pues sería fenomenal, de cómo hacer las leyendas, es decir, este tipo de variables como categorizarlas, con números, o yo que sé... es decir, lo ideal sería que lo hiciéramos todos parecido. Si esta variable es así como la registro, cómo registrar a hombres y mujeres, cómo registrar...no sé. Y que todas las características que fueran con códigos de colores, o de números... Cuanto más sea rutina, que todos lo hagamos igual, más fácil es luego fusionar los datos.” E008

“Un Excel donde cada campo está explicado en un diccionario de variables, donde cada código está explicado en un diccionario de códigos y la verdad es que la interpretación es muy sencilla. Es una presentación adecuada, además, con cada paciente con el número que le asigna esa base, un registro que no tiene ninguna vinculación con el paciente, y para nosotros luego es muy fácil cotejarlo con el archivo que tenemos nosotros” E023

Finalmente, mencionar también que, entre los que suelen compartir utilizando repositorios de datos, existía la percepción de que ciñéndose a las instrucciones de los repositorios y de las revistas ya se estaba haciendo correctamente. Y que son estos repositorios o revistas las responsables de que la información se organice y se solicite de manera correcta y ordenada.

“Lo habitual es que haya unas líneas...o sea, lo habitual es que haya unas bases de datos que cuando se crea, se crea con una serie de

líneas y de apartados. Hay que seguir eso. O sea, si creamos bases de datos nosotros al azar, pues no sé...yo te estoy hablando de repositorios que existen, nosotros nos metemos y cumplimos con las reglas del repositorio. Crear cosas así aisladas al final no sirve para nada, porque cada uno quiere distintas entradas y distintas salidas y al final eso es un lío tremendo. Nosotros usamos repositorios establecidos donde hacemos lo que todos hacen, metemos ahí los datos.” E017

“Si no hay nada especificado, eso sería culpa de la revista, porque normalmente cuando tú puedes acceder a esos datos, hay un Excel que explica qué es cada cosa, en el supplementary data. ¿Si no existe? Entonces es culpa de la revista, o de los revisores que no han hecho un buen corte. Pero bueno, yo con los papers que manejo no me ha pasado todavía. Si no lo entiendo, es porque yo no tengo conocimientos para entenderlo, pero no porque no esté bien explicado o porque falte información.” E020

Discusión

En este apartado se presenta la discusión en base a los resultados obtenidos. Para su mejor comprensión, se ha dividido en las tres dimensiones de análisis que se han utilizado en este estudio: los conocimientos y experiencias, las creencias y las actitudes. Además, se ha realizado una subdivisión en cada una de las dimensiones para discutir sobre temas específicos.

1. Conocimientos y experiencias

1.1 Experiencias

Tal y como se ha podido observar en los resultados, los conocimientos y experiencias distinguen a las personas entrevistadas en diferentes grupos. Por un lado, quienes no han compartido nunca con quien ha compartido alguna vez o lo hacen habitualmente; por otro, según el tipo de investigación y, por último, según los años de experiencia y la jerarquía.

En este sentido, en cuanto a los conocimientos y experiencias, el factor que resultó más determinante, el que más diferenciaba patrones, fue el tipo de investigación que llevaban a cabo. Se constata que el personal investigador de ramas más básicas suele compartir datos más a menudo, ya que la práctica totalidad manifestaban haber tenido relación con esta práctica. Mientras, el personal investigador aplicado tenía poca o ninguna experiencia.

Esto lleva a pensar que hay tipos de investigaciones que favorecen estas de prácticas y cabe preguntarse porqué. Si se atiende a la literatura científica, hay investigaciones que necesitan más que otras conocer el raw data que subyace al procesado de resultados (Sieber, 2015; Tenopir et al., 2011, 2015). Esto se observa claramente en este estudio, donde varios entrevistados manifiestan necesitar saber cómo lo han hecho otros para no duplicarse, o simplemente dan por hecho que es una práctica que tienen que llevar a cabo, otra más entre sus tareas. En áreas relacionadas con la

genética, esto se observa desde hace años debido a técnicas que generan grandes volúmenes de datos, como es la secuenciación masiva. En este sentido, cabe destacar que no es que no exista ningún tipo de discusión sobre la práctica, sí que existe, se trata más bien de que la discusión se encuentra en estadios más avanzados si se compara con otras áreas (Faure y Joly, 2016; Hernández, Quijada, Rodríguez-Lázaro, y Eiros, 2020; Pereira, Gibbs, y McGuire, 2014). En este sentido, cabe señalar un segundo nivel de diferenciación dentro de los básicos, ya que los que más frecuentemente habían tenido contacto con la práctica eran quienes pertenecían a laboratorios, empresas, o centros de investigación privados que hacían uso de este tipo de técnicas. Otros investigadores, más relacionados con el trabajo en la universidad con estudios donde se utilizan otro tipo de diseños, como los modelos con animales, manifestaban conocer la práctica y haberla llevado a cabo alguna vez, pero sin estar tan arraigada.

En el caso de los estudios aplicados (mayormente de corte clínico), como se ha mencionado, la presencia de la práctica de data sharing es escasa, lo que se ve reflejado en los conocimientos que tienen sobre la misma. Al haber tenido tan poco contacto o ninguno, pocos conocían mecanismos específicos para compartir datos, como los repositorios, o las distintas modalidades para compartir datos que pueden solicitarse en revistas o proyectos. Este desconocimiento o falta de cultura concuerda con trabajos como el de Nowacki (2013), en el que se refiere que, a pesar de que sí existen sistemas de uso público para uso compartido de datos en ensayos clínicos, esta práctica todavía no es la norma. Es, además, interesante mencionar que existe literatura donde se vinculan los beneficios del data sharing con la práctica clínica, pero no tanto gracias a compartir datos clínicos sino a compartir datos genéticos que acaban teniendo beneficios en la clínica (Vassilakopoulou y Aanestad, 2019). Esto podría indicar que el personal investigador aplicado valora la práctica, pero no tanto para aplicarla ellos mismos, sino para otras disciplinas o como un bien general.

Sin embargo, a pesar de no ser una práctica común, sí que se halla alguna referencia sobre compartir datos en el ámbito clínico. Estudios como el de Wilkinson et al., (2019) muestran que existe esa inquietud, aunque, como estos autores señalan, no está exenta de complicaciones. Estos autores ponen énfasis en varios aspectos que son de interés. En primer lugar, defienden que, para que pueda darse el uso compartido de

datos provenientes de ensayos clínicos, es altamente necesario entender todos los aspectos de la investigación, lo que solo sucede cuando se proporciona información muy detallada sobre los estudios originales. Obtener esta información detallada sería, entonces, un primer paso en la dirección correcta para proporcionar al personal investigador la información necesaria para reutilizar conjuntos de datos de ensayos. No obstante, destacan que lo justo para ambas partes (el que comparte y el que recibe) sería que los dueños originales de los datos pudieran seguir teniendo vinculación con los mismos. De esta manera, se propone que las personas que vayan a reutilizar los datos puedan colaborar estrechamente con el personal investigador primario, permitiéndoles conservar la propiedad y el control de los usos de sus datos y, al mismo tiempo, evitando que se malinterpreten las complejidades y los matices del ensayo.

Una última sugerencia que hacen Wilkinson et al., (2019) es todavía más interesante. Y es que ellos se preguntan si no sería una solución más útil y sostenible establecer unos estándares para el intercambio de datos de ensayos clínicos. Toman la idea de las lecciones aprendidas en cuanto a estandarización de los datos de los registros de salud electrónicos, así como de otras recomendaciones hechas para los ensayos clínicos no comerciales. Según ellos, esto ayudaría a sortear muchos obstáculos relacionados con los datos y permitirá un intercambio de datos de prueba más utilizable y utilitario.

Otro aspecto muy relevante sobre el personal investigador aplicado que ha emergido de las entrevistas es la mención a las bases de datos de acceso restringido. Como ya se ha mencionado, el personal investigador aplicado apenas tiene conocimiento de la existencia de infraestructuras para compartir datos, como los repositorios. Sin embargo, muchos de ellos relacionaban este concepto con una herramienta que ellos sí utilizaban: las bases de datos restringidas. Cuando se les pidió que profundizaran un poco más sobre este concepto, surgieron descripciones de bases de datos de carácter dinámico, que se utilizaban para realizar estudios donde participaban varios investigadores o varios grupos. Varios de ellos (sobre todo los que trabajan en hospitales) hicieron referencia a estudios multicéntricos, es decir, que no solo implicaban a diferentes investigadores, sino también a distintas instituciones, sobre todo hospitales. En estas bases de datos, según sus descripciones, se accede mediante

usuario y clave y lo más relevante es, quizás, el hecho (señalado por todos) de que, aun así, no puedes acceder a la base de datos completa, solo puedes acceder a tus propios datos y a veces, ni eso. Es decir, no puedes acceder a la base de datos completa ni hacer uso de ella sin permiso expreso y una justificación clara, que a veces se debía pedir al resto de los participantes y otras, al IP del proyecto. Es importante destacar también que, en ocasiones, incluso debían pedir permiso para acceder a sus propios datos una vez depositados. Un ejemplo de algo que podría parecerse es la experiencia del Data Access Compliance Office del repositorio del International Cancer Genome Consortium (ICGC-DACO), que también establece un acceso controlado bajo la premisa de que todas las personas que hagan uso del mismo están comprometidas con el data sharing, y que el acceso controlado es útil para ofrecer mejores mecanismos de seguridad y calidad a datos genómicos y clínicos que son potencialmente sensibles (B. M. Knoppers y Joly, 2018). Sin embargo, la diferencia entre ambos modelos (el que explican las personas entrevistadas y el del ICGC-DACO) difiere en la voluntad del data sharing. Mientras que el ICGC establece el acceso para aumentar la seguridad con intención de proteger la privacidad de los sujetos que participan en los estudios, pero está pensada para investigadores con ánimo de compartir datos, las bases de datos controladas que refieren las personas entrevistadas parecen tener otra finalidad además de esta. Esto se deduce teniendo en cuenta que, tal y como refieren, no tienen acceso ni a sus propios sujetos, por lo que difícilmente se podría pensar que hay una voluntad de compartir. Con respecto a estas bases de datos de acceso restringido, aunque en general no se mostraba descontento con esta práctica, ya que las personas entrevistadas la aceptaban como normal, si hubo un entrevistado que no estaba completamente de acuerdo. Esta disconformidad venía dada por el hecho de no poder acceder a sus propios datos y que, al solicitarlos expresamente, se pusieran más trabas que facilidades.

1.3 Revistas

Otro factor que se ha considerado clave es el papel de las revistas. En el apartado anterior se discutía sobre cuáles eran las experiencias y conocimientos de las personas entrevistadas. Es decir, qué cosas hacían y no hacían en relación al data sharing, así como su nivel de conocimiento de la práctica. Sin embargo, cuando se trató el tema de las revistas y las publicaciones (sus exigencias, sus propuestas o sus normativas sobre el data sharing), es cuando se empieza a comprender por qué hay una diferencia a veces tan marcada entre unos investigadores y otros.

Como se ha visto en los resultados, las personas entrevistadas que habían publicado en revistas que pedían datos eran también los que estaban más acostumbrados a esta práctica. A más conocimiento de la exigencia/petición de las revistas, más normalizada estaba la práctica. Del mismo modo, a cuanta más experiencia, más se utilizaban formas complejas de compartir datos, como es el uso de repositorios especializados. Mientras, las personas que decían no haber compartido nunca o haberlo hecho rara vez, coincidían en decir que las revistas no se lo habían pedido, o que no tenían constancia de ello. Dado que las revistas constituyen el principal vehículo a través del cual se hacen públicos y se difunden los resultados de las investigaciones y cumplen un papel esencial para la ciencia, no es de extrañar que se haya dado esta situación (González Alcaide y Ferri, 2014). No sería descabellado pensar, en este sentido, que el posicionamiento que tengan las revistas y las editoriales sobre el uso compartido de datos tiene un papel esencial. Si además se añade el elemento “factor de impacto” y el “*publish or perish*”, no es sorprendente tampoco que el personal investigador se vea influenciado para compartir datos o no hacerlo por los mandatos y recomendaciones que hagan las revistas y las editoriales a las que éstas pertenecen (Miller, Taylor, y Bedeian, 2011). En este sentido, varias de las personas entrevistadas referían que, si se quiere publicar en revistas con impacto, y éstas te piden que compartas los datos, pues no hay duda de que hay que hacerlo. Esta relación no es novedosa, ya que se observa claramente en diversos trabajos previos sobre la relación entre revistas con elevado factor de impacto y la presencia de artículos donde se comparten datos (Aleixandre-

Benavent et al., 2018; Aleixandre-Benavent, Moreno-Solano, Ferrer Sapena, y Pérez, 2016; Sixto-Costoya et al., 2020; Vidal-Infer et al., 2019, 2018).

1.4 Proyectos

Sin embargo, en el extremo opuesto, se ha observado que ninguna persona entrevistada recordaba que le hubieran solicitado los datos para los proyectos en los que participaba. Aunque esto es bastante llamativo si atendemos a las políticas sobre datos abiertos a nivel Internacional, comunitario y nacional, ya que éstas promueven el uso compartido de datos, no lo es tanto si se lee la letra pequeña.

En el caso de los proyectos de la UE, enmarcados mayormente en los denominados Horizon 2020 (H2020), su postura sobre el uso compartido de datos es clara: se defiende y se promueve, aunque de momento no es un mandato obligatorio. Básicamente, lo que desde la UE se indica sobre uso compartido de datos sigue la misma tendencia que sobre el OA en general: debe hacerse por sistema y por defecto. Uno de los argumentos más destacados por su lógica es el siguiente: no debería ser necesario pagar por la información financiada con fondos públicos cada vez que se accede o se utiliza la misma. Sin embargo, aunque la tendencia es similar, todavía no es tan contundente para los datos como para las publicaciones, ya que se sigue la máxima de los principios FAIR de “tan abierto como sea posible, tan cerrado como sea necesario”. Por este motivo, al preguntar sobre el tema de los proyectos y el data sharing, algunas de las personas entrevistadas manifiestan haber tenido que poner en OA la publicación, pero no los datos, ya que lo primero es obligatorio pero lo segundo no (European Commission, 2019; Sixto-Costoya, Aleixandre-Benavent, Vidal-Infer, Lucas-Domínguez, y Castelló-Cogollos, 2019).

En el caso de los proyectos nacionales, se tomará como referencia el análisis publicado por SPARC Europe y el Digital Curation Center (DCC) (Digital Curation Center y SPARC Europe, 2019; Ministerio de Industria Economía y Competividad, 2017), ya que ilustra muy bien cuál es el panorama actual en cuanto a políticas relacionadas con el uso compartido de datos en los distintos Estados Miembros, entre ellos España. Según este

análisis, España se encuentra entre los 14 de 28 que tiene algún tipo de plan o normativa referente a los datos de investigación. La referencia a los datos en el contexto de la ciencia española se hace en el Plan Estatal de Investigación Científica y Tecnológica y de Innovación, que comprende desde el año 2017 al 2020. No obstante, se opta por la recomendación y siempre haciendo alusión a la opcionalidad (Digital Curation Center y SPARC Europe, 2019; Ministerio de Industria Economía y Competividad, 2017). Estas políticas que optan por la no obligatoriedad podrían explicar el porqué de la negativa de las personas entrevistadas, cuando se indaga acerca de si algún proyecto en el que han participado les pedía compartir los datos.

Otra cuestión que salió a relucir cuando se cuestionó a las personas entrevistadas sobre los proyectos fue al preguntar a los júnior por este tema. En este caso, muchos de los júnior manifestaron que ellos no llevan el control de los proyectos, por lo que no podían dar una respuesta clara a esta cuestión. No obstante, referían que, de haber habido alguna exigencia de compartir datos, probablemente se hubieran enterado.

Esto puede ser debido a que, tal y como comentaron la mayoría de los júnior que decían haberse encargado del trabajo que supone preparar los datos para compartirlos, así como subirlos a las distintas plataformas, cuando lo pedían las revistas, probablemente también habrían formado parte del proceso, de haberlo habido, en los proyectos.

1.5 Esfuerzo percibido

En estrecha relación con los mandatos y peticiones por parte de las revistas y de los proyectos, es importante destacar la apreciación del esfuerzo percibido. Muchas de las personas entrevistadas hicieron referencia al esfuerzo que costaba todo el proceso de preparar los datos para que éstos pudieran ser compartidos, así como el manejo de las plataformas, en el caso de usarlas. Esta apreciación coincide con el estudio de Kim (2017), que también ha hallado en sus resultados (en este caso procedentes de un estudio cuantitativo), que el personal investigador en general percibían el hecho de compartir datos como un esfuerzo extra. En este sentido, muchas de las personas

entrevistadas referían una cierta sobresaturación de tareas que a veces es complicado de gestionar. Es interesante volver a destacar aquí la aportación del entrevistado E008, donde se relacionan el esfuerzo percibido con los incentivos: “Si te lo ponen como una traba más, como un obstáculo más, pues se hará a regañadientes porque no te queda más remedio (...) Podrán decidir si quieren más látigo o más zanahoria, pero si se puede poner zanahoria siempre es mejor”. Se podría decir que la valoración del esfuerzo en positivo o en negativo depende de la recompensa que se pueda percibir. Esta recompensa puede venir de muchas formas, que van desde considerar que se está haciendo lo mejor para la ciencia, el hecho de recibir más citas, o tener la posibilidad de colaborar con otras personas. Aunque se retomará el tema del esfuerzo/recompensa en los apartados de “creencias” y “actitudes”, es importante dejar constancia en esta sección de “experiencias”, ya que la percepción del esfuerzo forma parte de las experiencias vividas por las personas entrevistadas.

También dentro del esfuerzo percibido y de las experiencias relatadas por las personas entrevistadas, destaca la experiencia de algunos entrevistados junior, en este caso de los que sí habían tenido contacto con la práctica. Éstos reconocían que eran ellos los que llevaban a cabo el proceso de curado y preparación de los datos para ser depositados/subidos a donde la revista indicase. En este sentido, manifestaban (unos con mejor y otros con peor talante) que se habían tenido que aprender el procedimiento para compartir datos, en las diversas modalidades y plataformas que existen. Incidían en que había resultado un esfuerzo primero entender lo que se estaban requiriendo y después, aprender a llevarlo a cabo. En el caso de las plataformas como GEO, las instrucciones y su manejo no son siempre sencillas y los propios entrevistados junior reconocían que probablemente fuera más fácil de aprender para ellos que para sus jefes. De esto se deducen varias cuestiones. La primera, que a pesar de que prácticamente todas las personas entrevistadas manifestaron que la práctica de compartir datos era un trabajo extra, un esfuerzo que se une a todas las demás tareas y que es más frecuente que la asuman los investigadores en sus primeras etapas.

2. Creencias

2.1 Sobre las ventajas

Las ventajas, los puntos a favor de compartir datos, son comunes en todas las personas entrevistadas. Es decir, tanto cuando se les cuestiona sobre las ventajas en concreto, como cuando surge el tema de manera espontánea, los puntos positivos de compartir datos coinciden en gran medida entre las personas entrevistadas. Estos puntos fueron descritos ampliamente en la literatura científica sobre este tema y giran en torno a factores como la transparencia y la prevención de fraudes, la reproductibilidad, la calidad, la reutilización, el evitar la repetición de experimentos similares, la generación de nuevas hipótesis, la estimulación de descubrimientos adicionales, las comparaciones de series de datos, la posibilidad de sumar esfuerzos y la opción de abrir nuevas vías de colaboración (Aleixandre-Benavent et al., 2014b; Aleixandre-Benavent, Ferrer-Sapena, Alonso-Arroyo, y Peset, 2013; González, Saorín, Ferrer-Sapena, Aleixandre-Benavent, y Peset, 2013).

En este sentido, las personas entrevistadas seguían la línea de lo que se supone que son los preceptos y la finalidad del data sharing. Es decir, estaban de acuerdo con que compartir datos tiene la capacidad de mejorar los procesos, ahorrar recursos y, en última instancia, sería beneficioso para la ciencia en su conjunto y para la sociedad.

Esto se puede interpretar como que realmente la cultura de compartir datos, y la cultura del acceso abierto en general (que defiende todos estos preceptos), ha ido calando de alguna manera en las percepciones del personal investigador (Max-Planck-Gesellschaft, 2003; San José Montano, 2009). Después de veinte años desde las llamadas tres “B” (Declaración de Budapest, Declaración de Bethesda y Declaración de Berlín), las personas que hacen ciencia parece estar de acuerdo en que cuánto más “open”, mejor será para la ciencia. Aunque también cabe la posibilidad de que hayan expresado lo que es políticamente correcto (que la ciencia es un bien común, que todas las personas tienen derecho a beneficiarse de la misma, que la financiación con fondos públicos debe implicar más apertura de los resultados, etc.), lo cierto es que el movimiento Open Science y, dentro de éste, el de Open Data ha ido cobrando cada vez

más calado, tal y como se observa en diversos estudios basados en cuestionarios sobre las percepciones del personal investigador sobre el data sharing (Aleixandre-Benavent et al., 2020; Zhu, 2020).

Según los resultados de este estudio, se ha podido obtener una idea basada en la práctica diaria de por qué considera útil el personal investigador el compartir los datos de investigación. Estos ejemplos prácticos se obtuvieron de quienes han tenido algún contacto con el data sharing, bien de manera frecuente o de manera más ocasional. Como ya se ha mencionado, fue el personal investigador con líneas más básicas quienes mayormente habían tenido contacto con los datos que son compartidos de forma más abierta a la comunidad científica (por ejemplo, ofreciendo el número o link de acceso a un repositorio donde están depositados los datos, o bien como material suplementario a las publicaciones), mientras que el personal investigador aplicado tenían más experiencia compartiendo en bases de datos cerradas (con clave de acceso) y en el *peer to peer*.

Entre el personal investigador básico, éstos argumentaban que el poder acceder a los datos brutos generados en otros estudios les daba la oportunidad de informarse antes de comenzar un estudio ellos mismos, para saber lo que se había hecho y tratar de no repetir procesos, de la misma manera que sucede cuando se buscan otras publicaciones antes de hacer una nueva, para saber qué lo que ya hay y no solaparse. Además, mencionaron también la utilidad de consultar los datos brutos de otros no solo para no solaparse, sino también para conocer qué se hecho sobre algún tema y partir de un conocimiento más amplio para elaborar sus propios estudios; siguiendo el paralelismo de las publicaciones, sería lo mismo que cuando se consulta bibliografía sobre algún tema para formarse una idea acerca del mismo. Por otro lado, es interesante destacar que uno de las personas entrevistadas también manifestó consultarlos para su reutilización, ya que él no podía obtener por sus propios medios (debido a limitaciones tanto técnicas como humanas y materiales), lo la cual constituye uno de los pilares fundamentales sobre el que se sostiene el movimiento de Open Access: la reutilización. De manera general, todas las personas entrevistadas destacaron el hecho de que compartir sus datos, u observar que otros investigadores los comparten junto con sus publicaciones, otorga credibilidad a los estudios, los

ratifica y, de alguna manera, eleva su calidad y despeja dudas sobre los orígenes de la información procesada que se presenta en los resultados.

En el caso del personal investigador aplicado, es interesante comenzar, a modo de facilitar la comprensión sobre las diferencias entre las perspectivas, por el punto de vista de uno de las personas entrevistadas (E017). Esta perspectiva es muy ilustrativa, y sirve como una especie de puente, ya que se trata de un investigador al que no ha sido posible clasificar en un tipo de investigación (básica o aplicada), ya que participa de las dos de manera traslacional. Es interesante porque mientras afirmaba con total convencimiento la importancia de compartir los datos brutos de las investigaciones básicas en los repositorios que se recomendasen en las revistas, y, además, afirmaba que eso es lo que había que hacer cuando se quiere publicar en revistas con alto impacto (según índices de calidad como Journal Citation Report) y lo tenía totalmente normalizado, mostraba una posición muy distinta cuando se trataba de datos clínicos. En este caso, argumentaba que, al tratarse de investigaciones con pocos sujetos, no existía esa necesidad y que, en su opinión, no creía que hubiera repositorios especializados en este tipo de datos. Esta percepción sobre los datos clínicos ya ha sido documentada en la literatura (Shahin et al., 2020; T. Wilkinson et al., 2019), por lo que no resulta una novedad la perspectiva del entrevistado. Sin embargo, no deja de ser llamativo cómo un investigador que ha integrado la práctica de compartir datos, no es capaz de visualizar la utilidad de hacerlo con otro tipo de investigaciones. Cabe preguntarse, a este respecto, hasta qué punto los preceptos generales de las ventajas sobre el uso compartido de datos necesitan ser aterrizados a las necesidades y particularidades, no solo de cada disciplina, sino también de cada manera de hacer ciencia dentro de cada disciplina.

Si se centra la atención en el personal investigador mayormente aplicado, como se ha dicho, sus prácticas de compartir información bruta o de consultar la de otros, se circunscribe a bases de datos cerradas, normalmente limitadas a proyectos concretos, o al *peer to peer*. En este caso, su falta de experiencia en la práctica del data sharing en repositorios o en otras modalidades como el material suplementario dificultaba que se pudieran poner ejemplos concretos de su utilidad, más allá de su uso en estos grupos cerrados o entre colaboradores que ya se conocían con anterioridad. Aunque en

algunos casos, sobre todo en los que las muestras son pequeñas y difíciles de conseguir, el data sharing sí que podría ser una herramienta útil, así como que ciertamente beneficia la transparencia, en general el personal investigador aplicados, bien por desconocimiento, por saturación de otras tareas, como por otros miedos relacionados con privacidad y autoría (que se detallarán más adelante) se mostraban bastante reticentes.

No obstante, estos miedos, inseguridades y reticencias también se mostraron, a pesar de tener más formación y práctica, en el personal investigador básico. A continuación, se tratan estos aspectos, que se mostraron muy relevantes, con más detalle.

2.2 Sobre los inconvenientes y los miedos

Como se ha comentado en el anterior apartado, a pesar de que prácticamente todos reconocen lo que se consideran ventajas y puntos positivos de la práctica de compartir datos, algunos de manera más teórica (investigadores aplicados) y otros de forma más práctica (investigadores básicos), también surgieron dudas y miedos que, cuando los expresaban, las personas entrevistadas los detallaban con más profundidad que cuando hablaban de las ventajas. Esta realidad también fue observada en el estudio llevado a cabo por Bezuidenhout (2019) también cualitativo y a través de entrevistas semiestructuradas, pero en el contexto de varios países africanos. En su trabajo, Bezuidenhout destaca que efectivamente existe una especie de grieta o distancia entre la consideración de lo bueno que es compartir datos y las complicaciones, problemas y rechazos que puede generar. Ella denomina a esta dualidad *“the ideal and the real”*, que viene a significar que una cosa es lo que sería una situación ideal donde se aceptan y se llevan a cabo sin reticencias todos los aspectos positivos del data sharing comentados en el anterior apartado, y lo que es real, es decir, lo que es factible teniendo en cuenta las características sociales, políticas, económicas, geográficas y culturales de la ciencia en general y de los científicos en particular.

En el transcurso de las entrevistas, al entrar en temas relativos a estos factores de miedo y reticencias, lo primero que se detecta (y que se va confirmando a medida que

se van haciendo más entrevistas) es que existe algo que les preocupa que va más allá del compartir o no datos de investigación. Conceptos relacionados con la alta competitividad, la dificultad de acceso a los recursos y a la financiación, el elevado volumen de trabajo y el sistema de incentivos fueron recurrentes a la hora de contraponer las bondades de compartir datos con las preocupaciones y las reticencias.

En líneas generales, los miedos o desventajas (la mayoría de las veces entremezcladas) tenían que ver con temas de autoría, con el esfuerzo dedicado, las dificultades técnicas y de formación, la pérdida de ventaja competitiva y pérdida de control sobre tu trabajo, miedo a la mala interpretación que otras personas puedan hacer de los datos o de problemas de confidencialidad, temas que, igual que las ventajas y ya fueron tratados con bastante detalle en la literatura (Beno, Figl, Umbrich, y Polleres, 2017; Navarro-Molina y Melero, 2019). Sin embargo, lo que se ha comprobado en este estudio es que la importancia que cada investigador le daba a estos factores variaba en función del tipo de investigación que realizara (por ejemplo, investigadores aplicados más preocupados por la confidencialidad de los datos de los pacientes, o investigadores básicos más preocupados por la mala interpretación que otros puedan hacer de sus datos), o de sus años de experiencia y cargo ocupado (más tendencia entre los júnior a asumirlo y normalizarlo entre sus tareas y más tendencia entre los sénior a preocuparse por la pérdida de competitividad en los proyectos). Esta percepción de aceptación general al data sharing como idea o como filosofía, pero con reticencias y miedos significativos cuando se explora con un poco más de detenimiento, lo han observado también Linek, Fecher, Friesike, y Hebing (2017) en su trabajo sobre la influencia de la personalidad del personal investigador a la hora de compartir datos. Estos investigadores, que exploran desde una perspectiva psicológica las opiniones y percepciones del personal investigador para compartir datos, concluyen que el grado de implicación en el data sharing puede tener que ver con factores como la edad, el género y la personalidad. Para llegar a esta conclusión, basan su metodología en el marco de adecuación para la toma de decisiones en dilemas sociales. En esta línea, argumentan que las decisiones individuales sobre el intercambio de datos se basan en la percepción subjetiva de la situación del intercambio de datos en la academia, la identidad del investigador y la aplicación de

reglas y heurísticas individuales. Su interés en valorar si la personalidad del personal investigador influye en el intercambio de datos tiene que ver con el hecho de que se parte de la base de un sistema académico que puede caracterizarse como una economía de reputación.

Por otra parte, el trabajo de Inkpen et al., (2020) también refleja una reflexión basada en su punto de vista sobre la promoción del data sharing en los últimos años, que sirve para dar un marco contextual que ayuda a entender las reticencias a las que hacían referencia las personas entrevistadas. En su caso, los ejemplos se centran en áreas relacionadas con la Geografía ya que es su campo de trabajo, pero podrían aplicarse a cualquier otra área. Desde la experiencia que otorga el hecho de pertenecer a un área donde el data sharing ha estado a la orden del día en los últimos años, focaliza su discurso en los procesos de desigualdad y de relaciones de poder que hay detrás de la promoción del data sharing. Ellos argumentan que hay un componente de desigualdad que no se está teniendo en cuenta cuando se le pide al personal investigador que comparta sus datos. Para su razonamiento, parten de una contextualización de la ciencia y la actividad académica dentro de la economía capitalista. En este sentido, argumentan que es importante entender la actividad académica dentro de una dinámica neoliberal donde se valora sobre todo el beneficio individual por encima de lo colectivo y donde no se fomenta el establecimiento y fortalecimiento de vínculos de cooperación, sino de competición. Además, ponen en valor el hecho de que la generación de datos brutos conlleva una extracción de información emocionalmente agotadora y un desafío intelectual, que se suele llevar a cabo en una realidad a menudo desordenada y poco cooperativa. Este coste-beneficio fue destacado también en las entrevistas, donde recurrentemente se hacía referencia al esfuerzo que cuesta extraer datos para luego compartirlos “sin más”.

A este último respecto, a pesar de lo mencionado en el apartado sobre las ventajas, donde se explicaba que, en su mayoría, las personas entrevistadas coincidían en los puntos positivos de compartir datos con aquellos documentados en la literatura, luego ninguno refería compartirlos espontáneamente, de forma totalmente desinteresada. Todos tenían algún condicionante, que normalmente era el que la revista lo exigía para la publicación. En el caso de los júnior, otro condicionante también era que lo que ellos

quisieran hacer (compartir o no) poco importaba, ya que al final iba a depender de lo que sus superiores quisieran hacer, perspectiva que concuerda con Gewin (2016). Aunque del personal investigador júnior entrevistado ninguno mencionó que sus superiores les hubieran obligado a compartir datos en contra de su voluntad, sí que mencionaban que cuando tocaba hacerlo, ya sea por exigencia de una revista o cualquier otra razón, la tarea recaía sobre ellos. Esta situación era asumida con normalidad y achacada principalmente a dos factores: el conocimiento técnico y el sistema de reparto de tareas. En cuanto al conocimiento técnico, esto tenía que ver sobre todo con el cambio generacional, ya que los júnior asumían que estas exigencias, que conllevan una formación sobre el proceso y sobre las plataformas (subir los datos, establecer los metadatos, conocer los repositorios) lo iban a llevar ellos más fácilmente que los IP, que normalmente son de mayor edad. Sobre el reparto de tareas, esto tiene que ver con que, a menos estatus en la jerarquía académica, más probabilidades de encargarse de tareas más engorrosas o complicadas que a priori no reportan ningún beneficio, pero que son necesarias. Desde el punto de vista de Popkin (2019), trasladar esta tarea a el personal investigador joven o con menos experiencia y, por tanto, con menos poder de decisión, puede ser un error. Desde su perspectiva, sería mucho más práctico que las universidades o cualquier centro de investigación, contrataran a especialistas en tratamiento y curación de datos, antes que pretender que el personal investigador de menor rango se forme a medias sobre este tema y acaben asumiendo este trabajo. En cuanto a lo que la literatura refleja sobre lo que piensan el personal investigador júnior y el uso compartido de datos, se relaciona fundamentalmente con dos perspectivas, que normalmente se entrelazan: o bien se postula como un elemento ventajoso que aporta visibilidad y posibilidades de colaboración; o como lo contrario, una práctica que les pone en un aprieto al tener que exponer su material bruto a riesgo de que se lo puedan copiar y utilizar ganándoseles ventaja competitiva. En su trabajo, Gewin (2016) hace referencia a esta dualidad basada en que el personal investigador joven siente que compartir datos es beneficioso para su carrera – les da miedo que puedan perder ventaja en un mundo altamente competitivo. En este sentido, Gewin destaca que el personal joven debe hacer equilibrios entre tener en consideración las ventajas que ofrece el data sharing, como las oportunidades de recibir más citas y ganarse una reputación más allá de las métricas tradicionales, y la

posibilidad de volverse más vulnerables. En esta línea, habla sobre las repercusiones de compartir, especialmente en las primeras etapas de la carrera investigadora, haciendo referencia a la preocupación recurrente de que “todo el mundo conoce a alguien a quien le han fastidiado con el tema de los datos”. Esta alusión a historias que atemorizan sobre hacer uso de datos de otros ocasionando problemas al creador original, también se pudo detectar en alguna ocasión durante las entrevistas. A pesar de que este temor esté basado únicamente en “algo que le pasó a alguien”, realmente tiene calado y es necesario tenerlo en cuenta.

Estas contradicciones del personal investigador con menos años de experiencia se plasman perfectamente en los resultados del estudio de Tenopir et al., (2015), basado en un encuesta a investigadores y profesores con un carácter multidisciplinar. Según sus resultados, los encuestados más jóvenes son más favorables al intercambio y la reutilización, pero, sin embargo, ponen menos datos en abierto que los encuestados con más edad. Además de esto, si se compara con los encuestados de más edad, los jóvenes expresan más interés en utilizar los datos de otros y también en compartirlos, pero solo si pueden poner ciertas condiciones, como son obtener coautorías, tener oportunidad de colaboraciones, tener la oportunidad de revisar los resultados de otros cuando utilizan sus datos y hacer sugerencias al respecto, y garantías sobre los permisos legales. Estas condiciones concuerdan totalmente con lo que se ha observado a lo largo de las entrevistas que se han realizado para esta tesis, especialmente las relacionadas con las coautorías, colaboraciones y la supervisión de los datos. Sin embargo, habría que añadir el matiz diferenciador del personal investigador básico y aplicado. Mientras que los básicos mostraban un claro interés por ser los primeros en publicar, pero, una vez publicados, se mostraban más favorables a “desprenderse” de los datos y que estuvieran a libre disposición de otros; los aplicados hacían mucho más hincapié en mantener el control sobre los mismos y, sobre todo, en tener la posibilidad de revisarlos.

Otro dato peculiar que encontraron fue que, a pesar de que el personal investigador más joven estaba más interesados en compartir sus propios datos que los mayores, también lo estaban más en exigir que les tengan que solicitar una serie permisos para proporcionarlos. Además, se observó que los jóvenes tenían una mayor preocupación

por tener ellos la capacidad de ser los primeros que extraen resultados para una publicación, mientras que los mayores estaban más preocupados por temas de financiación. Esto último también se ha observado claramente en esta tesis, sin distinción además entre tipos de investigación: el personal investigador con más años de experiencia mostraba especial preocupación por el asunto de la financiación y de la concesión de proyectos. Esta preocupación abarcaba varios temas que, además, tenían cierta relación con las propias etapas de los proyectos: el antes de pedir, el durante y el después. En primer lugar, la ventaja competitiva que se tiene que tener para la obtención del proyecto en un contexto donde se prima la meritocracia y donde el acceso a la financiación es especialmente complicado y, como se ha dicho anteriormente, no se prima precisamente la solidaridad y la cooperación. En segundo lugar, una vez que se concede un proyecto, la reivindicación de que los miembros de ese proyecto sean los primeros en publicar resultados obtenidos de los datos que hayan sido generados en dichos proyectos, sin posibilidad de que, por haber compartido los datos brutos, otros se puedan adelantar. En tercer lugar, los propios compromisos adquiridos con las entidades financiadoras, en el sentido de que, ya que les había financiado, una vez extraídos resultados, no podían ponerlos en abierto al menos hasta que se cumplieran todos los requisitos que estuvieran estipulados en la convocatoria, sobre todo cuando lo que estaban involucrados eran o bien datos de humanos por tema de confidencialidad, o bien datos relacionados con patentes.

En referencia a las barreras que los investigadores y las investigadoras detectan cuando se aborda el tema del data sharing, el Institute of Medicine (IOM) de EEUU organiza en el año 2012 un workshop donde el objetivo era reunir a un grupo de agentes clave en la investigación biomédica, especialmente focalizado en la clínica, formado por empresas farmacéuticas, investigadores académicos, diferentes agencias gubernamentales e Institutos Nacionales de Salud, para indagar acerca de sus posturas sobre el uso compartido de datos e intentar llegar a acuerdos y puntos de vista comunes. De este workshop, surgieron una serie de resultados y planteamientos que son muy clarificadoras y muy útiles y que se plasmaron en un libro (Olson y Downey, 2013). Muchos de estos resultados están en línea con lo que se detectó en las entrevistas para este trabajo. Entre las secciones que abordaron, se encuentran las

barreras o reticencias que los actores reunidos encontraron para el data sharing y, más específicamente, éstas se dividieron en barreras legales, barreras culturales y barreras técnicas. En una línea parecida, Van Panhuis et al., (2014) realizan una revisión sistemática sobre la evidencia publicada acerca de las barreras para el data sharing en la salud pública. En los resultados de esta revisión, se diferenciaron 20 barreras potenciales que, a su vez, se agruparon en barreras técnicas, motivacionales, económicas, políticas, legales y éticas. Tomando estos dos documentos como referencia, a continuación, se describen estas barreras y se relacionan con los resultados de este estudio.

En lo referente a las cuestiones legales, éstas se relacionan con la privacidad de los datos de los pacientes y con temas relativos a la propiedad intelectual. Según los resultados obtenidos en el workshop, existe una gran incertidumbre legal con respecto a las regulaciones de la privacidad y el impacto de ésta en el data sharing. Tal y como y como se detecta en el workshop que origina este estudio, existen dos posiciones: la que aboga por la privacidad total y rechaza cualquier tipo de datos compartidos y la que ponen una balanza entre la utilidad social de compartir datos y las cuestiones de privacidad. En referente a los que abogan por la segunda opción, se refiere que, para que los datos puedan ser compartidos, se debe tener mucho cuidado a la hora de organizar, sistematizar y gestionar la información que se va a compartir, teniendo muy en cuenta aspectos como, por ejemplo, los formatos de la en los que la información se encuentra registrada para que los procesos puedan ser interoperables (sobre todo, pensando en los registros electrónicos cada vez más frecuentes). Por otra parte, Van Panhuis et al., (2014), además de estas cuestiones relacionadas con la protección de la identidad de los pacientes, destacaron otras preocupaciones como el uso que se puede hacer a veces de los derechos de autor para restringir en lugar de para ampliar el acceso a los datos. Concretamente, centran su atención en el contexto de la salud pública, donde, según su perspectiva, se puede dar la situación de que las agencias que recopilan datos de salud pública se sientan responsables de la protección de la privacidad individual y comunitaria y pueden sentir que el público les confiere una función de tutela o propiedad. Además, refieren que, en la práctica, a menudo no está bien documentado o no se sabe bien quién es el propietario de los datos de salud

pública. Esta percepción de sentirse en la responsabilidad de garantizar que los datos de pacientes que han sido obtenidos en el marco de proyectos financiados se utilicen adecuadamente también se ha detectado en una de las personas entrevistadas, que consideraba que el proceso antes de que se compartan los datos debía de ser muy cuidadoso y respetando todos los acuerdos.

En general, este aspecto de la privacidad y de la confidencialidad fue especialmente recalcado por las personas entrevistadas que llevaban a cabo investigaciones más aplicadas (tanto clínicas como psicosociales), donde la nota común era trabajar directamente con datos personales de pacientes (en el caso de los que eran personal sociosanitario) o de sujetos. Estos entrevistados se mostraban preocupados por los interrogantes y problemas que podrían surgir a la hora proteger la privacidad de los sujetos con los que investigaban. Cuando se les pedía que profundizaran acerca de este tema, mencionaban aspectos como los acuerdos de confidencialidad (consentimientos informados) que los pacientes/sujetos firman antes de comenzar una investigación que, de momento, no contemplan el supuesto de que los datos obtenidos puedan ser compartidos. Sin embargo, todas las personas entrevistadas (excepto una) que abordaron este tema estaban de acuerdo en que, probablemente, los pacientes/sujetos de estudio no tendrían ningún problema en que sus datos fueran compartidos en caso de que se les pidiese y que, además, seguramente mostrarían una actitud de buena disposición ya que con esto se podría contribuir a mejorar la ciencia. En este sentido, se concuerda con las impresiones obtenidas en el workshop, donde se resaltó que los pacientes suelen estar abiertos a compartir siempre que se les explique y se les agradezca debidamente su disposición. La visión de la persona entrevistada que no estaba de acuerdo con que los participantes de los estudios tuvieran buena disposición corresponde a una investigadora fundamentalmente cualitativa que maneja datos sensibles de las personas a las que investiga. Estos datos sensibles, que se encuadran dentro del área de Biología Reproductiva y, dentro de ésta, de la infertilidad, solían ser muy difíciles de obtener por la reticencia de las personas a ofrecerlos, debido a la vergüenza que pueden producir y al miedo a ser identificados. Por ello, esta entrevistada, aunque reconocía la utilidad de los datos compartidos a la

hora de garantizar la transparencia y la credibilidad de la información, se mostraba contraria de que se exigieran en todas las áreas y para todo tipo de investigaciones.

Otro aspecto que destacó entre las personas entrevistadas y que coinciden con otros estudios (Inkpen et al., 2020; Van Panhuis et al., 2014) es el miedo a que te juzguen o que no entiendan tu trabajo. Esta inquietud fue constatada por muchas de las personas entrevistadas, por lo que es una preocupación relevante. En este sentido, se deduce que sienten miedo a que los datos se vean como algo divorciado del resto del trabajo, que se pierda el contexto de su creación y que se puedan hacer malas interpretaciones de los mismos. Es por ello que muchas de las personas entrevistadas que argumentaban en esta línea (sobre todo los básicos) veían lógico poder hacer un seguimiento de sus datos y no perder totalmente el contacto con los mismos, en parte para ver qué se hace con los mismos por curiosidad científica, en parte para no perder una posibilidad de colaborar y, finalmente, para poder ejercer un control y un escrutinio sobre lo que se hace con los mismos. Además, también se descubrió el propio miedo a que, si otros tienen acceso a sus datos, puedan detectar fallos y, de alguna manera, poner en entredicho la investigación realizada con esos datos.

A un nivel más técnico y relacionado con el propio manejo de la práctica de compartir datos, se encuentra entre el personal investigador una barrera que está asociada con la dificultad que puede conllevar el preparar los datos para que estén en condiciones de ser compartidos. Ya se adelantaba en párrafos anteriores que, normalmente, eran el personal investigador más joven quienes solían encargarse de preparar los datos para compartirlos y conocer los medios para llevarlo a cabo (repositorios de datos, material suplementario, etc.). Las opiniones de estos investigadores sobre el trabajo extra que supone, así como del personal investigador sénior (la hubieran llevado a cabo o no) tendía hacia considerar la práctica bastante compleja. Ahora bien, dentro de considerarla compleja, no todos se lo tomaban con el mismo talante. Aquí se observó un patrón bastante común y es que, cuanto más integrada esté la práctica, cuántas más veces se haya hecho, más naturalizada se tenía. Sin embargo, todas las personas entrevistadas, sin excepción, referían que se agradecería que se simplificase lo más posible, que fuera más intuitiva y sencilla posible. En este sentido, la falta de protocolos estandarizados en algunas disciplinas hacía que el proceso resultara tedioso

y muchas veces realizado a última hora, cuando se encontraban con que una revista lo pedía. Esto producía una cierta incomodidad al tener que desenterrar los datos después de haber acabado la investigación y tener que prepararlos para compartirlos. Según los datos obtenidos en el workshop organizado por la IOM, se sugiere que los cambios para el futuro tienen que venir en esa dirección: entrenamiento para integrar la preparación de los datos para su compartición ya durante el proceso de investigación para que la información se tenga todavía fresca y sea un proceso fluido y que, si llega el caso de tener que compartirlos, ya estén preparados.

Otra barrera que fue también señalada por Van Panhuis et al., (2014) y que coincide con las personas entrevistadas, es la económica. En este sentido, se señala que la práctica de compartir datos cuesta dinero y tiempo, que han de invertir el personal investigador utilizando la financiación de la que dispongan. Entre las personas entrevistadas hubo quienes trataron este tema, argumentando que si es una tarea a mayores que se ha de hacer, debe llevar aparejada los recursos necesarios.

Finalmente, y también siguiendo el hilo de las impresiones que se obtuvieron en el workshop y Van Panhuis et al., (2014) en su revisión, y que coinciden en muchos aspectos con las detectadas en las entrevistas, surgen como barreras una posible falta de cultura de compartición de datos y los temas relacionados con la motivación. En primer lugar, con respecto al aspecto más cultural, es donde se observaron diferencias más evidentes, de nuevo, en el personal investigador aplicado dada su menor experiencia, pero también tuvieron lugar en las impresiones de algunos básicos. Cabe mencionar que a cuanto más básico, más afianzada estaba la cultura de compartir datos y a cuanto más próxima a la investigación con pacientes, menos cultura. A este respecto, en las entrevistas del personal investigador más básico (nivel celular) se observó una predisposición basada en parte en un convencimiento de que compartir datos era lo correcto y en parte en el seguimiento de unos mandatos claros de las revistas donde solían publicar. Sin embargo, en investigadores con estudios más preclínicos (fundamentalmente, investigación con animales), a la vez que la frecuencia de llevar a cabo la práctica de compartir datos era menor, también crecía la sensación de que era un incordio que, de vez en cuando, les pedía alguna revista y que hacían porque había que hacerlo para poder publicar. Cuando la investigación ya llegaba a

humanos, como se ha dicho anteriormente, la práctica de poner los datos en abierto era casi nula y la cultura de llevarla a cabo, por extensión, prácticamente inexistente. Sobre todo, en estos dos últimos casos, se observaron muchas reticencias y pocos incentivos, así como excusas entremezcladas con la percepción de que era algo positivo para los pacientes y para la propia ciencia, pero incómodo. La percepción final que se obtiene sobre la cultura es que el personal investigador necesita algo más que regulaciones y mandatos para aceptar de buen grado esta práctica. Se necesitan, básicamente, incentivos. Y estos incentivos no son otros que la conversión de todas las barreras, dificultades y problemas que se han ido comentando en este apartado, en facilidades. Es decir, a nivel más específico, diseñar regulaciones legales precisas y que se ajusten a las necesidades y demandas; ofrecer recursos económicos, materiales y humanos dedicados concretamente a la práctica de compartir datos; garantizar incentivos en forma de citaciones que contabilicen de alguna manera en beneficio de los investigadores; ofrecer formación específica para comprender la práctica de compartir datos y llevarla a cabo correctamente; y promover que las revistas, editoriales, repositorios, financiadoras y otros agentes demandantes de los datos ofrezcan facilidades para que el proceso sea lo más sencillo, eficaz y eficiente posible. A nivel más general, un incentivo es que el personal investigador sienta que se tiene en cuenta que la ciencia es muy amplia y que las disciplinas que la conforman muy variadas, por lo que no se puede pedir un “tabla rasa” en cuanto a compartir datos, sino que se debe adaptar la práctica a las necesidades y particularidades de cada disciplina. Por último, a un nivel más idealista o utópico (teniendo en cuenta las lógicas capitalistas en las que la ciencia se ve inmersa), promover que el uso compartido de datos se encuadre dentro de un clima de colaboración y de intercambio de conocimientos entre investigadores y se aleje de los preceptos que predominan actualmente de competitividad e individualismo.

3. Actitudes

3.1 Disposición a compartir datos

La intención de crear un espacio dedicado a calibrar las actitudes hacia la práctica del data sharing fue concebido para detectar, una vez abordados los conocimientos y las creencias, hasta dónde estarían dispuestos a llegar las personas entrevistadas con respecto a compartir sus datos de investigación. Esto parte de la premisa de que, una cosa es saber o no saber hacer algo (conocimientos), otra es lo que una persona crea que es mejor o peor, más o menos beneficioso (creencias) y otra (a veces bastante distinta) lo que está dispuesta a hacer. Las actitudes son, tal y como se entienden en este trabajo, la cristalización de las creencias, los conocimientos y las experiencias. O, dicho de otro modo, lo que sus creencias y conocimientos les hacen llevar a hacer en la práctica.

En general, en relación con la disposición y la actitud hacia compartir datos se detectan tres tendencias: el factor de la publicación, la vinculación “emocional” con los datos, el deseo de control hacia los mismos y los factores de beneficio/colaboración.

En primer lugar, el factor de la publicación se puede entender muy bien si se piensa en la frase “*publish or perish*” tan famosa en el mundo de ciencia y de la comunicación científica, que refleja de alguna manera la presión por publicar que existe actualmente (Fanelli, 2010). Esta presión hace referencia a la inercia marcada por diferentes agentes dentro de la academia según la cual se debe estar continuamente publicando como manera de demostrar el talento y el estatus. En este sentido, los investigadores que no publican frecuentemente, o que se focalicen en otras actividades que no tienen como resultado las publicaciones, son concebidos como menos prestigiosos y pueden perder de alguna manera categoría en el imaginario colectivo (Siegel, Brand, Rossi, & Lubowitz, 2018).

Esta competencia y presión por publicar que impera en la academia, basada en producir cuanto más mejor, tiene entre otras consecuencias aspectos como el sesgo científico. Esto es, una sobrerrepresentación de resultados “positivos” y una escasez

de publicación de resultados “negativos” por temor a que éstos no sean publicables o que no obtengan citas, aunque esto sea en detrimento de la ciencia, ya que resultados relevantes podrían quedar en el olvido (Echevarría, Malerba, & Arechavala-Gomez, 2020; Faschinger, 2019). De hecho, existe la paradoja de que muchos investigadores apoyarían que se publiquen los resultados negativos, sin embargo no lo hacen debido al temor de que reciban poco impacto (Echevarría et al., 2020).

Según el estudio de Fanelli (2010), este fenómeno de sesgo hacia la publicación solo de datos “positivos” se ve incrementada en las instituciones de países donde la competencia académica y las presiones para publicar son altos. Según él, a más excelencia académica, también existe más sesgo. También de acuerdo con esta posición están Siegel, Brand, Rossi, y Lubowitz (2018) para el ámbito biomédico dentro del contexto de EEUU. Tomando como referencia a los jóvenes residentes de medicina, señalan hasta qué punto se ha ido implementando en criterio de publicar cuanto más mejor, incluso para poder acceder a estos primeros puestos para realizar su residencia. Señalan que, actualmente, en biomedicina impera la cantidad sobre la calidad y que los incentivos para conseguir cualquier logro en ciencia (financiación para proyectos, ascensos, acreditaciones, acceso a diversos puestos de trabajo) se basan en tener una cantidad considerable de publicaciones, dejando un poco de lado el que éstas puedan contener calidad. Por otro lado, en el contexto europeo se sigue una línea parecidas, tal y como se puede observar en el trabajo de Goyanes y Rodríguez-Gómez (2018) sobre el “*publish or perish*” en el contexto español. Desde su punto de vista, este sistema basado en publicar o morir condiciona muchos de los esfuerzos en la ciencia e implica, a su vez, una adaptación del currículum a las peticiones de las agencias de acreditación. Mencionan, en este sentido, efectos coloquialmente conocidos como “publicacionismo”, “incrementalismo” o “efecto Aneca”⁵, que ilustran bastante bien esta tendencia para publicar mucho en vez con más calidad (cantidad vs calidad). El estudio, que se basa en un cuestionario enviado a personal docente e investigador del

⁵ La Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) es un Organismo Autónomo, adscrito al Ministerio de Ciencia, Innovación y Universidades, que ha sido creado por el artículo 8 de la Ley 15/2014, de 16 de septiembre, de racionalización del Sector Público y otras medidas de reforma administrativa, procedente de la conversión de la Fundación Agencia Nacional de Evaluación de la Calidad y Acreditación en organismo público, que tiene como objetivo contribuir a la mejora de la calidad del sistema de educación superior mediante la evaluación, certificación y acreditación de enseñanzas, profesorado e instituciones (ANECA, 2021).

campo de la Comunicación en universidades españolas acerca de temas relacionados con la presión por publicar, revela datos tan interesantes como que existe una diferencia significativa en cuanto a los años de experiencia, esto es, a cuantos más años de experiencia menos presión por publicar. Asimismo, también muestra que la causa más importante para la presión por publicar proviene de la ANECA, es decir, de factores externos. Dentro de esta presión externa por publicar, el personal con situaciones menos estables (profesorado indefinidos y contratados temporales) tienen una percepción de presión estadísticamente mayor por parte de las agencias de acreditación que el resto de los grupos, mientras que los doctorandos y doctorandas, así como el profesorado asociado, perciben que la presión proviene más por parte de los compañeros.

Teniendo en cuenta todo este contexto, no es de extrañar que una de las condiciones principales para compartir los datos pase por tener en cuenta la publicación de los resultados en prácticamente todas las personas entrevistadas. Tal y como se ha comprobado en las entrevistas, claramente los investigadores se encuentran sujetos a las publicaciones, ya que, de primeras, es la principal barrera que se observa para liberar los datos.

Una vez se ponen en la situación de que los datos ya han sido utilizados por ellos mismos (es decir, que ya han publicado sus resultados), pueden empezar a plantearse compartirlos. Ahí es donde comienzan a visualizarse diferentes puntos de vista dependiendo de sus características. Estas características tienen que ver con si el personal investigador ya ha compartido datos alguna vez y también con la frecuencia con la que lo hayan hecho, ya que cuando lo hacían regularmente lo asumían como una tarea normal necesaria para la publicación, pero cuando era una práctica esporádica solía verse más como un estorbo.

Sin embargo, cabe destacar que, aunque el fin último de compartir datos suela ser cumplir con la exigencia de la revista para poder publicar, a cuanta más práctica compartiendo datos, más sentimiento de que era algo útil para la ciencia y para su área. No obstante, si la práctica solo se había llevado a cabo de manera esporádica mayor era el malestar y la sensación de que era una traba más, por lo que se deduce que hacerlo regularmente permite no solo normalizarla, sino también valorar su

utilidad. Por otro lado, dentro de las personas que nunca habían compartido sus datos más allá del *peer to peer* entre personas conocidas, o en bases de datos de acceso restringido, se detecta que el mayor desconocimiento lleva a mayores reticencias, en parte porque a veces se les hacía incluso difícil imaginar qué era de lo que se estaba hablando.

Concretizando un poco más, cabe destacar que, por normal general, el personal investigador básico no solía poner más requisito que el haber publicado ellos primero los resultados sobre esos datos. Una vez hubieran publicado, y si los datos además ya estaban subidos a repositorios como GEO, no tenían ningún inconveniente en dar directamente el número de acceso para que otras personas pudiera acceder a los datos. Cuando se trataba de investigadores básicos pero el uso de repositorios no era la norma general, también se solía encontrar acuerdo en compartir una vez se hubiera publicado, pero en algún caso con algo más de reticencia debido al trabajo que puede costar preparar los datos para compartirlos. De esto se deduce que el depósito en repositorios puede tener la ventaja de ayudar a introducir la práctica de compartir datos en el proceso de investigación, durante y no después, para favorecer el que luego sea más sencillo compartirlo si llega el caso de que cualquier agente (la revista, otros investigadores e investigadoras, la normativa de los proyectos...) lo solicita o lo exige.

Con respecto a compartir antes de publicar, la cuestión se complica un poco más. La disposición ya no es tan grande y otros factores entrarían en juego, como el solicitar colaboraciones en algunos casos, o que se citen los datos en otros. El problema de la citación de los datos es que realmente, a día de hoy, aunque existe normativa para la misma y hay revistas y editoriales que al promueven (Federer et al., 2018) no ofrecen ningún rédito científico más allá del mero reconocimiento (Buneman et al., 2020). Esto es relevante ya que, si no fuera así y sí que ofreciera algún tipo de contabilización efectiva, podría funcionar como incentivo a la hora de querer compartir antes de tener resultados publicados. Todo esto tiene sentido si se vuelve a pensar en que la ciencia y los científicos están inmersos, como ya se explicaba anteriormente, en una lógica capitalista y competitiva donde el “hacer por hacer” o “dar por amor al arte” es bastante minoritario.

Por otro lado, con respecto a la vinculación emocional con los datos, se ha elegido este concepto para explicar un sentimiento de pertenencia que iba más allá de la propiedad en sí. Este sentimiento se ha detectado en las entrevistas sobre todo a los investigadores aplicados y se hacía patente por la manera en la que hablaban de sus datos, como si se tratase de algo muy cercano a ellos y ellas, su creación. Cuando se profundizó sobre este aspecto, se intuyó que era un tema con varias aristas, que tenían que ver sobre todo con ejercer un control sobre los datos, una supervisión sobre a quién se van a compartir y para qué, y cómo podrían ellos y ellas tomar parte en esos posibles nuevos trabajos con sus datos. Mientras que el personal investigador básico planteaba alguna de estas opciones (sobre todo la de establecer colaboraciones en nuevos trabajos) cuando se planteaban una posible solicitud de los datos antes de haber sacado ellos y ellas publicaciones de resultados, los aplicados lo consideraban para cualquier momento del proceso (tanto antes como después de publicar). La explicación que se ha encontrado para este fenómeno es la cercanía de los datos con información directamente sobre sujetos. Es decir, cuanto más cerca estén los datos de ser información sobre personas, más celo sobre los mismos incluso después de haber publicado. Del mismo modo, cuando más “lejos” (investigación con animales, o a nivel celular o molecular), más desvinculación “emocional” con los mismos.

La constatación sobre esto se puede encontrar en la literatura científica que se ha escrito alrededor del data sharing y su relación con los datos personales de pacientes o usuarios de diferentes servicios sociosanitarios. Como ya se hacía referencia en el marco teórico (pg. 57), varios trabajos apuntan en esta dirección. En el caso concreto de los estudios con características similares a los de esta tesis (estudios cualitativos), se puede observar claramente esta tendencia a focalizar el objetivo en torno a los datos de los pacientes. En este sentido, el centro del interés se focaliza en analizar la opinión de los pacientes/usuarios de servicios de salud, así como del personal sanitario, investigadores y otros agentes clave que manejan este tipo de datos. El objetivo de estos trabajos suele ser abordar los límites del consentimiento informado, la delicadeza en el tratamiento de la información confidencial y los mecanismos que existen o que deberían existir para manejarlos y gestionarlos de la manera más apropiada posible. Además, reflejan varias realidades teniendo en cuenta las

diferencias entre países (con más y menos recursos), o de la titularidad de la institución (pública o privada), así como la funcionalidad de la misma (centros de investigación, hospitales o empresas de productos sanitarios como las farmacéuticas) (Broes et al., 2017; Cheah et al., 2018; Colledge et al., 2014; Saunders et al., 2014).

Otro aspecto que se destacó como importante para las personas entrevistadas es el del beneficio de las colaboraciones que podrían extraer gracias a la práctica de compartir sus datos. Tal y como se obtuvo la información sobre este aspecto en las entrevistas, ésta podría formularse tanto en negativo como en positivo. En negativo porque, con respecto a compartir los datos antes de haber publicado, solo era viable (para la gran mayoría de entrevistados y entrevistadas) si se establecía un proceso de negociación, según la cual ellos esperaban obtener algún beneficio de la compartición. Es decir, se trata de un “no lo haría, excepto...”, y es por ello que se dice que, en principio, parece una formulación en negativo. Sin embargo, también es importante poner en valor el que no se cierran en banda a hacerlo, lo que en una formulación en positivo se trataría de un “lo haría si...”. Esto remite en cierto modo a aquello que se ha comentado en varias ocasiones a lo largo de la discusión de que, en muchas ocasiones, según lo que se ha observado en el transcurso de las entrevistas, la clave está en los matices. Los investigadores no van, por norma general, a aceptar compartir sus datos sin establecer condiciones, ni tampoco se van a negar a hacerlo en redondo sin abrir una puerta a la negociación. Como se ha observado en los resultados, estarían dispuestos en muchas ocasiones a abrir negociaciones sobre compartir sus datos con otros colegas, incluso aunque no los conozcan de nada. Pero lo harían con condiciones en forma de colaboraciones, que pueden ir desde compartir autorías hasta la misma participación de nuevos proyectos donde se estén utilizando sus datos permitieran, así como también citas apropiadas que contabilicen de forma similar a como sucede con los artículos.

Particularmente, se ha observado en la literatura que el personal investigador de Ciencias de la Salud ni es totalmente ajeno a este tipo de prácticas ni tampoco está en contra de la misma (R Alexandre-Benavent et al., 2020; Tenopir et al., 2020). En los trabajos que se han realizado en el marco del proyecto DATASEA y que han servido para dar contexto a esta tesis se puede observar perfectamente esta realidad. Por

ejemplo, en el caso del estudio realizado en el CWTS sobre la presencia de datasets en ORCID, donde una de las variables que se estudiaron fue la pertenencia a las distintas áreas de conocimiento, se observó que las *Medical and Life Sciences* junto con las *Natural Sciences* constituían el 85,6% de un total de siete áreas (Sixto-Costoya, 2021) .

En ocasiones, además, constituyen incluso un grupo en la vanguardia, cuyas prácticas, infraestructuras y modalidades de compartir podrían servir de ejemplo e inspiración para otras áreas, como las Ciencias Sociales o Humanidades. Que existan mecanismos para facilitar estas negociaciones podría ser tan buena idea como en su día fue el establecimiento de la práctica de compartir preprints como manera de sortear las imposiciones de revistas y editoriales debido al copyright para compartir postprints.

Además, tal y como se detectó en las entrevistas, el establecimiento de colaboraciones no solo estaba enfocado en sacar beneficio de compartir datos, sino también debido a la curiosidad científica de participar en otros estudios utilizando los mismos datos pero con otros objetivos, y saber qué cosas nuevas se podrían obtener. Tendencias que, en definitiva, de lo que están hablando es de más cooperación y colaboración, complicadas en el panorama actual, pero no imposibles. A principios de la década de los 2000 el movimiento del OA comenzó a gestarse y a día de hoy se puede decir que ya está implantado y que pocos investigadores afirmarían no conocerlo (Abadal, 2012; Laakso et al., 2011; San José Montano, 2009). La reivindicación del data sharing ligado a internet y al movimiento del OA ya se empieza a mencionar en la Declaración de Berlín del año 2003 y desde ese momento ha evolucionado hasta estar presente tanto en la política de las editoriales y de las revistas, como en las políticas científicas de muchos países (Digital Curation Center & SPARC Europe, 2019; Ministerio de Ciencia e Innovación, 2020). Sin embargo, su implantación no es tan amplia como lo ha sido el OA a las publicaciones científicas. Una de las causas puede ser este matiz de publicar los datos antes o después de la publicación y otra, sin duda, es la falta de incentivos.

Aquí está la clave, que ya apuntaba el entrevistado E008 en su frase “podrán decidir si quieren más látigo o más zanahoria”, que viene a significar que el personal investigador va a compartir si no hay más remedio. Es decir, si se exige para poder publicar, promocionar u obtener financiación. Pero, por la salud de la ciencia en general y del personal investigador en particular siempre va a ser mejor “la zanahoria”.

Lo que significa que, para que el uso compartido de datos funcione y cale no solo en los hábitos, sino también en la creencia de que es lo moralmente ético y, además, beneficioso para el avance de la ciencia en su conjunto, se necesitan algo más que imposiciones. Hasta el momento actual, esto no ha sido así.

3.2 Cómo deberían compartirse los datos

Este apartado fue diseñado para obtener una visión sobre la parte más técnica de la práctica de compartir datos. Se abordó de manera específica al final de la entrevista y sirvió (aunque no de manera intencionada) para cerrar el encuentro de una forma amena y ligera después de haber estado hablando sobre aspectos más profundos y sesudos como son las creencias y actitudes, a veces delicadas por ser muy personales. Las preguntas técnicas, sin embargo, sirvieron para relajar el ambiente y para que las personas entrevistadas se *explayaran* sobre aspectos que controlaban. Esto sirvió tanto si habían tenido contacto como si no con la práctica de compartir datos, ya que todas podían hacerse fácilmente a la idea de qué es lo que les gustaría técnicamente si se vieran en la tesitura de utilizar datos de otras personas.

Las tres peticiones más comunes giraban en torno a que los datos estuvieran auténticamente brutos, que se contextualicen con el resto del estudio y que la información de los datos (la presentación del contenido) estuviera lo más estandarizada posible.

La primera, que los datos estén realmente brutos, tiene sentido si se tiene en cuenta que el fin último de la práctica de compartir los datos es asegurar la transparencia, la replicabilidad de los resultados y la reutilización de los datos (Peset et al., 2017). Difícilmente se puede asegurar ninguna de los tres aspectos si los datos ya están procesados. Sin embargo, en la entrevista esto solía venir a colación de que más de uno se había encontrado con datos compartidos supuestamente como brutos, que finalmente no lo eran, ya que eran tablas o figuras procesadas. Asegurar hoy en día esta crudeza de los datos es complicada por varios motivos. En el caso de los datos brutos compartidos como material suplementario, porque muchas veces el *peer*

review o bien no alcanza a la revisión del material suplementario, o lo hace solo superficialmente (Pop & Salzberg, 2015) y es el posterior lector quien ha de discernir si los datos son válidos o no. En el caso de la estandarización de la información en el material suplementario, sucede algo similar, y es que las revistas no suelen incluir en sus “*author guidelines*” explicaciones sobre cómo deben presentarse formalmente los datos de investigación en caso de ser adjuntados como material suplementario. En el caso de los repositorios, aunque a priori se espera de ellos que tengan algún filtro de calidad esto no siempre sucede. Aunque hay algunos como GEO que realmente sí tienen estos criterios, ya que al menos mantienen un control realizado por personas para asegurar que se cumplan unos requisitos básicos mínimos, otros como ZENODO aclaran que todo lo que se encuentra depositado en su plataforma está “*as is*”, es decir, que no pasa por ningún tipo de control (Edgar, Domrachev, & Lash, 2002; ZENODO, 2021). En este sentido, cabe destacar que el personal investigador de básica que utilizaba GEO (la gran mayoría utilizaban este repositorio) daban por sentado que si seguían las instrucciones ya lo estaban haciendo correctamente, haciendo hincapié además en lo minucioso (y a veces complicado) que era el proceso de subir los datos a esta plataforma. Estos procesos de estandarización de la información, aunque ellos no utilizaran este concepto, están relacionados con los metadatos, pieza fundamental en la práctica de compartir datos. Los metadatos es información que acompaña a los datos y que les otorga un contexto entendible y por ello han de ser lo suficientemente completos para que una persona que no ha sido la creadora original de los datos tenga toda la información necesaria para comprenderlos (Mayernik, 2019; Torres-Salinas et al., 2012).

En relación también con el contexto entendible, pero desde una perspectiva menos técnica y más filosófica y ética, Inkpen et al., (2020) también reflexionan sobre la importancia del contexto en lo que se refiere a los datos de investigación. Desde su punto de vista, los datos no son nada, no valen nada, sin su contexto. Pero no solo porque técnicamente no se entenderían y costaría muchísimo sacarles una utilidad en términos de replicabilidad o transparencia, sino que además tendría una serie de implicaciones para sus creadores. Según su opinión, es un error considerar los datos como una entidad separada y transferible, ya que en la gran mayoría de los casos esta

no es la realidad en la investigación. En este sentido, argumentan que los datos raramente pueden ser entendidos e interpretados como un fenómeno aislado, sino que tienen que ser encuadrados dentro del esfuerzo intelectual y las habilidades técnicas que los han generado. Según estos autores, entender el esfuerzo intelectual que ha generado los datos es tan valioso, o más incluso, que cualquier información adicional (técnicas, procedimientos...) que se pueda dar sobre los mismos. Es más, discuten también sobre lo que denominan la *“deification of data”*, que viene a ser el convencimiento que puede existir en la comunidad científica de que los datos brutos son algo completamente neutro que no se ve afectado por ningún factor subjetivo. Es decir, argumentan que podría ser un error considerar los datos como una especie de embarcación vacía donde cada cual puede verter sus propias interpretaciones, sin considerar (de nuevo) el contexto en el que pudieron ser creados. De manera parecida, Kitchin (2014) cuestiona incluso que unos mismos datos puedan ser transferibles entre contextos, entendiendo en este caso por contextos a todos aquellos elementos que engloban a las preguntas de investigación, las particularidades y vivencias del grupo que las crea, sus necesidades o sus presiones. Según este razonamiento, para utilizar de algún modo los datos brutos de otras personas no solo habría que tener un buen conjunto de metadatos que los definiese a la perfección, sino que habría que conocer también el contexto en el que fueron creados y valorar si estos datos tendrían o no cabida en un contexto nuevo. Todo este razonamiento se ha visto reflejado cuando las personas entrevistadas mostraban su preocupación por que otras personas pudieran no entender o malinterpretar sus datos.

Conclusions

The first conclusion of this study is that these types of studies are necessary. Before this thesis was carried out, several studies that explored the data sharing practices among researchers were done. But they were mostly completed from general perspectives. These studies are very useful but uncompleted if we want to know the real reasons of the data sharing behaviors.

Following the previous experience with the DATASEA project in the field of data sharing, there is a need of a qualitative study. Until now, the scientific evidence found after exploring the data sharing phenomenon has been mainly focused on three issues: 1) the enthusiasm for the advantages of this practice and, as a result, the big progress for Science; 2) the inconveniences and barriers; and 3) the technical challenges. Regarding to these topics, on the one hand several studies have been focused on issues such as the technical difficulties of the development of appropriate digital infrastructures for data sharing, the complications related to learn new ways to organize raw data and metadata to make it comprehensible for sharing, and the developing of new deliverables such as the Data Management Plans. On the other hand, there are also a significant number of studies addressed to understand the role of researchers and other professionals as users and creators of data. These studies, mostly quantitative and based on questionnaires completed by researchers, provide us with valuable information and an insight into the researcher's perspectives on this practice. Nevertheless, all these studies try to understand this phenomenon of data sharing offering general visions for a complex reality, that is, global responses for global questions that help us to have a top-down understanding.

In this thesis we have tried to change things. Starting from the idea that data sharing is, first, a practice that should start from researchers, we have tried to get as close as possible to them. We assume that as good as this practice is, there is not much that can be done if people creating the data do not understand/know/agree with the data sharing practice. For this reason, it is necessary to explore the subjective reality of researchers and try to formulate questions closer to “why” rather than to “what”, that is, closed responses to open questions that help us in this top-down understanding.

This thesis is focused on the Spanish context of Health Sciences research. Subsequently, the thesis is an attempt to answer the general objective set out: “To investigate the sharing of raw data among researchers in the field of Health Sciences from a qualitative perspective”.

As for the specific objectives, we have achieved to have a close idea of what researchers know about the data sharing practices. Besides to know that basic researchers have in general more experience than clinicians in almost all the investigated aspects (journals, infrastructures, formats...), the more experience they show the higher agreement with the practice they have. However, it was very obvious the great influence that the requirements of the journals have over the researchers and their data sharing practices.

Regarding to the beliefs of the researchers related to the data sharing practices, there were two clear trends. They mostly agreed with the advantages of open data and they were able to recognize several of them (the reproducibility of the data, the transparency, the reuse, material and human savings). However, they delved much deeper into the fears, barriers and inconveniences. This inequality between the advantages and inconveniences shows the long road ahead in order to reverse this situation.

When their attitudes were analyzed, these inequalities between advantages and barriers were particularly evident. The approach to the different attitudes basically served to clarify how far they would be willing to go in a data sharing situation. All of them, without exception, set some kinds of conditions. These conditions could be a basic acknowledgement in the final publication, or continuing to have full control over the shared data. Anyhow, none of the people interviewed considered that the sharing of data should be a totally gratuitous gesture and subject to good will.

In addition, two different typologies of research in Health Sciences were considered: basic and applied. The relevance of this differentiation was observed in the 5 Journal Citation Reports categories that were selected to build the sample. In this case, the differences among the interviewees were very notable. Clearly, the belonging of one or the other research typology was indicative of the data-sharing practices undertaken.

Basic researchers were more aware about data sharing than the applied ones, since many of the latter barely even knew what this practice consisted of. This shows how important is to consider the differences between disciplines even when they are in the same area, in this case, Health Sciences.

Finally, the study considered the academic hierarchy and years of experience as one of the objectives of this thesis. Although the individuals from the lower ranks and less years or experience were somewhat more open-minded about open data and data sharing practices and were generally more familiar with the technical aspects of the practice, they were equally concerned about the drawbacks as the more senior interviewees so far. On the one hand it should be kept in mind that their scientific careers are just beginning and the pressure to publish and gain a CV is very high. On the other hand, senior researchers were concerned about the high level of competition for projects and funding, considering that putting their data in the open could be a competitive inconvenience.

As a final general observation and taking into account all the issues analyzed in this thesis, it is important to highlight a reality that the interviews revealed about the practice of data sharing. In political science, Giuseppe Tomasi di Lampedusa (1896-1957) exposed an interesting paradox known as "gatopardismo" or "lampedusiano". This exposition says the following "change everything so that nothing changes" or what is the same "If we want everything to remain as it is, everything must change". In the explosion of the data sharing movement, accompanied by Open Data or Open Science, many things seem to change for the better: a more open and democratic science, where the scientific community and even the whole society can enjoy and take advantage of scientific findings. But really, if we do not change the basis of the problems (high and sometimes unhealthy race to publish, high pressure of competition in scientific careers, great power of publishing lobbies, or high prices even to publish in open access, among others) the open data movement and the positive momentum through data sharing practices has the risk to remain a nice dream with an unsatisfactory final where everything seemed to change, to finally stay the same.

Limitaciones y propuestas de futuro

La principal limitación de este estudio es que, por razones de cercanía geográfica y, por supuesto, por la sobrevenida crisis por la COVID-19, no se han podido realizar entrevistas que abarcasen un rango más amplio del territorio español. Desde un principio, la idea fue que la entrevistas se realizasen de manera presencial, así que el espacio geográfico hubo de limitarse a las posibilidades de desplazamiento. Aun así, se tuvieron que realizar dos entrevistas por Skype que ya estaban programadas, pero que no pudieron realizarse presencialmente. Por otra parte, también hubiera sido deseable contar con más representación de otras áreas de Ciencias de la Salud que pudiesen enriquecer más la muestra y ofrecer otras perspectivas, además de las analizadas en este estudio.

Sin embargo, estas limitaciones también pueden representar oportunidades de futuro. Tras los resultados obtenidos en este trabajo, se considera muy oportuno realizar otros estudios de tipo cualitativo que exploren la perspectiva del personal investigador sobre el uso compartido de datos en contextos diferentes. Por ejemplo, se podrían añadir investigadores de Ciencias de la Salud de otros países de Europa, para comparar el contexto español con países cuya organización política y social puede ser semejante. Además, también podría ser muy útil e interesante focalizar estudios en el grupo de los investigadores e investigadoras con menos experiencia y de menos rango en la escala académica. Esto se debe a la sencilla razón de que estas personas son el futuro de la ciencia y representan el relevo generacional. Lo que este grupo de investigadores interiorice ahora sobre la práctica de compartir será de vital importancia en la ciencia del mañana. Por ello, vale la pena una focalización en sus perspectivas sobre la práctica de compartir datos si realmente se desea que ésta tenga proyección en las próximas décadas. Por otro lado, esta tesis se focalizó en el personal investigador de Ciencias de la Salud, pero quedan otras áreas de la ciencia que sería muy interesante abordar, como las Ciencias Sociales, las Humanidades, las Ciencias Naturales o las Ingenierías. Abrir este marco de actuación a más áreas tendría la ventaja no solo de ofrecer más información y otras visiones sobre las perspectivas de los investigadores e investigadoras sobre el fenómeno de compartir datos, sino que también permitirían

detectar las particularidades de cada área y, por consiguiente, realizar propuestas de actuación adecuadas a las necesidades.

Por último, una propuesta de futuro que deriva de este estudio tiene que ver con la incidencia política. Como se ha comentado en las conclusiones, y también a lo largo del trabajo, desde que el tema de compartir datos y el movimiento de Open Data y Open Sciences con el que se asocia se puso “de moda” (es necesario recordar que la práctica existir, existe desde hace al menos un siglo), empezando por declaraciones como la de Berlín y siguiendo por otras muchas declaraciones y recomendaciones, poco se ha incidido desde el ámbito de la política, en la situación particular del personal investigador. Uno de los resultados de esta tesis es que ninguna de las personas entrevistadas va a compartir sus datos “por amor al arte”, y esto es normal desde el punto de vista del sistema meritocrático que predomina la ciencia actualmente. Los investigadores e investigadoras pueden estar muy de acuerdo con las ventajas de esta práctica, pero lógicamente esperan algo más. Este algo más puede venir dado por más incentivos para compartir, por plenas garantías de la protección de los datos que comparten (en términos de confidencialidad y anonimato de los datos sensibles), por ayudas y soporte frente a las dificultades técnicas, o incluso por un relajamiento de los requisitos y las presiones para acceder a contratos, becas o proyectos, lo que podría conllevar más cooperación. Todo esto es un suponer, ya que, de momento, ninguna acción de política científica, al menos en el caso español, va en esta dirección. Resultados como los de esta tesis pueden orientar a legisladores y personas encargadas de la toma de decisiones en el ámbito científico sobre qué es lo que piensan los investigadores, cuáles son sus inquietudes y sus miedos y hasta dónde están dispuestos a llegar en la compartición de sus datos. Es decir, cuánto hay de “utopía” y cuánto hay de posibilidades reales en esta práctica.

Referencias bibliográficas

- Abadal, E. (2012). *Acceso abierto a la ciencia*. Barcelona: Editorial UOC. Recuperado desde <https://core.ac.uk/download/pdf/11889005.pdf>
- Abadal, E., y Rius Alcaraz, L. (2006). Revistas científicas digitales: características e indicadores. *RUSC. Universities and Knowledge Society Journal*, 3(1), 20. <https://doi.org/10.7238/rusc.v3i1.272>
- Ajzen, I., y Fishbein, M. (2005). The Influence of Attitudes on Behavior. *The Handbook of Attitudes*, (January 2005), 187–236. <https://doi.org/10.4324/9781410612823-13>
- Aleixandre-Benavent, R., Vidal-Infer, A., Alonso-Arroyo, A., González De Dios, J., Ferrer-Sapena, A., y Peset, F. (2014). Disponibilidad en abierto de los artículos y de los datos brutos de investigación en las revistas pediátricas españolas. *Anales de Pediatría*, 82(1), e90–e94. <https://doi.org/10.1016/j.anpedi.2013.11.014>
- Aleixandre-Benavent, R, Ferrer-Sapena, A., Alonso-Arroyo, A., y Peset, F. (2013). Compartir datos de investigación en cardiología. *Revista Española de Cardiología*, 66(12), 1007–1008.
- Aleixandre-Benavent, R, Lucas-Domínguez, R., Sixto-Costoya, A., y Vidal-Infer, A. (2018). The Sharing of Research Data in the Cell y Tissue Engineering Area: Is It a Common Practice? *Stem Cells and Development*, 27(11), 717–722. <https://doi.org/10.1089/scd.2018.0036>
- Aleixandre-Benavent, R, Vidal-Infer, A., Alonso-Arroyo, A., Ferrer-Sapena, A., Peset, F., y García García, A. (2015). Gestión de los datos brutos de investigación en los investigadores españoles en ciencias de la salud. *Trauma Fund MAPFRE*, 26(1), 73–81.
- Aleixandre-Benavent, R, Vidal-Infer, A., Alonso-Arroyo, A., Peset, F., y Ferrer-Sapena, A. (2020). Research Data Sharing in Spain: Exploring Determinants, Practices, and Perceptions. *Data*, 1–14. <https://doi.org/https://doi.org/10.3390/data5020029>

- Aleixandre-Benavent, R, Vidal-Infer, A., Alonso Arroyo, A., Valderrama Zurián, J. C., Bueno Cañigral, F., y Ferrer Sapena, A. (2014). Public availability of published research data in substance abuse journals. *International Journal of Drug Policy*, 25(6), 1143–1146. <https://doi.org/10.1016/j.drugpo.2014.07.007>.The
- Aleixandre-Benavent, R, Moreno-Solano, L. M., Ferrer Sapena, A., y Pérez, E. A. S. (2016). Correlation between impact factor and public availability of published research data in Information Science and Library Science journals. *Scientometrics*, 107(1), 1–13. <https://doi.org/10.1007/s11192-016-1868-7>
- Alonso Gamboa, J. O. (2017). Transformación de las revistas académicas en la cultura digital actual. *Revista Digital Universitaria*, 18(3), 1–15.
- Alsheikh-Ali, A. A., Qureshi, W., Al-Mallah, M. H., y Ioannidis, J. P. A. (2011). Public availability of published research data in High-Impact journals. *PLoS ONE*, 6(9), 2009–2012. <https://doi.org/10.1371/journal.pone.0024357>
- American Journal of Political Science. (s.f.). AJPS Verification Policy. Recuperado de <https://ajps.org/ajps-verification-policy/>
- Anane-sarpong, E., Wangmo, T., Leonie, C., Tanner, M., y Simone, B. (2018). “You cannot collect data using your own resources and put it on open access ”: Perspectives from Africa about public health data sharing. *Developing World Bioethics*, 18, 394–405. <https://doi.org/10.1111/dewb.12159>
- ANECA. (2021). ANECA. Recuperado desde <http://www.aneca.es/ANECA>
- Arning, U. (2015). GMS publishes your research findings and makes the related research data available through Dryad. *GMS Zeitschrift Fur Medizinische Ausbildung*, 32(3), 1–8. <https://doi.org/10.3205/zma000976>
- Banks, D. (2009). Starting science in the vernacular. Notes on some early issues of the Philosophical Transactions and the Journal des Sçavans, 1665-1700. *ASp*, (55), 5–22. <https://doi.org/10.4000/asp.213>
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., ... Soboleva, A. (2013). NCBI GEO: Archive for functional genomics data sets -

- Update. *Nucleic Acids Research*, 41(D1), 991–995.
<https://doi.org/10.1093/nar/gks1193>
- Beck, A. (1976). *Cognitive therapy and the emotional disorders*. Madison: International Universities Press.
- Bell, J. (1993). *Como realizar un proyecto de investigação*. Lisboa: Gradiva publicações.
- Beno, M., Figl, K., Umbrich, J., y Polleres, A. (2017). Open data hopes and fears: Determining the barriers of open data. *Proceedings of the 7th International Conference for E-Democracy and Open Government, CeDEM 2017*, 69–81.
<https://doi.org/10.1109/CeDEM.2017.22>
- Berners-Lee, T. (2009). Linked Data. Recuperado desde
<https://www.w3.org/DesignIssues/LinkedData.html>
- Bezuidenhout, L. (2019). To share or not to share: Incentivizing data sharing in life science communities. *Developing World Bioethics*, 19(1), 18–24.
<https://doi.org/10.1111/dewb.12183>
- Birney, E., Vamathevan, J., y Goodhand, P. (2017). Genomics in healthcare: GA4GH looks to 2022. *BioRxiv 203554*, <http://dx.doi.org/10.1101/203554>.
<https://doi.org/10.1101/203554>
- Bohner, G., y Dickel, N. (2011). Attitudes and attitude change. *Annual Review of Psychology*, 62, 391–417. <https://doi.org/10.1146/annurev.psych.121208.131609>
- British Medical Journal. (s.f.). Article types and preparation. Recuperado de
<https://www.bmj.com/about-bmj/resources-authors/article-types>
- Broes, S., Lacombe, D., Verlinden, M., y Huys, I. (2017). Sharing human samples and patient data: Opening Pandora’s box. *Journal of Cancer Policy*, 13(July), 65–69.
<https://doi.org/10.1016/j.jcpo.2017.07.006>
- Broes, S., Verbaanderd, C., Casteels, M., y Lacombe, D. (2020). Sharing of Clinical Trial Data and Samples : The Cancer Patient Perspective. *Frontiers in Medicine*, 7, 1–8.
<https://doi.org/10.3389/fmed.2020.00033>
- Buneman, P., Christie, G., Davies, J. A., Dimitrellou, R., Harding, S. D., Pawson, A. J., ...

- Wu, Y. (2020). Why data citation isn't working, and what to do about it. *Database*, 2020, 1–16. <https://doi.org/10.1093/databa/baaa022>
- Calvete, E., y Cardeñoso, O. (2001). Creencias, Resolución de Problemas Sociales y Correlatos Psicológicos. *Psicothema*, 13(1), 95–100.
- Canadian Institutes of Health Research. (2015). CIHR Open Access Policy. Recuperado desde <http://www.cihr-irsc.gc.ca/e/46068.html#5>
- Chan, L., Arunachalam, S., y Kirsop, B. (2009). Open access: A giant leap towards bridging health inequities. *Bulletin of the World Health Organization*, 87(8), 631–635. <https://doi.org/10.2471/BLT.09.064659>
- Cheah, P. Y., Jatupornpimol, N., Hanboonkunupakarn, B., Khirikoekkong, N., Jittamala, P., Pukrittayakamee, S., Day, D., Parker, M., y Bull, S. (2018). Challenges arising when seeking broad consent for health research data sharing: a qualitative study of perspectives in Thailand. *BMC Medical Ethics*, 1–11.
- Cho, J. (2019). Subject analysis of LIS data archived in a Figshare using co-occurrence analysis. *Online Information Review*, 43(2), 256–264. <https://doi.org/10.1108/OIR-12-2017-0369>
- Clarivate Analytics. (2021). InCites Journal Citation Reports. Recuperado desde <https://jcr.clarivate.com/>
- Colledge, F., Persson, K., Elger, B., y Shaw, D. (2014). Sample and data sharing barriers in biobanking: consent, committees, and compromises. *Annals of Diagnostic Pathology*, 18(2), 78–81. <https://doi.org/10.1016/j.anndiagpath.2013.12.002>
- Comisión Económica para América Latina y el Caribe. (2019). Gestión de datos de investigación. Recuperado de <https://biblioguias.cepal.org/gestion-de-datos-de-investigacion/citacion>
- Commission, E. (s.f.). Cómo elegir el formato correcto para los datos abiertos. Recuperado desde <https://www.europeandataportal.eu/elearning/es/module9/#/id/co-01>
- Da Costa, M. P., y Leite, F. C. L. (2016). Open access in the world and Latin America: A

review since the Budapest Open Access Initiative. *Transinformacao*, 28(1), 33–45.
<https://doi.org/10.1590/2318-08892016002800003>

Datasea. (s.f.). Preparando un plan de gestión de datos. Recuperado desde
<http://www.datasea.es/es/research-data/investigador/preparando-plan-gestion-datos>

Deutsche Forschungsgemeinschaft. (2019). Handling of Research Data. Recuperado desde
https://www.dfg.de/en/research_funding/proposal_review_decision/applicants/research_data/index.html

Deverka, P. A., Majumder, M. A., Villanueva, A. G., Anderson, M., Bakker, A. C., Bardill, J., Boerwinkle, E., Bubela, T., Evans, B. J., Garrison, N., Gibbs, R.A., Gentleman, R., Glazer, D., Goldstein, M., Greely, H., Harris, C., Knoppers, B.M., Koenig, B.A., Kohane, I.S., La Rosa, S., Mattison, J., O'Donnell, C.J., Rai, A.K., Rehm, H.L., Rodriguez, L., Shelton, R., Simoncelli, T., Terry, S.F., Watson, M. S., Wilbanks, J., Cook-Deegan, R., y McGuire, A. L. (2017). Creating a data resource: What will it take to build a medical information commons? *Genome Medicine*, 9(1), 1–5.
<https://doi.org/10.1186/s13073-017-0476-3>

Diez Patricio, A. (2017). Más sobre la interpretación (II). Ideas y creencias. *Revista de La Asociación Española de Neuropsiquiatría*, 37(131), 127–143.
<https://doi.org/10.4321/S0211-57352017000100008>

Digital Curation Center, y SPARC Europe. An Analysis of Open Science Policies in Europe , v4 (2019).

Dryad. (2019). About Dryad. Recuperado desde
<https://datadryad.org/pages/organization>

Echevarría, L., Malerba, A., y Arechavala-Gomez, V. (2020). Researcher's Perceptions on Publishing "Negative" Results and Open Access. *Nucleic Acid Therapeutics*, 31(3), 185–189. <https://doi.org/10.1089/nat.2020.0865>

Edgar, R., Domrachev, M., y Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, 30(1),

- 207–210. <https://doi.org/10.1093/nar/30.1.207>
- Elliott, K. C., y Resnik, D. B. (2019). Making Open Science Work for Science and Society. *Environmental Health Perspectives*, 127(7), 075002.
<https://doi.org/10.1289/ehp4808>
- Elsevier. (2019). Research data. Recuperado de
<https://www.elsevier.com/about/policies/research-data>
- English, R., y Joseph, H. (2008). An open access landmark. *CyRL News*, (February), 82–85.
- European Commission. (s.f.). Data management. Recuperado desde
https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm#
- European Commission. (2016a). H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020, (July 2016).
- European Commission. (2016b). H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020, (July 2016). Recuperado de
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf?utm_content=bufferc22c5yutm_medium=socialyutm_source=twitter.comyutm_campaign=buffer
- European Commission. (2019). H2020 Online Manual. Recuperado de
https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm
- European Commission. (2019). Open Access. Recuperado desde
<https://ec.europa.eu/research/openscience/index.cfm?pg=openaccess>
- Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLoS ONE*, 5(4).
<https://doi.org/10.1371/journal.pone.0010271>
- Faschinger, C. (2019). Do not hesitate and publish negative results and look for long-

- term results! *Graefe's Archive for Clinical and Experimental Ophthalmology*, 257(12), 2697–2698. <https://doi.org/10.1007/s00417-019-04465-1>
- Faure, D., y Joly, D. (2016). NGS Data Sharing and Access. In *Insight on Environmental Genomics* (pp. 27–35). <https://doi.org/10.1016/b978-1-78548-146-8.50003-4>
- Federer, L. M., Belter, C. W., Joubert, D. J., Livinski, A., Lu, Y. L., Snyders, L. N., y Thompson, H. (2018). Data sharing in PLOS ONE: An analysis of Data Availability Statements. *PLoS ONE*, 13(5), 1–12. <https://doi.org/10.1371/journal.pone.0194768>
- Ferguson, L. (2014). How and Why Researchers Share Data (and Why They don't). Recuperado de <https://www.wiley.com/network/researchers/licensing-and-open-access/how-and-why-researchers-share-data-and-why-they-dont>
- Fernández Esquinas, M. (1998). *Entrada: entrevista; Diccionario de Sociología*. Madrid: Alianza Editorial.
- Fernandez, M. E., McCurdy, S. A., Arvey, S. R., Tyson, S. K., Morales-Campos, D., Flores, B., y Sanderson, M. (2009). HPV knowledge, attitudes, and cultural beliefs among Hispanic men and women living on the Texas-Mexico border. *Ethnicity and Health*, 14(6), 607–624. <https://doi.org/10.1080/13557850903248621>
- Ferrante, M., Saulle, R., Ledda, C., Pappalardo, R., Fallico, R., La Torre, G., y Fiore, M. (2013). Prevalence of smoking habits, attitudes, knowledge and beliefs among Health Professional School students: a cross-sectional study. *Ann Ist Super Sanità*, 49(4), 143–149. <https://doi.org/10.4415/ANN>
- Figshare. (2019). About Figshare. Recuperado desde <https://figshare.com/about>
- Fiske, M., Kendall, P. L., y Merton, R. K. (2002). Propósitos y criterios de la entrevista focalizada (traducción de Consuelo del Val y Javier Callejo). *Empiria. Revista de Metodología de Ciencias Sociales*. <https://doi.org/10.5944/empiria.1.1998.740>
- Forni, F., Frey, A., y Quaranta, G. (2012). Frédéric Le Play : Un precursor de las Metodologías Cualitativas en Ciencias Sociales. *Miríada*, 1(1), 59–103. Recuperado de <https://p3.usal.edu.ar/index.php/miriada/article/view/429>

- Fountadion, O. K. (s.f.). ¿Qué son los datos abiertos? Recuperado desde <https://opendatahandbook.org/guide/es/>
- Garrison, N. A., Barton, K. S., Porter, K. M., Mai, T., Burke, W., y Carroll, S. R. (2019). Access and Management: Indigenous Perspectives on Genomic Data Sharing. *Ethnicity y Disease*, 29, 659–668. <https://doi.org/10.18865/ed.29.S3.659>
- Generalitat de Catalunya. (s.f.). ¿Qué son los datos abiertos? Recuperado desde http://governobert.gencat.cat/es/dades_obertes/que-son-les-dades-obertes/
- Gewin, V. (2016). An open mind on open data. *Nature*, 529(7584), 117–119. <https://doi.org/10.1038/NJ7584-117A>
- Giofrè, D., Cumming, G., Fresc, L., Boedker, I., y Tressoldi, P. (2017). The influence of journal submission guidelines on authors' reporting of statistics and use of open research practices. *PLoS ONE*, 12(4), 1–15. <https://doi.org/10.1371/journal.pone.0175583>
- González Alcaide, G., y Ferri, J. G. (2014). La colaboración científica: principales líneas de investigación y retos de futuro ; Scientific collaboration: main research lines and future challenges. *Revista Española de Documentación Científica* 37(4), Octubre-Diciembre 2014, 37(4), e062. <https://doi.org/10.3989/redc.2014.4.1186>
- González, L.-M., Saorín, T., Ferrer-Sapena, A., Aleixandre-Benavent, R., y Peset, F. (2013). Gestión de datos de investigación: infraestructuras para su difusión. *El Profesional de La Información*, 22(5), 415–423. <https://doi.org/10.3145/epi.2013.sep.06>
- Gøtzsche, P. C. (2012). Strengthening and Opening Up Health Research by Sharing Our Raw Data. *Circulation: Cardiovascular Quality and Outcomes*, 5(2), 236–237. <https://doi.org/10.1161/circoutcomes.112.965277>
- Government Publishing Office. Consolidated Appropriations Act, 2008. United States of America.
- Goyanes, M., y Rodríguez-Gómez, E. F. (2018). Why do we publish? Prevalence, motivations, and consequences of publish or perish. *Profesional de La*

- Informacion*, 27(3), 548–558. <https://doi.org/10.3145/epi.2018.may.08>
- Grando, A., Murcko, A., Kaufman, D., Whitfield, M. J., y Maupin, J. (2020). Mental health professional perspectives on health data sharing: Mixed methods study. *Health Informatics Journal*, 1–16. <https://doi.org/10.1177/1460458219893848>
- Grupo de editores de revistas españolas sobre Ciencias de la Salud. (2018). Declaración de Sant Joan d’Alacant en defensa del Acceso Abierto a las publicaciones científicas, del grupo de editores de revistas españolas sobre ciencias de la salud (GERECS). *Revista Española de Nutrición Humana y Dietética*, 22(1), 1–3.
- Haeusermann, T., Fadda, M., Blasimme, A., Greshake, B., Vayena, E. (2018). Genes wide open : Data sharing and the social gradient of genomic privacy. *AJOB Empirical Bioethics*, 9(4), 207–221. <https://doi.org/10.1080/23294515.2018.1550123>
- Hernández-Pérez, T., y García-Moreno, M. A. (2013). Datos abiertos y repositorios de datos: nuevo reto para los bibliotecarios. *Profesional de La Informacion*, 22(3), 259–263. <https://doi.org/10.3145/epi.2013.may.10>
- Hernández, M., Quijada, N. M., Rodríguez-Lázaro, D., y Eiros, J. M. (2020). Bioinformatics of next generation sequencing in clinical microbiology diagnosis. *Revista Argentina de Microbiología*, 52(2), 150–161. <https://doi.org/10.1016/j.ram.2019.06.003>
- Hofner, B., Schmid, M., y Edler, L. (2016). Reproducible research in statistics: A review and guidelines for the Biometrical Journal. *Biometrical Journal*, 58(2), 416–427. <https://doi.org/10.1002/bimj.201500156>
- Holroyd, E. A., Taylor-Piliae, R. E., y Twinn, S. F. (2003). Investigating Hong Kong’s Filipino domestic workers’ healthcare behavior, knowledge, beliefs and attitudes towards cervical cancer and cervical screening. *Women and Health*, 38(1), 69–82. https://doi.org/10.1300/J013v38n01_05
- Huang, C., Yue, X., Chen, J., Xu, W., y Li, J. (2017). The effect of “open access” on journal impact factors: A causal analysis of medical journals. *ISSI 2017 - 16th International Conference on Scientometrics and Informetrics, Conference*

- Proceedings*, 533, 611–621. <https://doi.org/10.1016/j.physa.2019.122043>
- Hunt, S. L., y Bakker, C. J. (2018). A qualitative analysis of the information science needs of public health researchers in an academic setting. *Journal of the Medical Library Association*, 106(2), 184–197. <https://doi.org/10.5195/jmla.2018.316>
- Ibáñez, J. (1986). *Más allá de la Sociología: el grupo de discusión: técnica y crítica*. Madrid: Siglo XI. Madrid: Siglo XXI.
- Inkpen, R., Gauci, R., y Gibson, A. (2020). The values of open data. *Area*, 1–7. <https://doi.org/10.1111/area.12682>
- International Cancer Genome Consortium. (2019). Data Access Compliance Office. Recuperado desde <https://icgc.org/daco>
- Juan Voutssas, M. (2012). Aspectos para el desarrollo de una revista científica digital. *Investigacion Bibliotecologica*, 26(58), 71–100. <https://doi.org/10.22201/iibi.0187358xp.2012.58.35253>
- Kim, Y. (2017). Fostering scientists' data sharing behaviors via data repositories, journal supplements, and personal communication methods. *Information Processing and Management*, 53(4), 871–885. <https://doi.org/10.1016/j.ipm.2017.03.003>
- Kim, Y., y Burns, C. S. (2016). Norms of data sharing in biological sciences: The roles of metadata, data repository, and journal and funding requirements. *Journal of Information Science*, 42(2), 230–245. <https://doi.org/10.1177/0165551515592098>
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data and Society*, 1(1), 1–12. <https://doi.org/10.1177/2053951714528481>
- Klump, J., Bertelmann, R., Brase, J., Diepenbroek, M., Grobe, H., Höck, H., ... Wächter, J. (2006). Data publication in the open access initiative. *Data Science Journal*, 5(June), 79–83. <https://doi.org/10.2481/dsj.5.79>
- Knoppers, B. M., y Joly, Y. (2018). Introduction: the why and whither of genomic data sharing. *Human Genetics*, 137(8), 569–574. <https://doi.org/10.1007/s00439-018-1923-y>
- Knoppers, Bartha M, Harris, J. R., Budin-Ijøsne, I., y Dove, E. S. (2014). A human rights

- approach to an international code of conduct for genomic and clinical data sharing. *Hum Genet*, 133, 895–903. <https://doi.org/10.1007/s00439-014-1432-6>
- Kurata, K., Morioka, T., Yokoi, K., y Matsubayashi, M. (2013). Remarkable Growth of Open Access in the Biomedical Field: Analysis of PubMed Articles from 2006 to 2010. *PLoS ONE*, 8(5), e60925– e60925. <https://doi.org/10.1371/journal.pone.0060925>
- Laakso, M., Welling, P., Bukvova, H., Nyman, L., Björk, B. C., y Hedlund, T. (2011). The development of open access journal publishing from 1993 to 2009. *PLoS ONE*, 6(6), e20961- e20961. <https://doi.org/10.1371/journal.pone.0020961>
- Leung, T. I., y Dumontier, M. (2019). FAIR Principles for Clinical Practice Guidelines in a Learning Health System. *Studies in Health Technology and Informatics*, 264, 1690–1691. <https://doi.org/10.3233/SHTI190599>
- Linek, S. B., Fecher, B., Friesike, S., y Hebing, M. (2017). Data sharing as social dilemma: Influence of the researcher’s personality. *PLoS ONE*, 1-4. <https://doi.org/10.1371/journal.pone.0183216>
- Longo, D. L., y Drazen, J. M. (2016). Data Sharing. *The New England Journal of Medicine*, 374(3), 276–277. <https://doi.org/10.1056/NEJMe1516564>
- Mannheimer, S., Pienta, A., Kirilova, D., Elman, C., y Wutich, A. (2019). Qualitative Data Sharing: Data Repositories and Academic Libraries as Key Partners in Addressing Challenges. *American Behavioral Scientist*, 63(5), 643–664. <https://doi.org/10.1177/0002764218784991>
- Max-Planck-Gesellschaft. (2003). Berlin declaration on open access to knowledge in the sciences and humanities. In *Berlin Open Access Conference, Max Planck Society and Max Planck Institute for the History of Science* (p. 4). Berlin.
- Mayernik, M. S. (2019). Metadata accounts: Achieving data and evidence in scientific research. *Social Studies of Science*, 49(5), 732–757. <https://doi.org/10.1177/0306312719863494>
- McDonald, L., Schultze, A., Simpson, A., Graham, S., Wasiak, R., y Ramagopalan, S.

- (2017). A review of data sharing statements in observational studies published in the BMJ: A cross-sectional study. *F1000Research*, 6(0), 1708.
<https://doi.org/10.12688/f1000research.12673.1>
- McKiernan, G. (2000). arXiv.org: the Los Alamos National Laboratory e-print server. *International Journal on Grey Literature*, 1(3), 127–138.
<https://doi.org/10.1108/14666180010345564>
- Melero, R. (2005). Acceso abierto a las publicaciones científicas: definición, recursos, copyright e impacto. *El Profesional de La Información*, 14(4), 255–266.
- Melero, R., y Abad Garcia, M. F. (2008). Revistas open access: características, modelos económicos y tendencias. *Bid. Textos Universitaris De Biblioteconomia I Documentació*, (20), 17.
- Michener, W. K. (2015). Ten Simple Rules for Creating a Good Data Management Plan. *PLoS Computational Biology*, 11(10), 1–9.
<https://doi.org/10.1371/journal.pcbi.1004525>
- Miller, A. N., Taylor, S. G., y Bedeian, A. G. (2011). Publish or perish: academic life as management faculty live it. *Career Development International*, 16(5), 422–445.
<https://doi.org/10.1108/13620431111167751>
- Milne, M. J., y Adler, R. W. (1999). Exploring the reliability of social and environmental disclosures content analysis. *Accounting, Auditing y Accountability Journal*, 12(2), 237–256.
- Ministerio de Ciencia e Innovación. Resolución de 11 de noviembre de 2020 de la Presidencia de la Agencia Estatal de Investigación, por la que se aprueba la convocatoria 2020 de ayudas a «Proyectos de I+D+i» (2020). Spain. Recuperado de
https://www.ciencia.gob.es/stfls/MICINN/Ayudas/PE_2017_2020/PE_Orientada_Retos_Sociedad/FICHEROS/Proyectos_IDI_Retos_Investigacion/ConvocatoriaPID2020_Resolucion20201111.pdf
- Ministerio de Industria Economía y Competividad. Plan Estatal de Investigación Científica y Tecnológica y de Innovación 2017-2020 (2017). España. Recuperado

- de
<http://www.ciencia.gob.es/stfls/MICINN/Prensa/FICHEROS/2018/PlanEstatallDI.pdf>
- Morain, S. R., Kass, N. E., y Grossmann, C. (2017). What allows a health care system to become a learning health care system: Results from interviews with health system leaders. *Learn Health System*, 1, 1–8. <https://doi.org/10.1002/lrh2.10015>
- National Institutes of Health. (2015). *National Institutes of Health Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research*. Recuperado de <http://www.ncbi.nlm.nih.gov/pmc/>
- National Institutes of Health. (2018). Request for Information (RFI) on Proposed Provisions for a Draft Data Management and Sharing Policy for NIH Funded or Supported Research. Recuperado desde <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-19-014.html>
- Nature Communications. (2018). Data sharing and the future of science. *Nature Communications*, 9(1), 9–10. <https://doi.org/10.1038/s41467-018-05227-z>
- Nature Genetics. (2016). FAIR principles for data stewardship. *Nature Genetics*, 48(4), 343. <https://doi.org/10.1038/ng.3544>
- Navarro-Molina, C., y Melero, R. (2019). Motivación, barreras e incentivos para la compartición y reutilización de los datos de investigación. Visión de los investigadores. *El Profesional de La Información*, 28(5), 1–15. <https://doi.org/10.3145/epi.2019.sep.16>
- Nowacki, A. S. (2013). Data sharing and the development of the cleveland clinic statistical education dataset repository. *Journal of Statistics Education*, 21(1). <https://doi.org/10.1080/10691898.2013.11889660>
- OECD. Declaration on Access to Research Data From Public Funding, OECD § (2004). Recuperado de <http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=157>
- OECD. (2019). Recommendation of the Council on Resource Productivity, (April).

- Office Management and Budget. Uniform Administrative Requirements for Grants and Agreements With Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations (1999). EEUU. Recuperado de <https://georgewbush-whitehouse.archives.gov/omb/circulars/a110/a110.html>
- Olson, S., y Downey, A. S. (2013). *Sharing clinical research data*. Washington, DC: The National Academies Press.
- Open Society Institute. (2002). Budapest Open Access Initiative. Recuperado desde <https://www.budapestopenaccessinitiative.org/read>
- Opoku, S. Y., Benwell, M., y Yarney, J. (2012). Knowledge, attitudes, beliefs, behaviour and breast cancer screening practices in Ghana, West Africa. *Pan African Medical Journal*, 11, 28. <https://doi.org/10.11604/pamj.2012.11.28.548>
- Ostherr, K., Borodina, S., Bracken, R. C., Lotterman, C., Storer, E., y Williams, B. (2017). Trust and privacy in the context of user-generated health data. *Big Data y Society*, 1–11. <https://doi.org/10.1177/2053951717704673>
- Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., ... Dierolf, U. (2013). Making research data repositories visible: The re3data.org registry. *PLoS ONE*, 8(11). <https://doi.org/10.1371/journal.pone.0078080>
- Pereira, S., Gibbs, R. A., y McGuire, A. L. (2014). Open access data sharing in genomic research. *Genes*, 5(3), 739–747. <https://doi.org/10.3390/genes5030739>
- Peset, F., Aleixandre-Benavent, R., Blasco-Gil, Y., y Ferrer-Sapena, A. (2017). Datos abiertos de investigación. Camino recorrido y cuestiones. *Anales de Documentacion*, 20(1), 1–12. <https://doi.org/10.6018/ANALESDOC.20.1.272101>
- Petro-Nustas, W., Kulwicki, A., y Zumout, A. F. (2002). Students' knowledge, attitudes, and beliefs about AIDS: A cross-cultural study. *Journal of Transcultural Nursing*, 13(2), 118–125. <https://doi.org/10.1177/104365960201300204>
- Piwowar, H. A., y Chapman, W. W. (2008). A review of journal policies for sharing research data. In *Open Scholarship: Authority, Community, and Sustainability in the Age of Web 2.0 - Proceedings of the 12th International Conference on*

- Electronic Publishing, ELPUB 2008* (pp. 1–14).
<https://doi.org/10.1038/npre.2008.1700.1>
- PLOS One. (2019). Data Availability. Recuperado de
<https://journals.plos.org/plosone/s/data-availability>
- Plummer, K. (1983). *Documents of Life*. London: Allen.
- Pop, M., y Salzberg, S. L. (2015). Use and mis-use of supplementary material in science publications. *BMC Bioinformatics*, 16(1), 1–4. <https://doi.org/10.1186/s12859-015-0668-z>
- Popkin, G. (2019). Data sharing and how it can benefit your scientific career. *Nature*, 569(7756), 445–447. <https://doi.org/10.1038/d41586-019-01506-x>
- Price, A., Schroter, S., Clarke, M., y McAneney, H. (2018). Role of supplementary material in biomedical journal articles: Surveys of authors, reviewers and readers. *BMJ Open*, 8(9), 1–7. <https://doi.org/10.1136/bmjopen-2018-021753>
- PubMed Central. (2020). PMC Overview. Recuperado desde
<https://www.ncbi.nlm.nih.gov/pmc/about/intro/>
- Quinn, M., Forman, J., Harrod, M., Winter, S., Fowler, K. E., Krein, S. L., ... Arbor, A. (2019). Electronic health records, communication, and data sharing: challenges and opportunities for improving the diagnostic process. *Diagnosis*, 6(3), 241–248. <https://doi.org/10.1515/dx-2018-0036.E>
- Real Academia Española. (2019). Real Academia Española. Madrid. Recuperado de
<https://dle.rae.es/?id=LYf3Ibz>
- Reinhardt, C., y Cook, T. (1979). *Beyond qualitative versus quantitative methods C S, Thomas D Cook*. Beverly Hills: Sage.
- Riley, J. (2017). *Understanding metadata. What is Metadata, and what is for?* Baltimore: National Information Standards Organization (NISO). Recuperado de
http://www.niso.org/apps/group_public/download.php/17446/UnderstandingMetadata.pdf
- Robinson-Garcia, N., Mongeon, P., Jeng, W., y Costas, R. (2017). DataCite as a novel

- bibliometric source: Coverage, strengths and limitations. *Journal of Informetrics*, 11, 841–854 . <https://doi.org/10.1016/j.joi.2017.07.003>
- Rokeach, M. (1970). *Beliefs, Attitudes, and Values*. San Francisco: Jossey-Bass.
- Ruiz Olabuénaga, J. (2012). *Metodología de la investigación cualitativa*. Bilbao: Universidad de Deusto.
- Ruiz Olabuénaga, J. (2012). *Teoría y práctica de la investigación cualitativa*. Bilbao: Universidad de Deusto.
- San José Montano, B. (2009). Acceso abierto (open access), un modelo necesario de comunicación científica. *Pediatría de Atención Primaria*, 11(42), 299–311. <https://doi.org/10.4321/s1139-76322009000200011>
- Sandelowski, M., y Leeman, J. (2012). Writing usable qualitative health research findings. *Qualitative Health Research*, 22(10), 1404–1413. <https://doi.org/10.1177/1049732312450368>
- Saunders, P. A., Wilhelm, E. E., Lee, S., y Merkhofer, E. (2014). Data sharing for public health research : A qualitative study of industry and academia. *Communication y Medicine*, 11(2), 179–187. <https://doi.org/10.1558/cam.v11i2.18310>
- Sayogo, D. S., y Pardo, T. A. (2013). Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. *Government Information Quarterly*, 30(1), 19–31. <https://doi.org/10.1016/j.giq.2012.06.011>
- Schmitt, N., Olsen, J. E., Hansen, H. S., Høyer, K. L., Moresco, K., Sørensen, A. H., y Boer, S. den. (2016a). *Policy for Research Data Management*. Recuperado de https://healthsciences.ku.dk/research/responsible-conduct-of-research/SUND_policy_for_research_data_management_FINAL.pdf
- Schmitt, N., Olsen, J. E., Hansen, H. S., Høyer, K. L., Moresco, K., Sørensen, A. H., y Boer, S. den. (2016). *Policy for Research Data Management*.
- Scientific Software Development GmbH. (2021). Atlas.ti. Berlin: ATLAS.ti GmbH.
- Shahin, M. H., Bhattacharya, S., Silva, D., Kim, S., Burton, J., Podichetty, J., Klaus Romero, D., y Conrado, D. J. (2020). Open Data Revolution in Clinical Research:

- Opportunities and Challenges. *Clinical and Translational Science*, 13(4), 665–674.
<https://doi.org/10.1111/cts.12756>
- Sieber, J. (2015). Data Sharing in Historical Perspective. *Journal of Empirical Research on Human Research Ethics*, 10(3), 215–216.
<https://doi.org/10.1177/1556264615594607>
- Sieber, J.. (1988). Data sharing - Defining problems and seeking solutions. *Law and Human Behavior*, 12(2), 199–206. <https://doi.org/10.1007/BF01073128>
- Siegel, M. G., Brand, J. C., Rossi, M. J., y Lubowitz, J. H. (2018). “Publish or Perish” Promotes Medical Literature Quantity Over Quality. *Arthroscopy - Journal of Arthroscopic and Related Surgery*, 34(11), 2941–2942.
<https://doi.org/10.1016/j.arthro.2018.08.029>
- Sixto-Costoya, A, Aleixandre-Benavent, R., Lucas-Domínguez, R., y Vidal-Infer, A. (2020). The Emergency Medicine Facing the Challenge of Open Science. *Data*, 5(28), 7. <https://doi.org/10.3390/data5020028>
- Sixto-Costoya, A, Aleixandre-Benavent, R., Vidal-Infer, A., Lucas-Domínguez, R., y Castelló-Cogollos, L. (2019). *Data sharing: qué son y cómo pueden compartirse los datos de investigación. Manual de recomendación para gestores de la información* (Vol. 53). Sociedad Española de Documentación e Información Científica.
- Sixto-Costoya, A, Lucas-Domínguez, R., Aleixandre-Benavent, R., y Vidal-Infer, A. (2021). Is Sharing Datasets the Answer to the New Challenges of Reproductive Biology Research? *Reproductive Sciences*, 10–12. <https://doi.org/10.1007/s43032-021-00484-8>
- Sixto-Costoya, A., Robinson-Garcia, N., van Leeuwen, T. N., & Costas, R. (2021). Exploring the relevance of ORCID as a source of study of data sharing activities at the individual-level: a methodological discussion. *Scientometrics*.
<https://doi.org/10.1007/s11192-021-04043-5>
- Smith, E., Haustein, S., Mongeon, P., Shu, F., Ridde, V., y Larivière, V. (2017). Knowledge sharing in global health research – the impact, uptake and cost of

- open access to scholarly literature. *Health Research Policy and Systems*, 15(1), 73.
<https://doi.org/10.1186/s12961-017-0235-3>
- Sociedad Max Planck. (2003). La Declaración de Berlín sobre acceso abierto. *GeoTrópico*, 1(2), 152–154.
- Stanley, B., y Stanley, M. (1988). Data sharing: The primary researcher's perspective. *Law and Human Behavior*, 12(2), 173–180.
- Suber, P. (2002). Open access to the scientific journal literature. *Journal of Biology*, 1(1), 3. <https://doi.org/10.1186/1475-4924-1-3>
- Suber, P. (2004). A Very Brief Introduction to Open Access.
- Suber, P. (2008). An open access mandate for the National Institutes of Health. *Open Medicine*, 2(2), 14–16.
- Szajnfarber, Z., y Gralla, E. (2017). Qualitative methods for engineering systems: Why we need them and how to use them. *Systems Engineering*, 20(6), 497–511.
<https://doi.org/10.1002/sys.21412>
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., y Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE*, 6(6), 1–21. <https://doi.org/10.1371/journal.pone.0021101>
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., y Dorsett, K. (2015). Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLoS ONE*, 10(8), 1–24.
<https://doi.org/10.5061/dryad.1ph92>
- Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., Christian, L., Grant, B., y Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLoS ONE*, 15(3), 1–26.
<https://doi.org/10.1371/journal.pone.0229003>
- Thelwall, M., y Kousha, K. (2017). Do journal data sharing mandates work? Life sciences evidence from Dryad. *Aslib Journal of Information Management*, 69(1), 36–45. <https://doi.org/10.1108/AJIM-09-2016-0159>

- Torres-Salinas, D., Robinson-García, N., y Cabezas-Clavijo, Á. (2012). Compartir los datos de investigación en ciencia: introducción al *data sharing*. *El Profesional de La Información*, 21(2), 173–184. <https://doi.org/10.3145/epi.2012.mar.08>
- Tuckerman, J., Kaufman, J., y Danchin, M. (2020). How to use qualitative methods for health and health services research. *Journal of Paediatrics and Child Health*, 56(5), 818–820. <https://doi.org/10.1111/jpc.14849>
- Ulrichsweb. (2019). Ulrichsweb. Recuperado desde <https://ulrichsweb.serialssolutions.com/search/-1987618610>
- University of Bath. (s.f.). Research Data Service Homepage: Research Data Home. Recuperado desde <https://library.bath.ac.uk/research-data>
- University of Leeds. (s.f.). Research data management explained. Recuperado desde https://library.leeds.ac.uk/info/14062/research_data_management/61/research_data_management_explained
- Valles, M. (2000). *Técnicas cualitativas de investigación social. Reflexión metodológica y práctica profesional*. Madrid: Síntesis.
- Van Maanen, J. (1983). *Qualitative Methodology*. London: Sage.
- Van Panhuis, W. G., Paul, P., Emerson, C., Grefenstette, J., Wilder, R., Herbst, A. J., ... Burke, D. S. (2014). A systematic review of barriers to data sharing in public health. *BMC Public Health*, 14(1), 1–9. <https://doi.org/10.1186/1471-2458-14-1144>
- Vargas Arbeláez, E. J. (2014). La política editorial de acceso abierto: horizontes de la democratización del conocimiento para el desarrollo. *Equidad y Desarrollo*, (21), 145. <https://doi.org/10.19052/ed.2908>
- Vassilakopoulou, P., y Aanestad, M. (2019). Communal data work: Data sharing and re-use in clinical genetics. *Health Informatics Journal*, 25(3), 511–525. <https://doi.org/10.1177/1460458219833117>
- Vidal-Infer, A., Aleixandre-Benavent, R., Lucas-Domínguez, R., y Sixto-Costoya, A. (2019). The availability of raw data in substance abuse scientific journals. *Journal*

- of Substance Use*, 24(1), 1–5. <https://doi.org/10.1080/14659891.2018.1489905>
- Vidal-Infer, A., Tarazona, B., Alonso-Arroyo, A., y Aleixandre-Benavent, R. (2018). Public availability of research data in dentistry journals indexed in Journal Citation Reports. *Clinical Oral Investigations*, 22(1), 275–280. <https://doi.org/10.1007/s00784-017-2108-0>
- Vivar, C. G., Arantzamendi, M., López-Dicastillo, O., y Gordo Luis, C. (2010). La Teoría Fundamentada como Metodología de Investigación Cualitativa en Enfermería. *Index de Enfermería*, 19(4), 283–288. Recuperado de http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1132-12962010000300011&lng=es&synrm=isoyt&lng=es
- Wang, Z., Lachmann, A., y Ma'ayan, A. (2019). Mining data and metadata from the gene expression omnibus. *Biophysical Reviews*, 11(1), 103–110. <https://doi.org/10.1007/s12551-018-0490-8>
- Wicherts, J. M., Borsboom, D., Kats, J., y Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61(7), 726–728. <https://doi.org/10.1037/0003-066X.61.7.726>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Mons, B. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 1–9. <https://doi.org/10.1038/sdata.2016.18>
- Wilkinson, T., Sinha, S., Peek, N., y Geifman, N. (2019). Clinical trial data reuse - Overcoming complexities in trial design and data sharing. *Trials*, 20(1), 10–13. <https://doi.org/10.1186/s13063-019-3627-6>
- World Medical Association. (2013). World Medical Association Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects. *Journal of the American Medical Association*, 310(20), 2191–2194. <https://doi.org/10.1001/jama.292.11.1359>
- Zenodo. (2019). About Zenodo. Recuperado desde <https://about.zenodo.org>

Zenodo. (2021). General Policies v1.0. Recuperado desde
<https://about.zenodo.org/policies/>

Zhu, Y. (2020). Open-access policy and data-sharing practice in UK academia. *Journal of Information Science*, 46(1), 41–52. <https://doi.org/10.1177/0165551518823174>

Anexos

1. El guion

CONOCIMIENTOS

1. Contextualización de los datos de investigación
2. Proyectos/programas/publicaciones en revistas etc, en los que estás implicado recomienda o exige que se lleve a cabo esta práctica
3. Papel de la institución/organización a la hora de ofrecer recursos en forma de guías, soporte técnico, formación, supervisión...
4. Diferencia entre financiación pública y privada en los datos brutos
5. Conocimiento sobre infraestructuras digitales para compartir datos
6. Planes y protocolos
7. Incentivos

CREENCIAS

8. Desde una perspectiva global de la ciencia, posible importancia de compartir datos
9. En su área, posible importancia de compartir datos
10. Dificultad para integrar esta práctica
11. Valoración de los incentivos. Coste – Beneficio
12. Barreras tecnológicas (accesibilidad y sencillez de las infraestructuras)
13. Apoyo institucional, valoración
14. Aspectos positivos y negativos según su opinión
15. Ventajas
16. Inconvenientes

ACTITUDES

17. Nivel de disposición a la hora de compartir

18. Preferencias técnicas y formales relacionadas con el proceso de compartir los datos. Qué te gustaría que una dataset contuviera para ser comprensible, reusable e interoperable.

2. Las redes

A continuación, se muestran las redes realizadas con el software Atlas.ti a partir de los fragmentos de las entrevistas que se consideraron relevantes de acuerdo con las etiquetas diseñadas.

La organización de las redes es la siguiente:

- En un primer nivel, se diferencian las tres dimensiones de análisis estudiadas: los conocimientos, las creencias y las actitudes de las investigadores e investigadores.
- En un segundo nivel, se diferencian por tipo de investigación (básica y aplicada).
- En un tercer nivel, se subdivide en investigadores e investigadoras júnior y sénior.

Cabe señalar que la organización se presenta de esta manera para facilitar su lectura y comprensión, pero que no es inamovible. Para realizar en análisis de los resultados fue necesario combinar las redes de acuerdo con las necesidades de información que iban surgiendo para contestar a los objetivos del estudio. Por ejemplo, para indagar acerca de las creencias de el personal investigador júnior fue necesario disponer de todas las redes confeccionadas a partir de las entrevistas a este grupo de investigadores e investigadoras independientemente del tipo de investigación. O, si lo que se pretendía es averiguar acerca de las actitudes en general de el personal investigador sobre su disposición a compartir datos se necesitó resituar las redes alrededor de esta dimensión de análisis antes para luego poder diferenciar por tipo de investigación. De este modo, mediante relocalaciones y re combinaciones de las redes, se fueron dando contestación a todos los interrogantes surgidos.

Por otro lado, en referencia al diseño de la red, mencionar que cada investigador e investigadora tiene tres redes, cada una correspondiente con una dimensión de análisis

(conocimientos, actitudes o creencias). En el centro de cada red se presenta la etiqueta correspondiente a cada persona entrevistada y, alrededor de la misma, se muestran con diseño orgánico-radial los fragmentos de las entrevistas que se consideraron relevantes para esa dimensión. Estos fragmentos forman los discursos que luego sirven para detectar los patrones y las tendencias que son presentadas en la sección de “Resultados”. Las frases de las entrevistas que son tomadas como ejemplo en los resultados forman parte de estas redes y fueron seleccionadas por resultar especialmente relevantes e ilustrativas. Además, en las redes pueden presentarse relaciones entre los fragmentos. A este respecto, las relaciones entre los fragmentos (que pueden ser frases o párrafos) son:

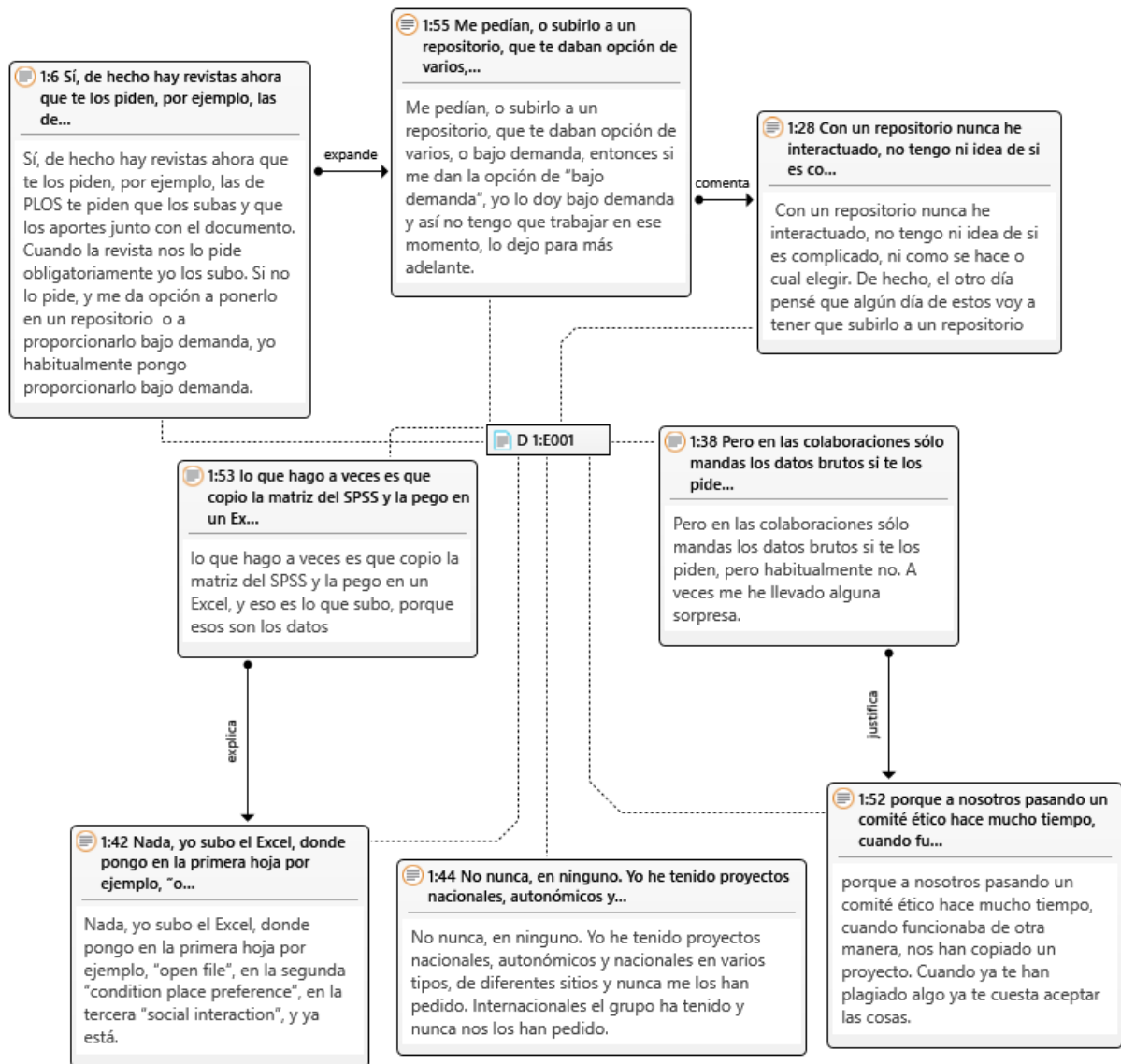
- **“Comenta”**: se utiliza esta relación cuando un fragmento toma una idea de otro y continúa una línea argumentativa parecida sin cambiar de tema.
- **“Explica”**: se utiliza cuando un fragmento sirve para explicar a otro, ofreciendo una ampliación de la información, un mayor nivel de detalle.
- **“Justifica”**: se utiliza cuando un fragmento justifica lo que se dijo en otro. Sirve para cuando se introduce algún tema con potencial de ser controvertido.
- **“Expande”**: es similar a la relación “explica”, pero se diferencia en que, mientras que la explicación supone entrar en un mayor de detalle, la expansión no necesariamente trata el tema con más detenimiento, sino que sería una continuación.
- **“Contradice”**: se utiliza cuando un fragmento niega de alguna lo que se dice en otro.

CONOCIMIENTOS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 001 (E001)

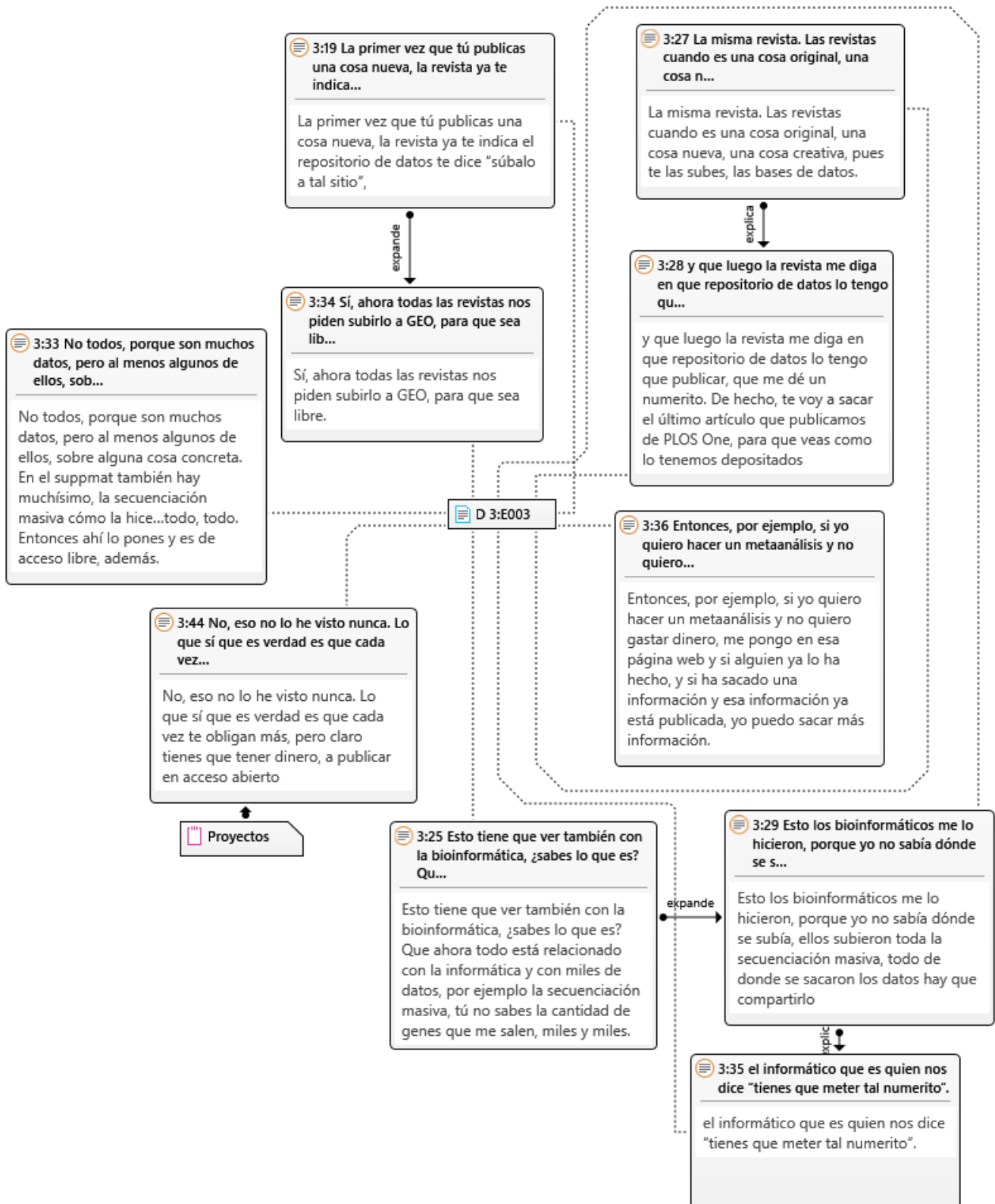


CONOCIMIENTOS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 003 (E003)

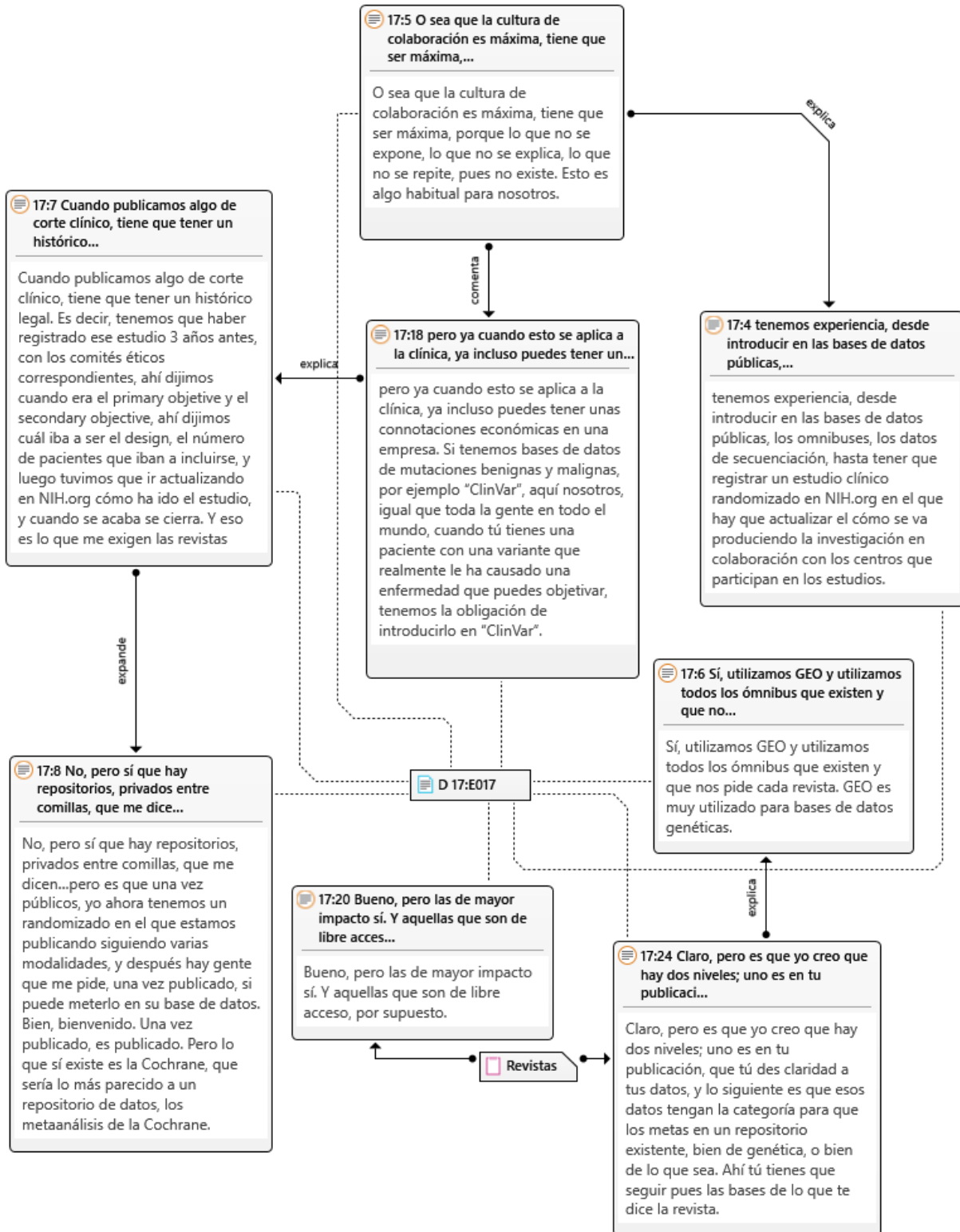


CONOCIMIENTOS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADO 017 (E017)

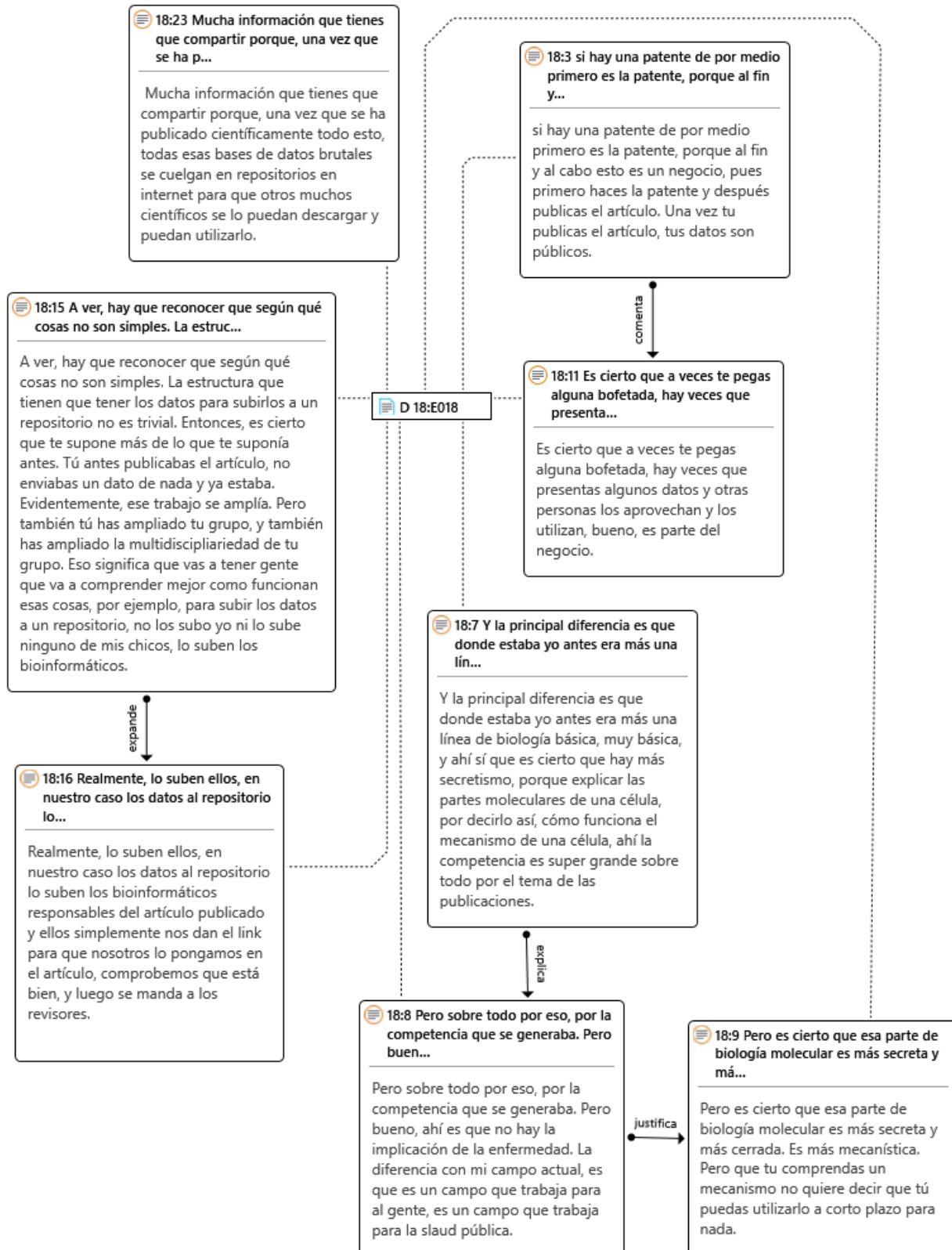


CONOCIMIENTOS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADO 018 (E018)

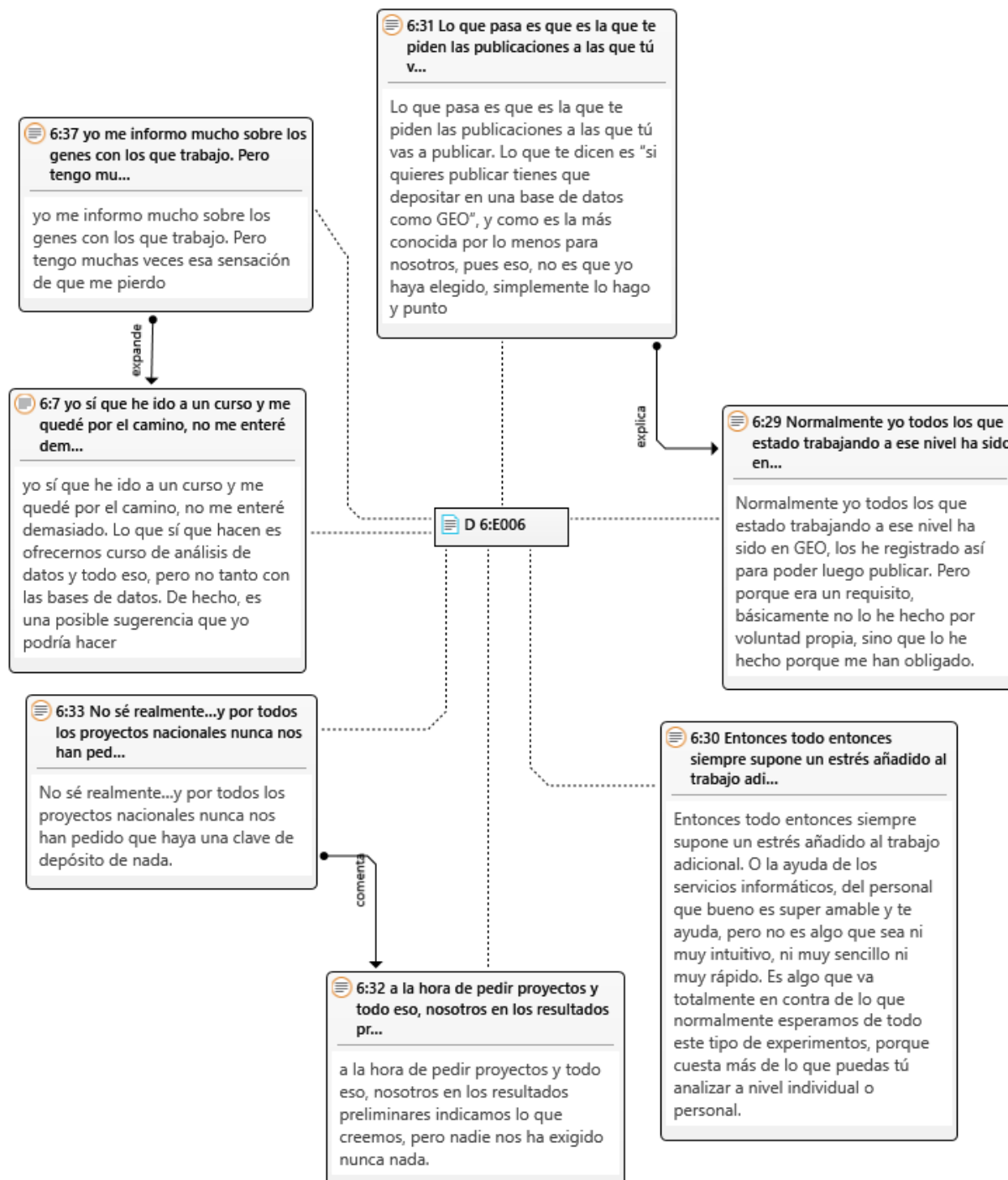


CONOCIMIENTOS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 006 (E006)

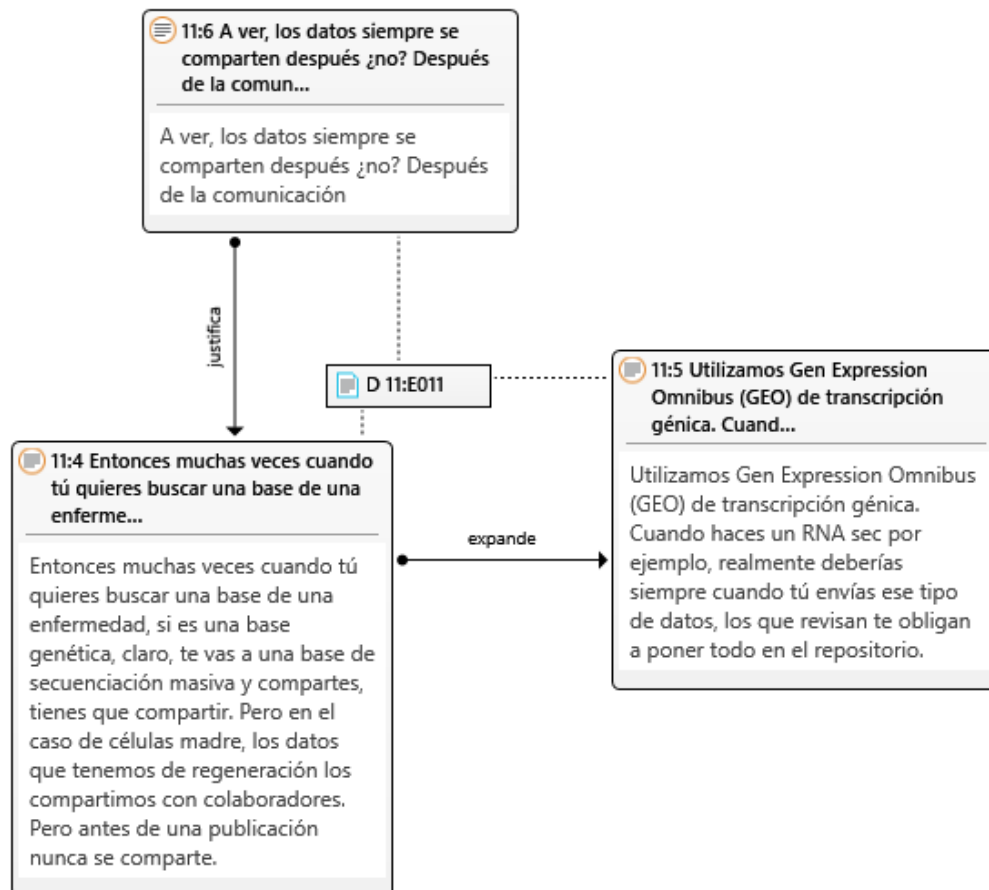


CONOCIMIENTOS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 011 (E011)

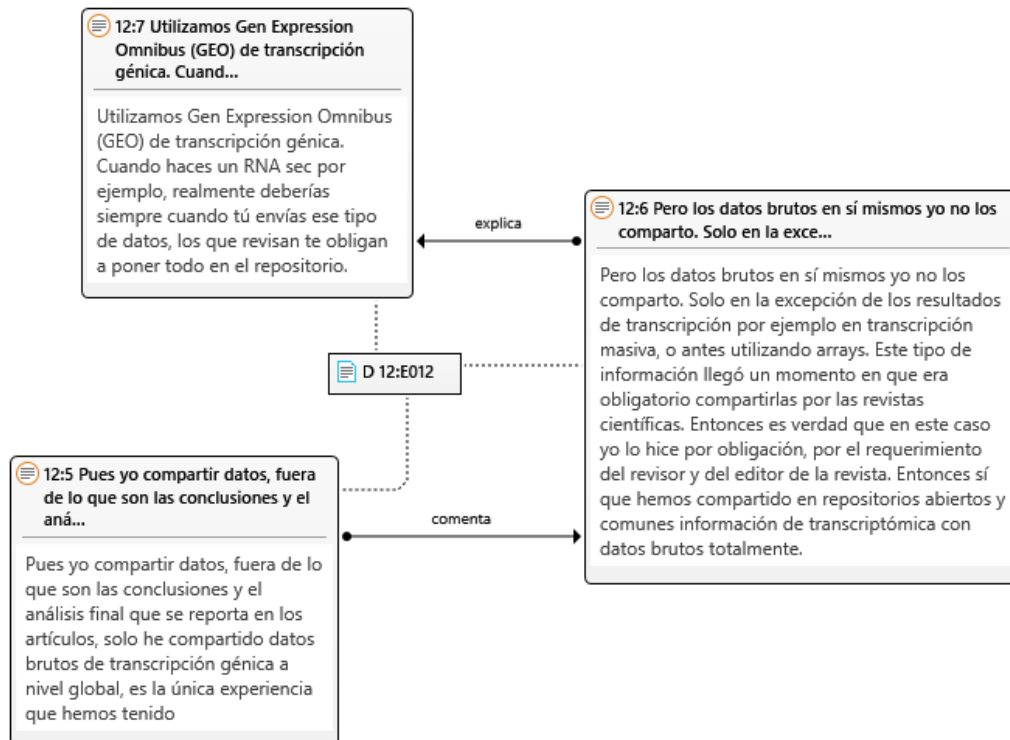


CONOCIMIENTOS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADO 012 (E012)

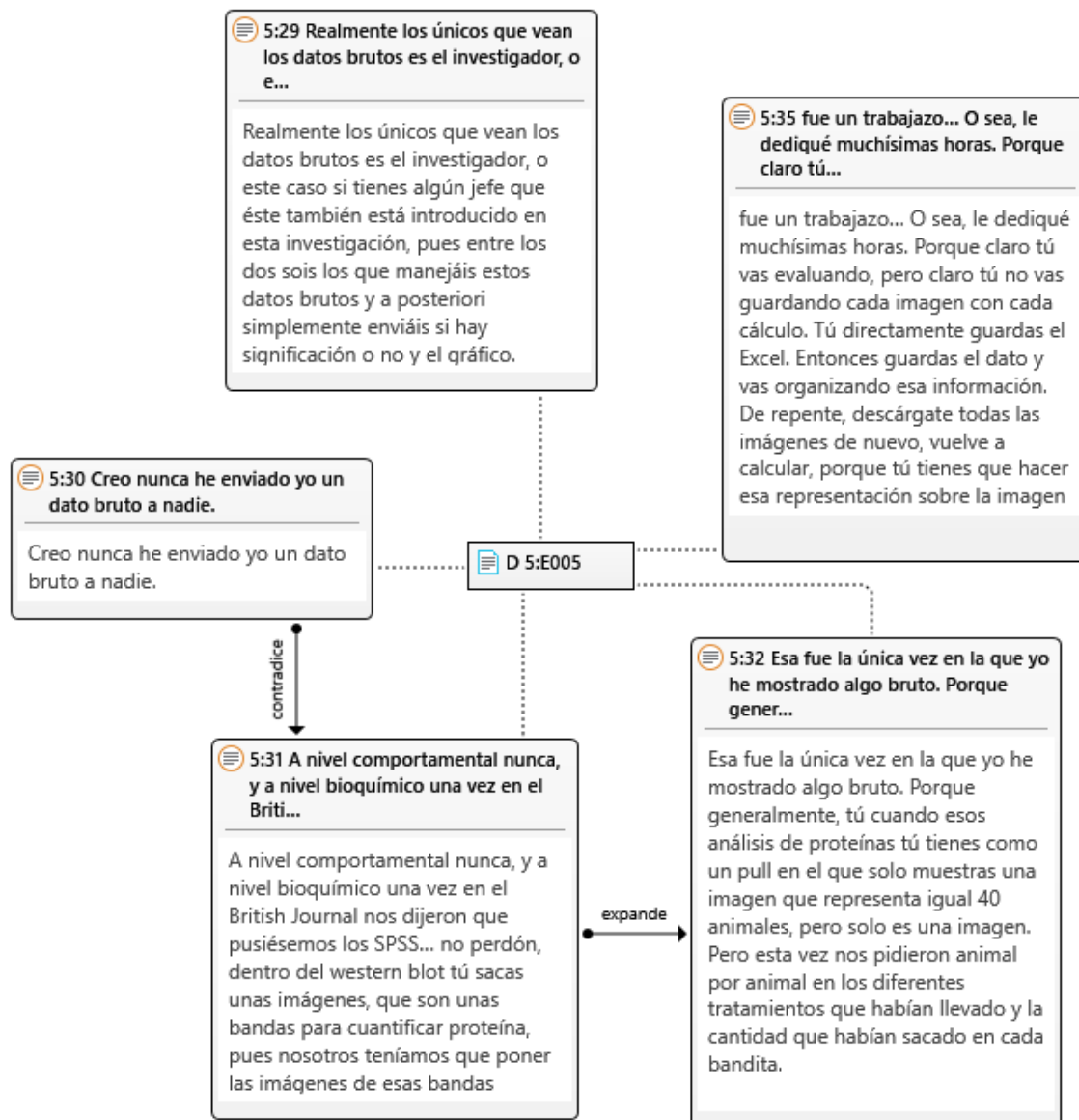


CONOCIMIENTOS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 005 (E005)

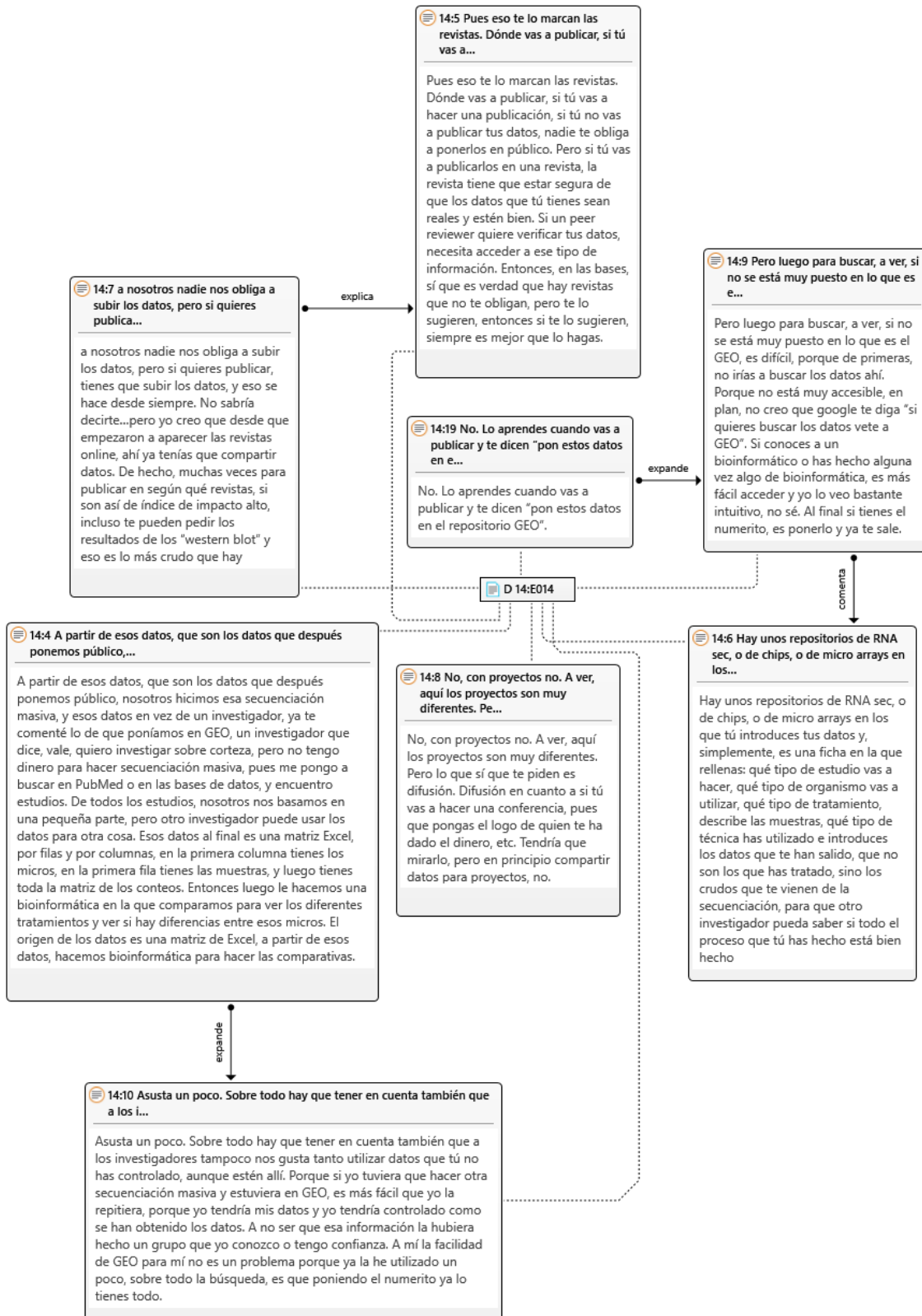


CONOCIMIENTOS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 014 (E014)

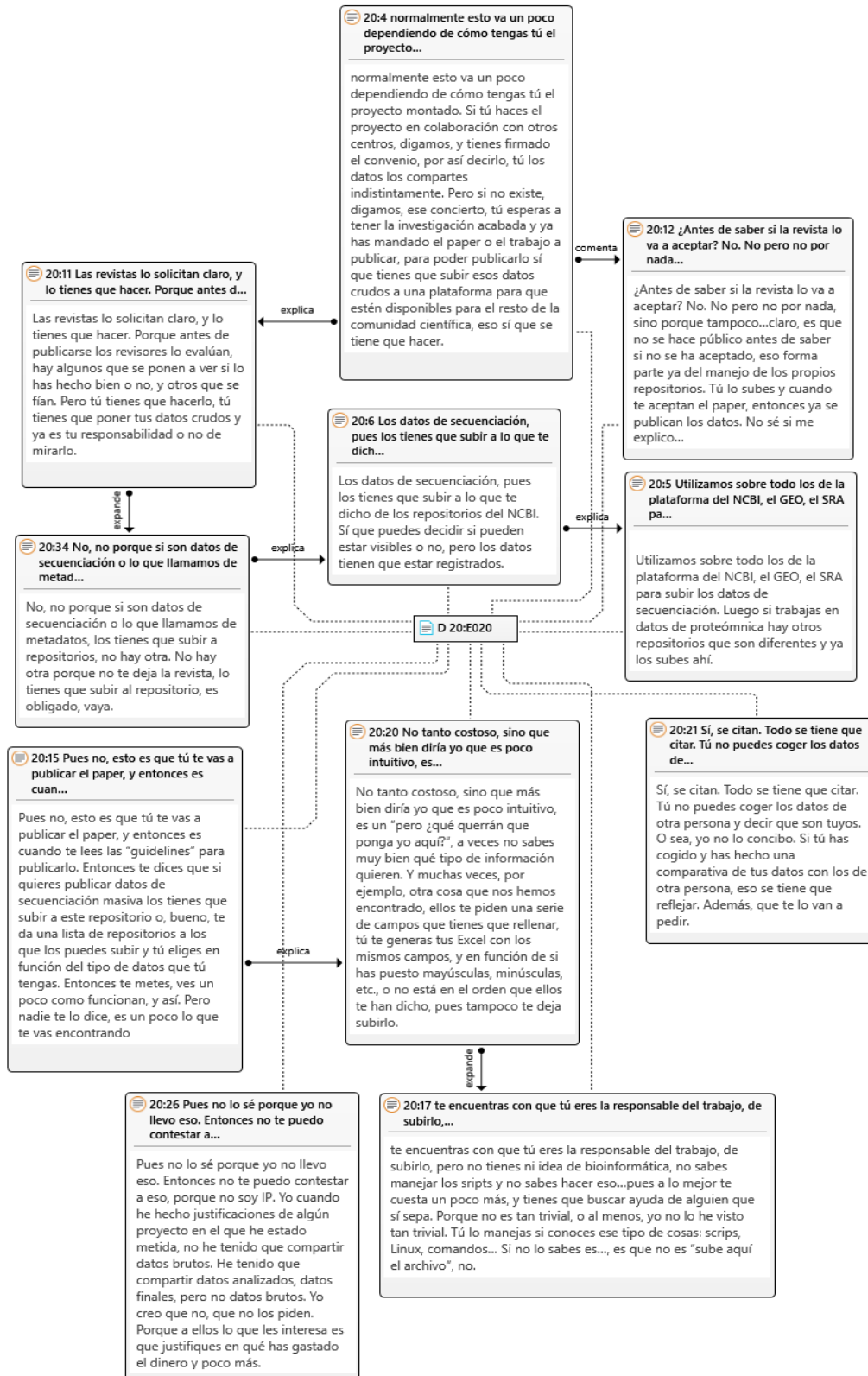


CONOCIMIENTOS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 020 (E020)

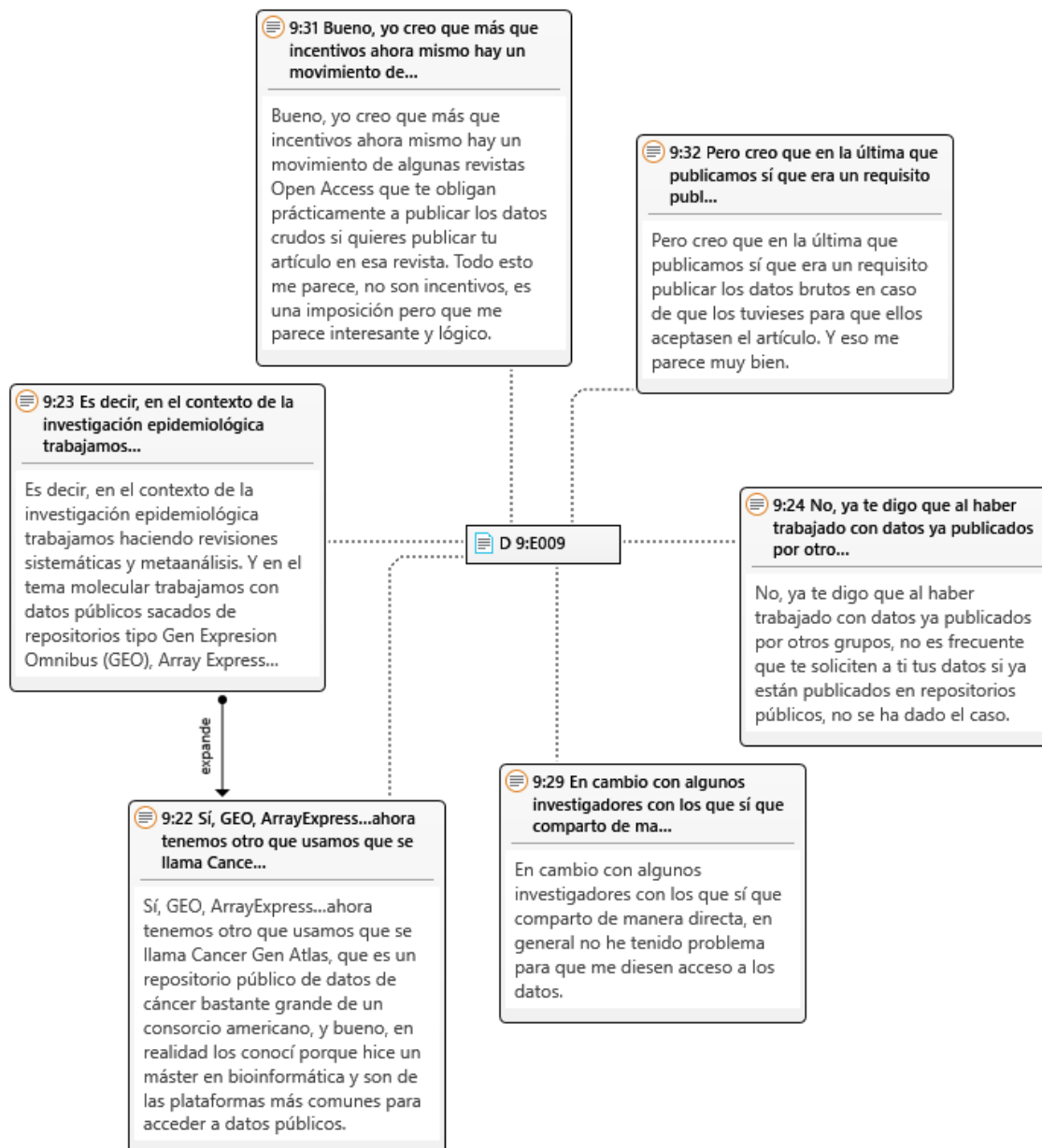


CONOCIMIENTOS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 009 (E009)

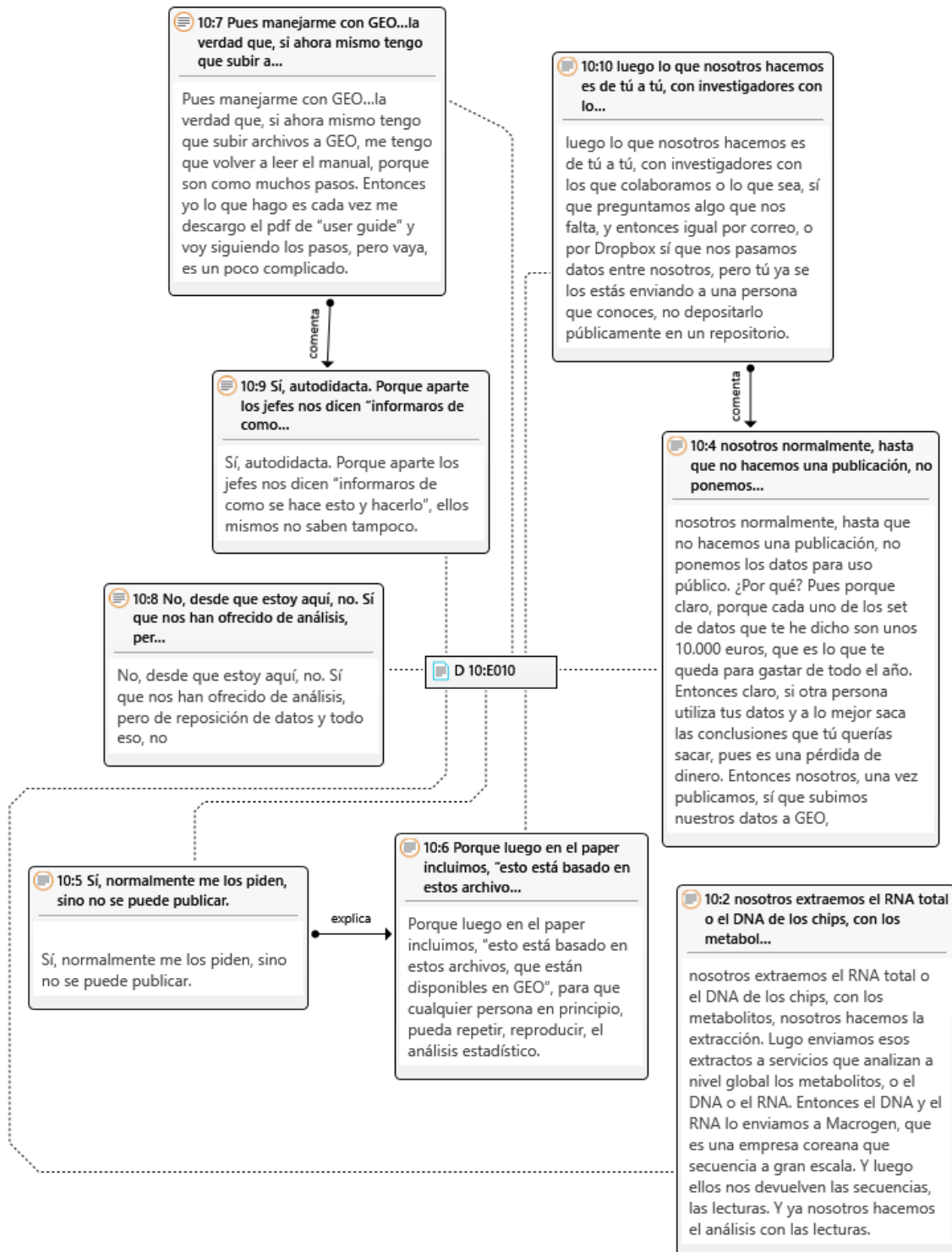


CONOCIMIENTOS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 010 (E010)

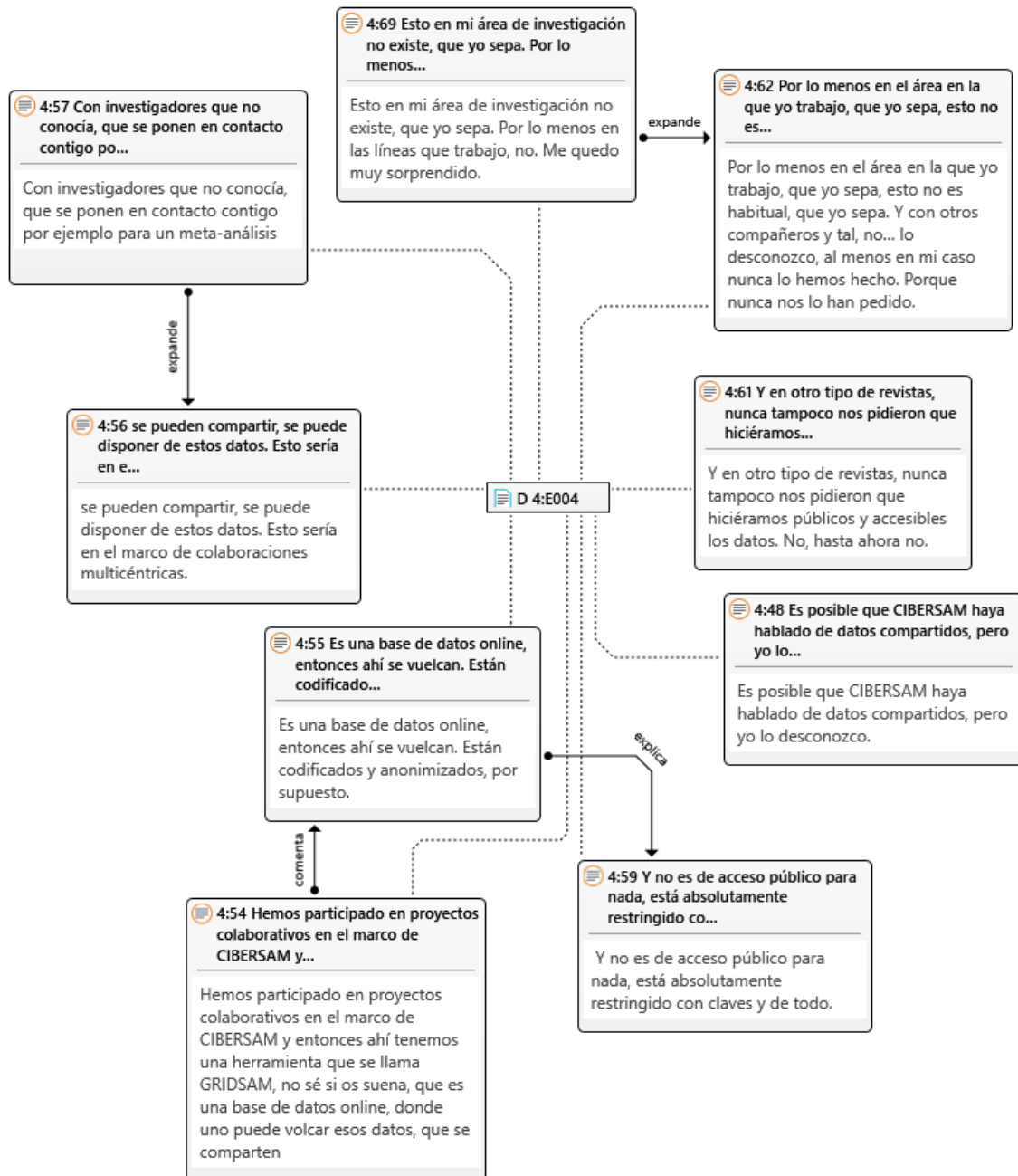


CONOCIMIENTOS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADO 004 (E004)

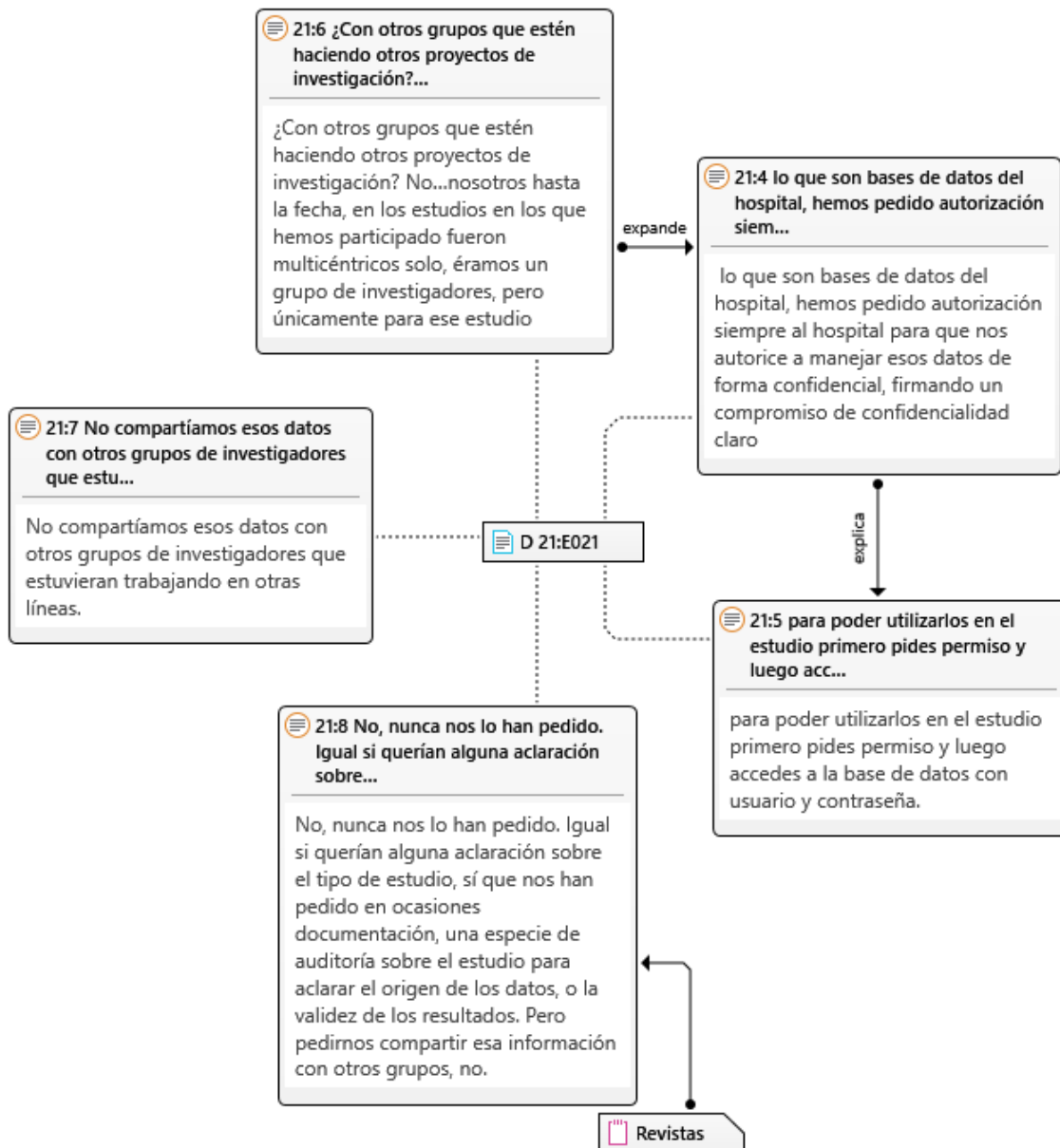


CONOCIMIENTOS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADO 021 (E021)

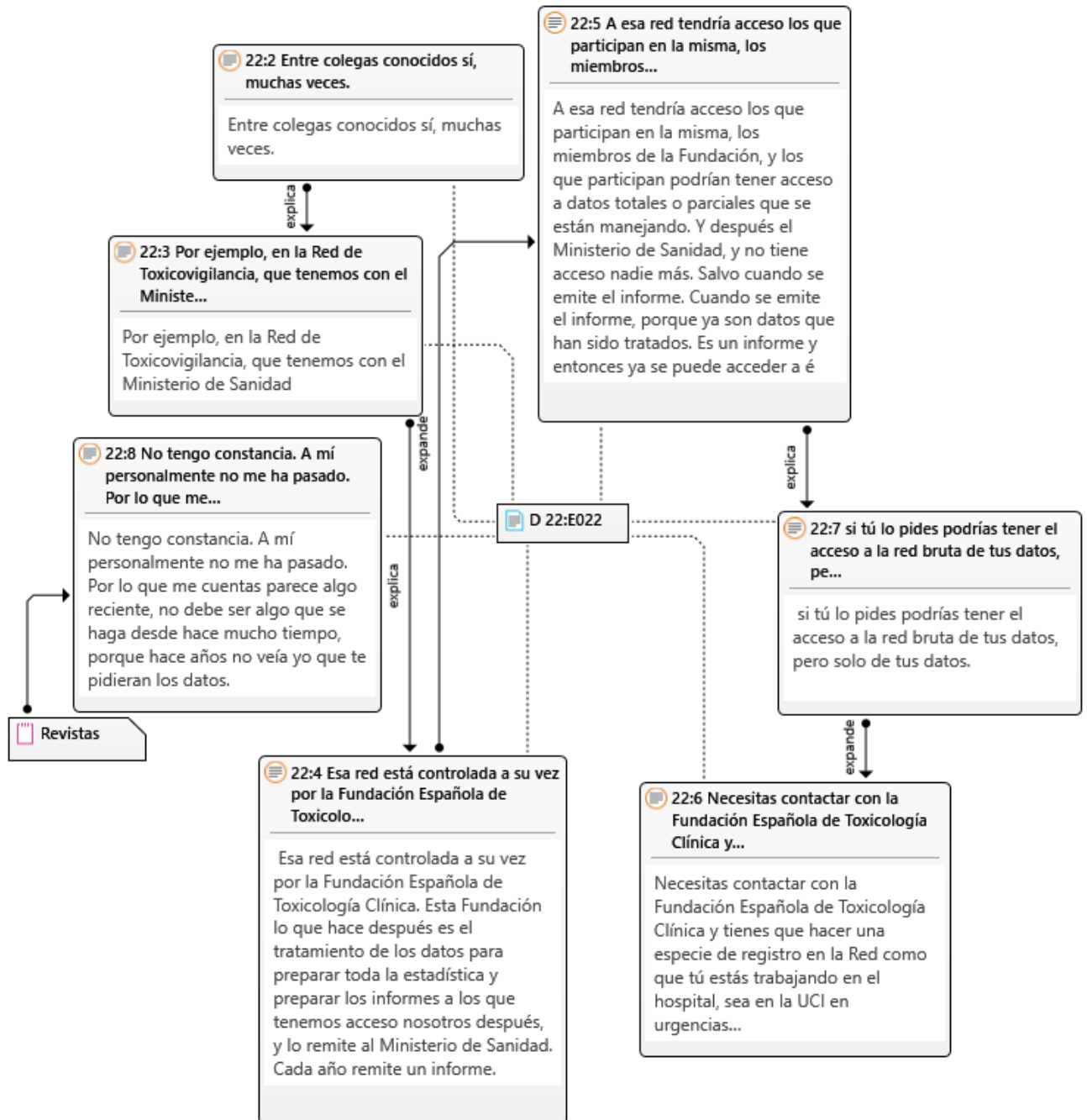


CONOCIMIENTOS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADO 022 (E022)

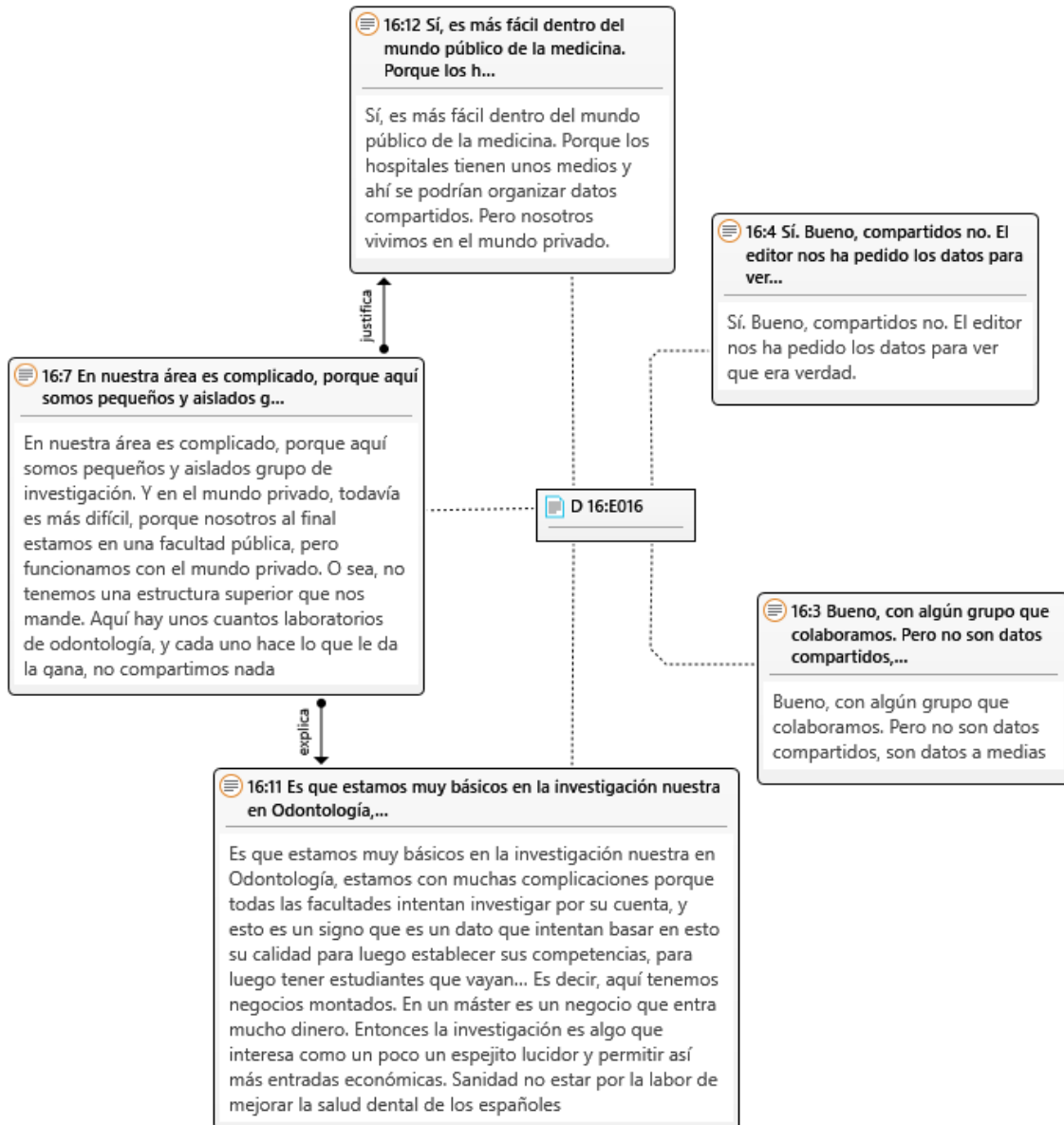


CONOCIMIENTOS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADO 016 (E016)

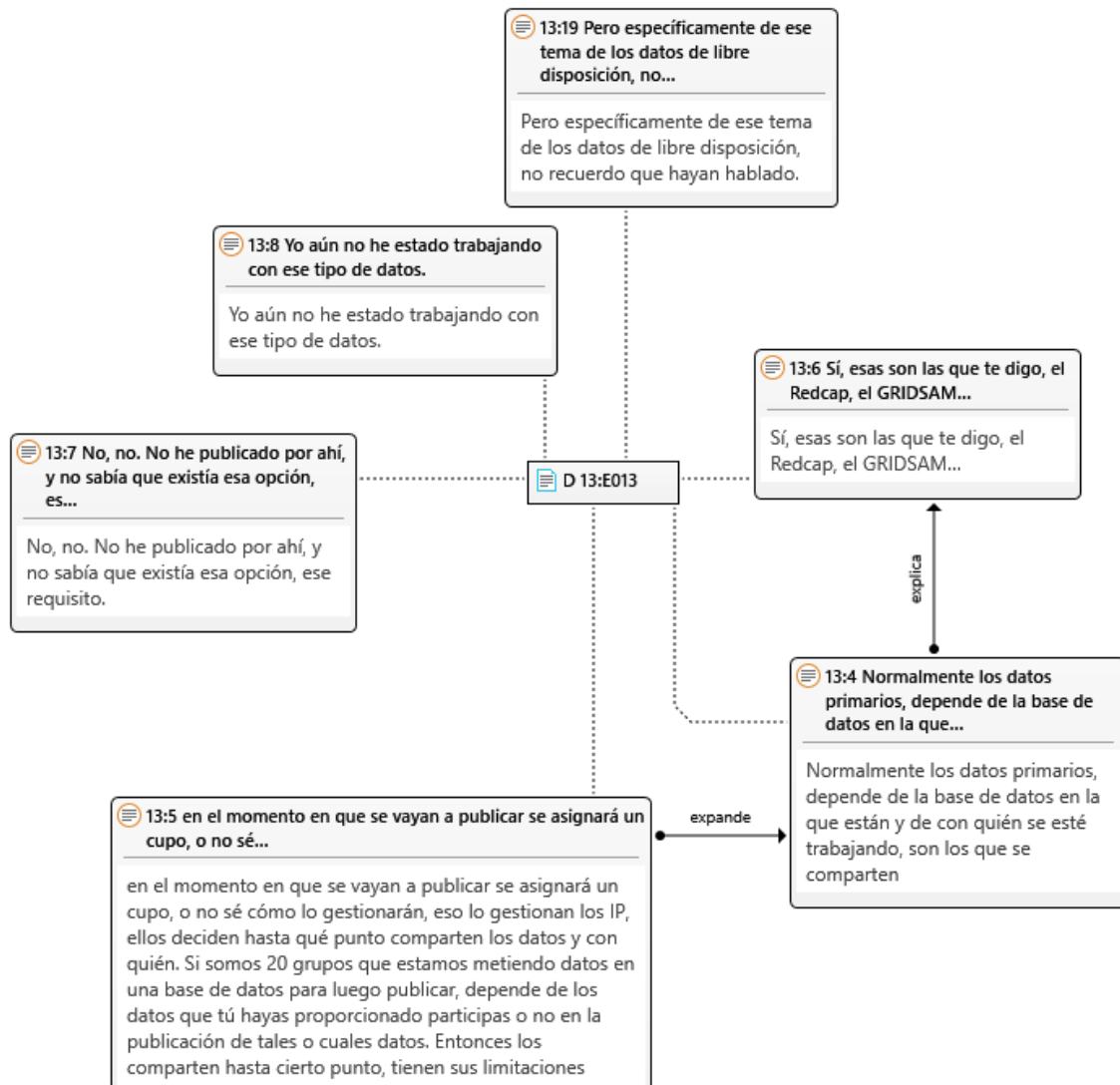


CONOCIMIENTOS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 013 (E013)

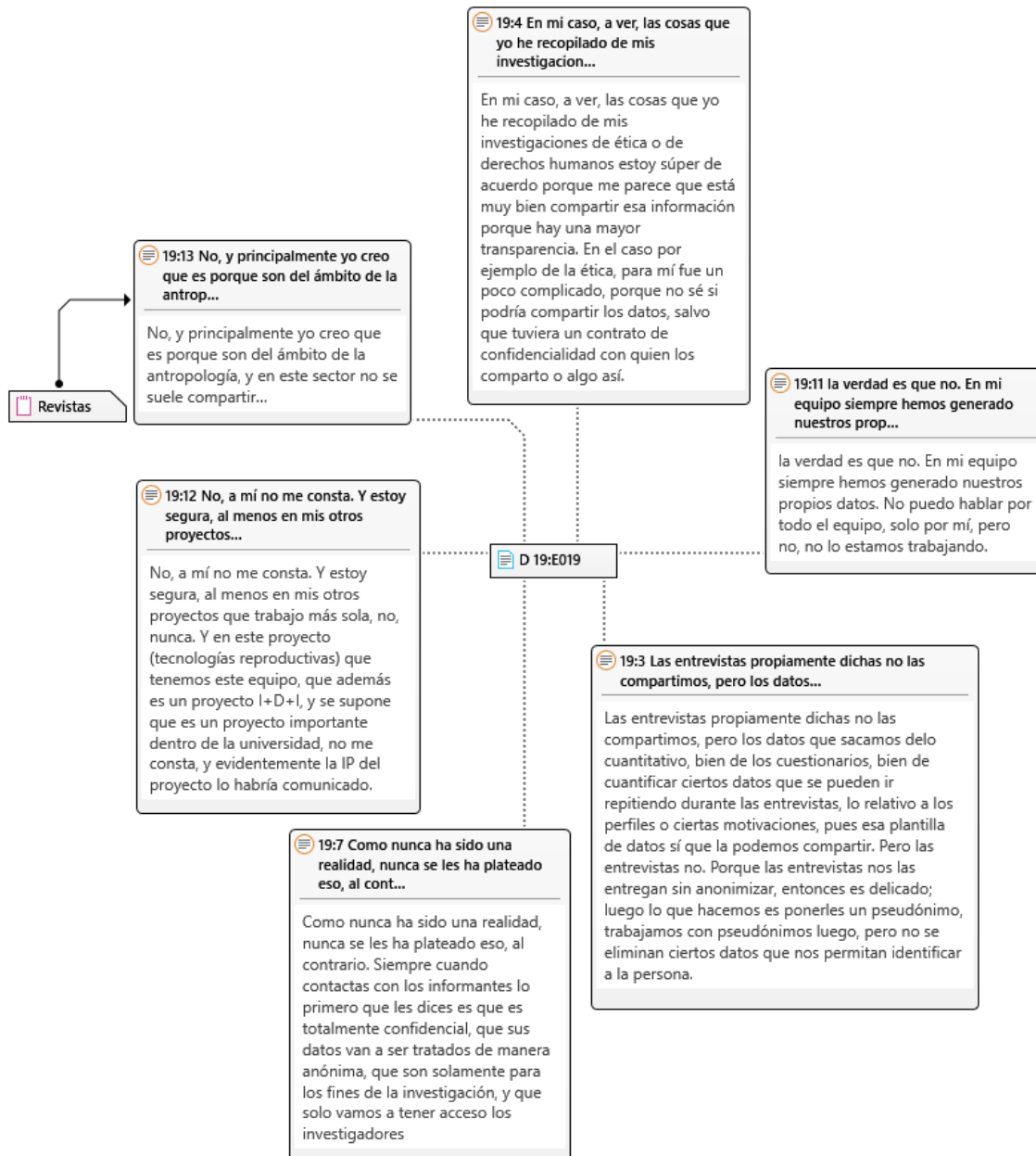


CONOCIMIENTOS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 019 (E019)

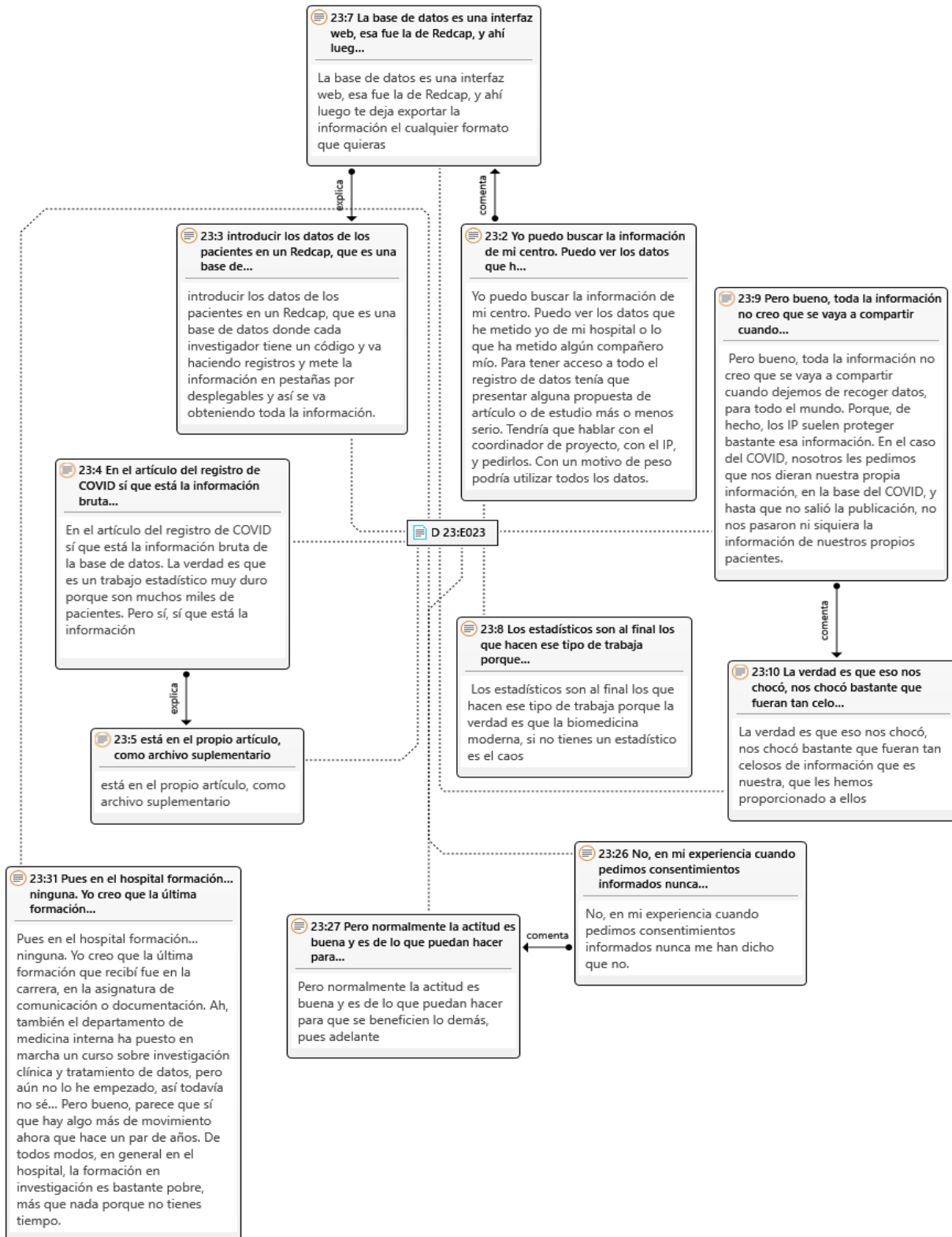


CONOCIMIENTOS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 023 (E023)

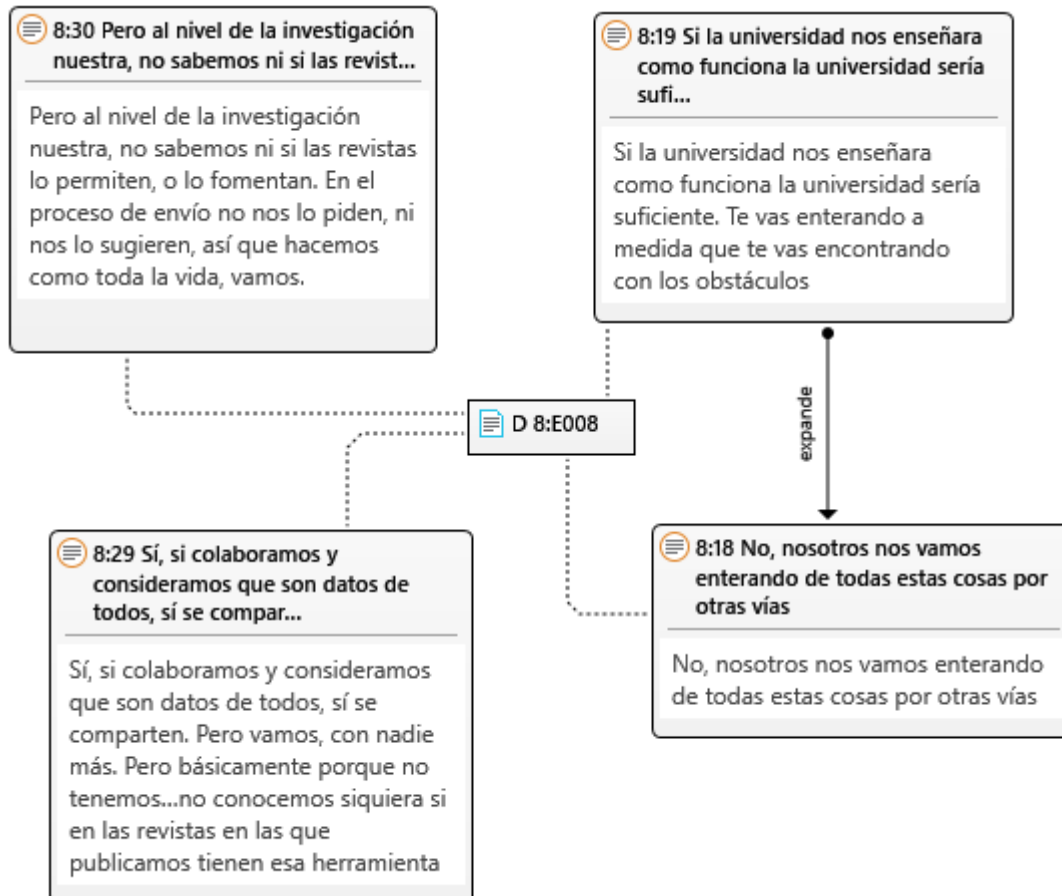


CONOCIMIENTOS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 008 (E008)

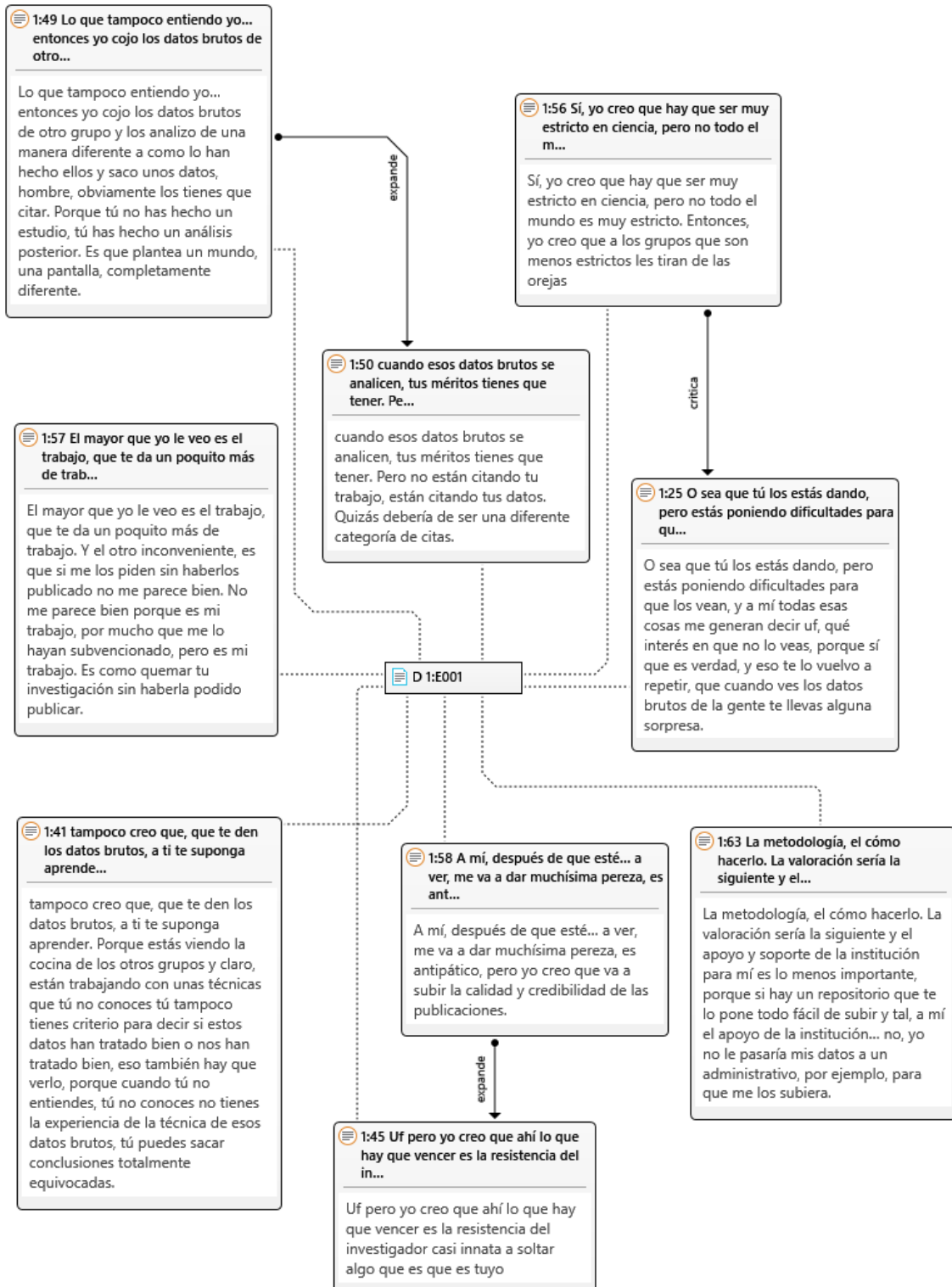


CREENCIAS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 001 (E001)

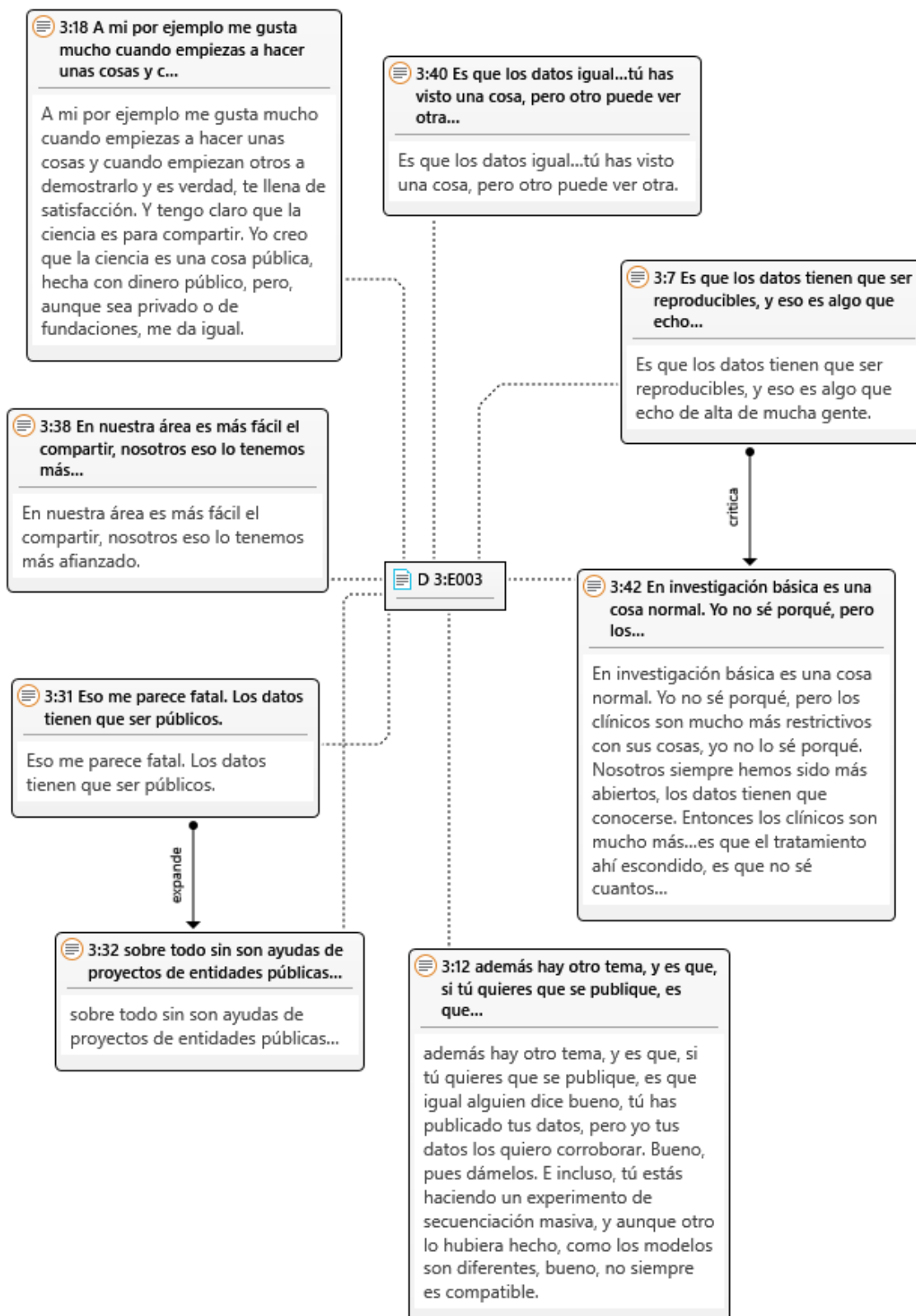


CREENCIAS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 003 (E003)

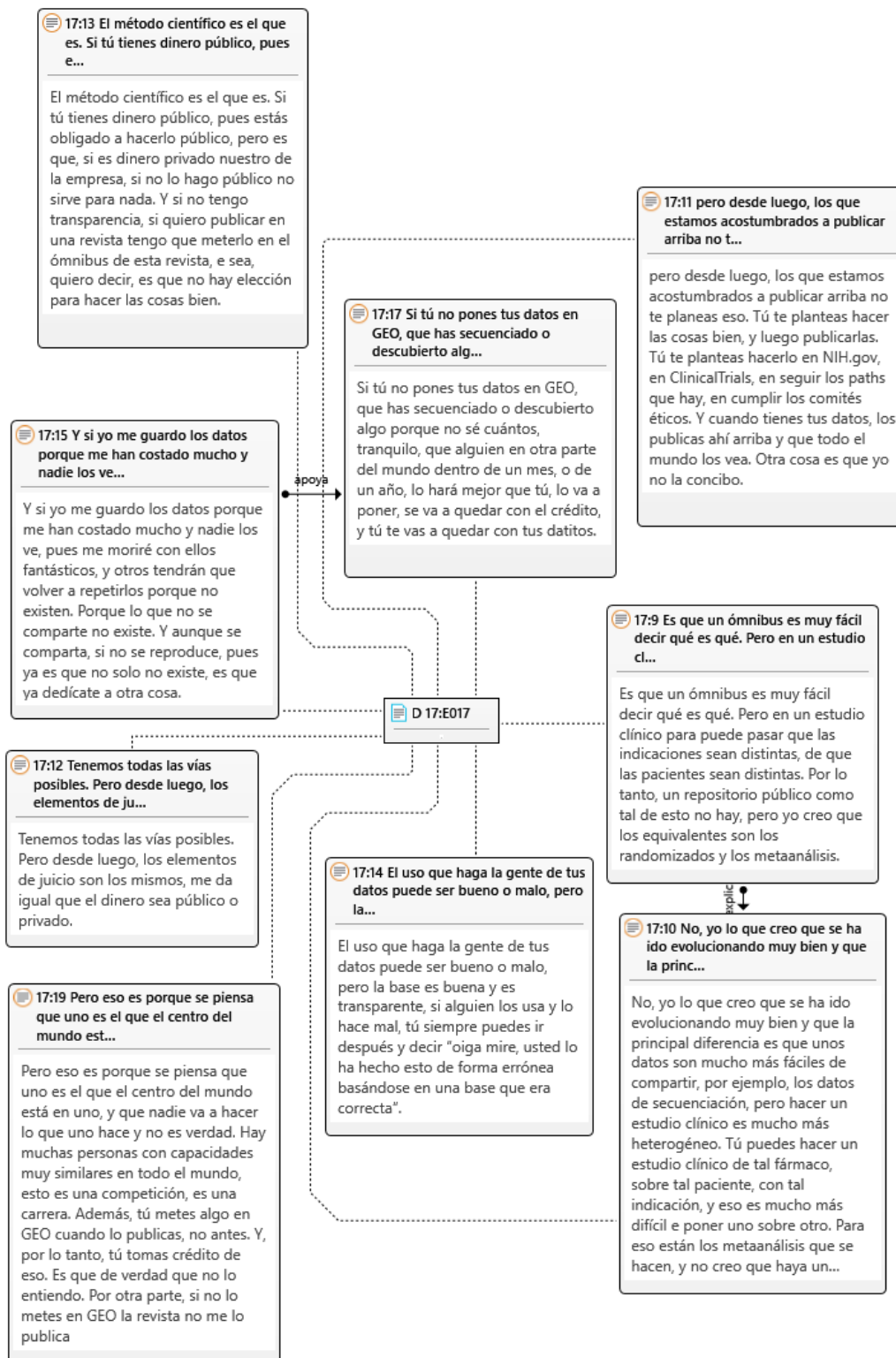


CREENCIAS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 017 (E017)

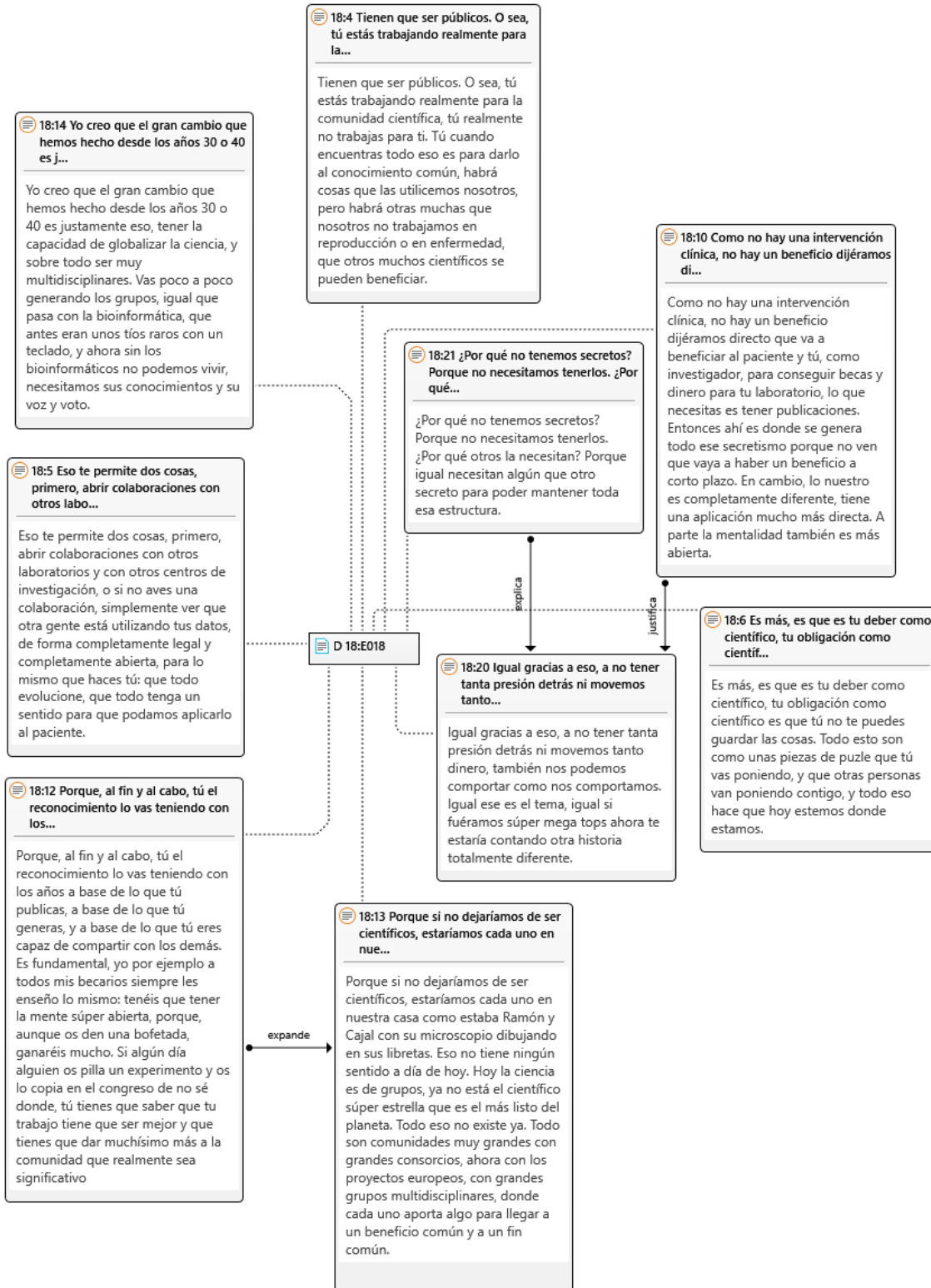


CREENCIAS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 018 (E018)

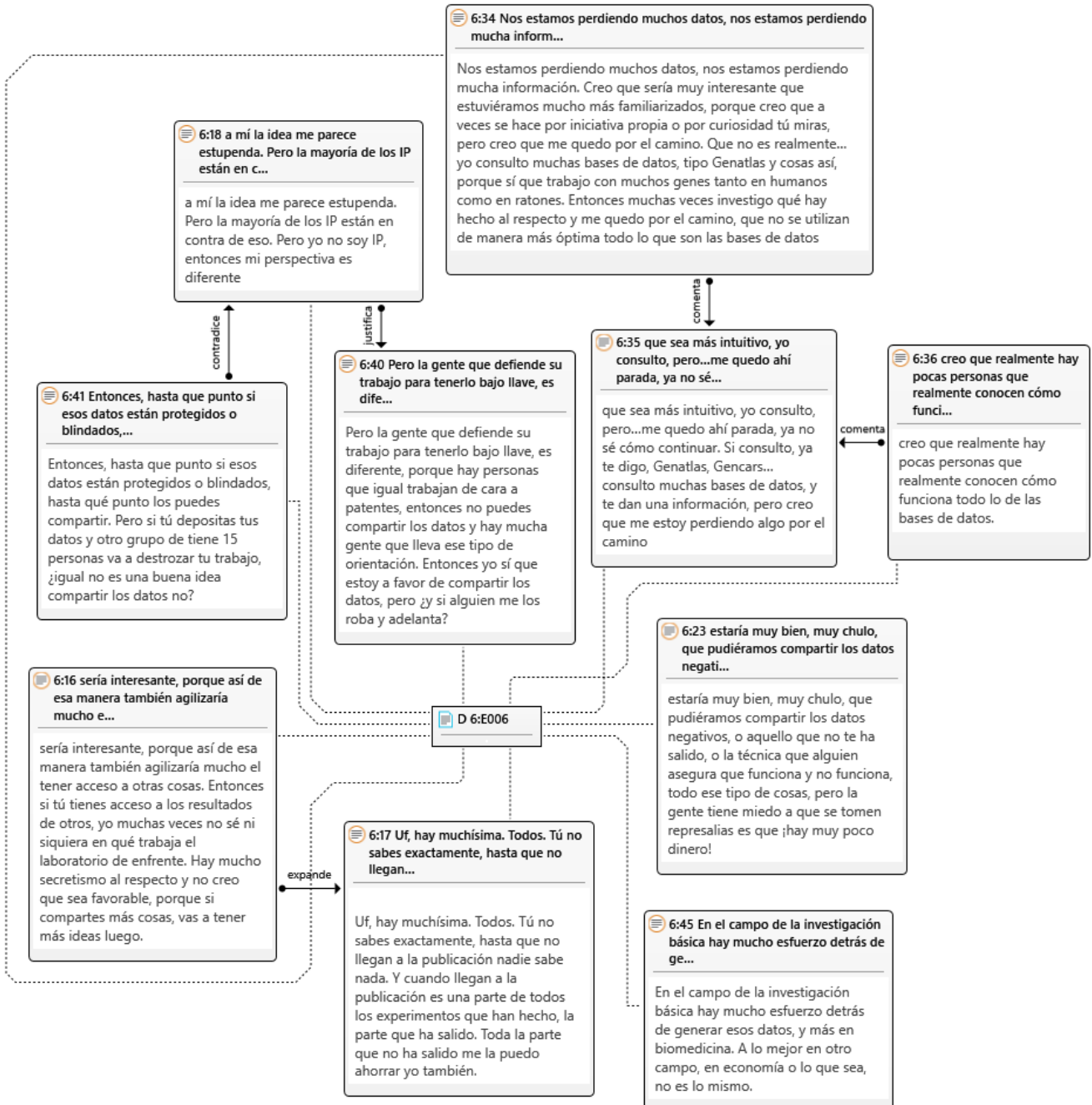


CREENCIAS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 006 (E006)

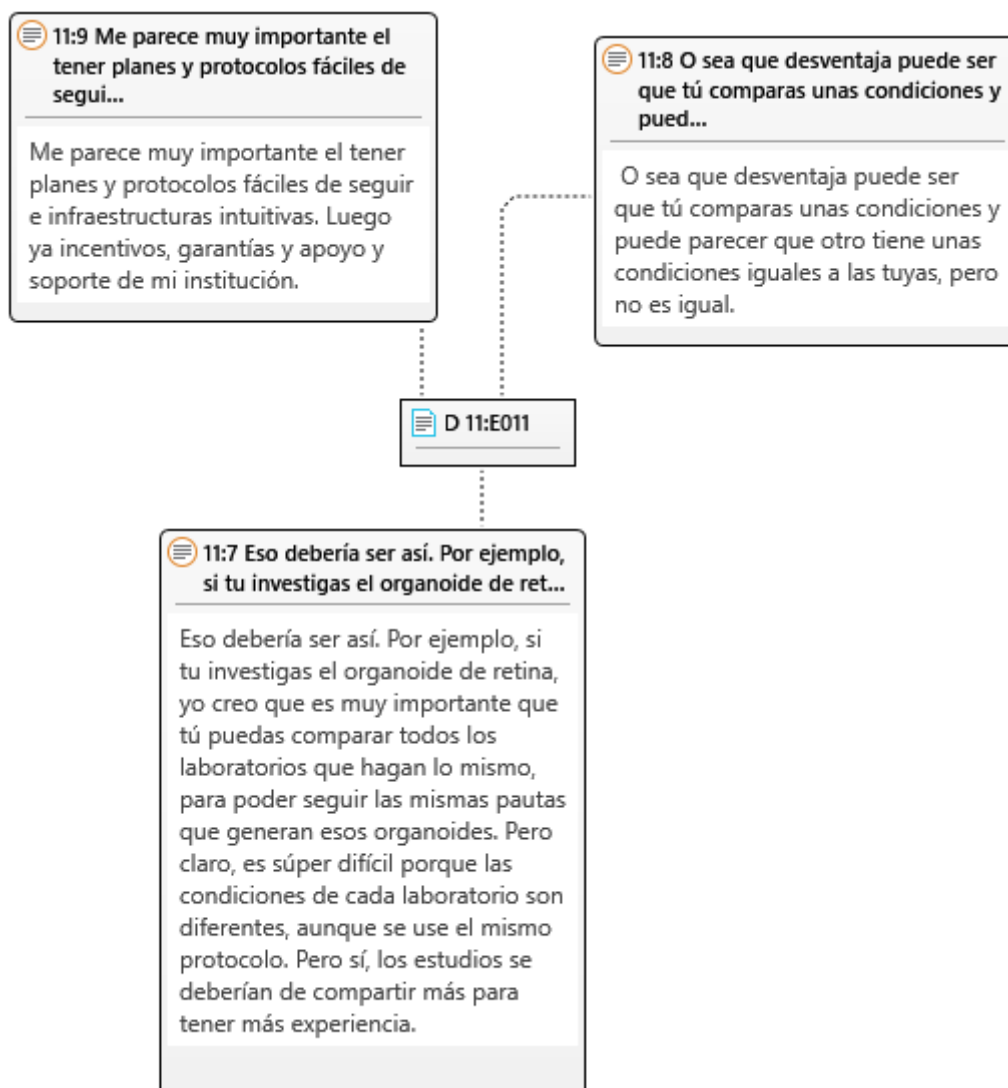


CREENCIAS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 011 (E011)

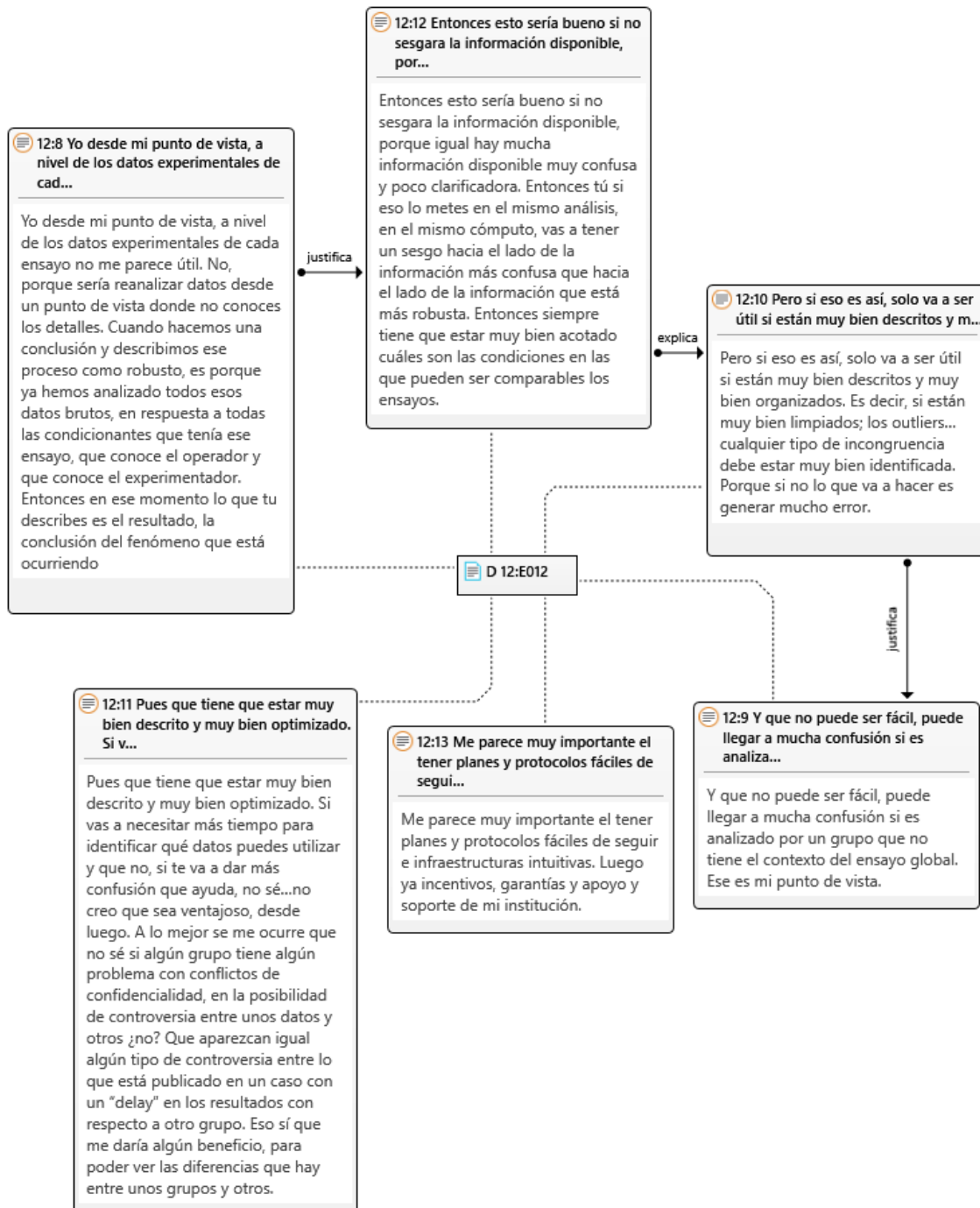


CREENCIAS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 012 (E012)

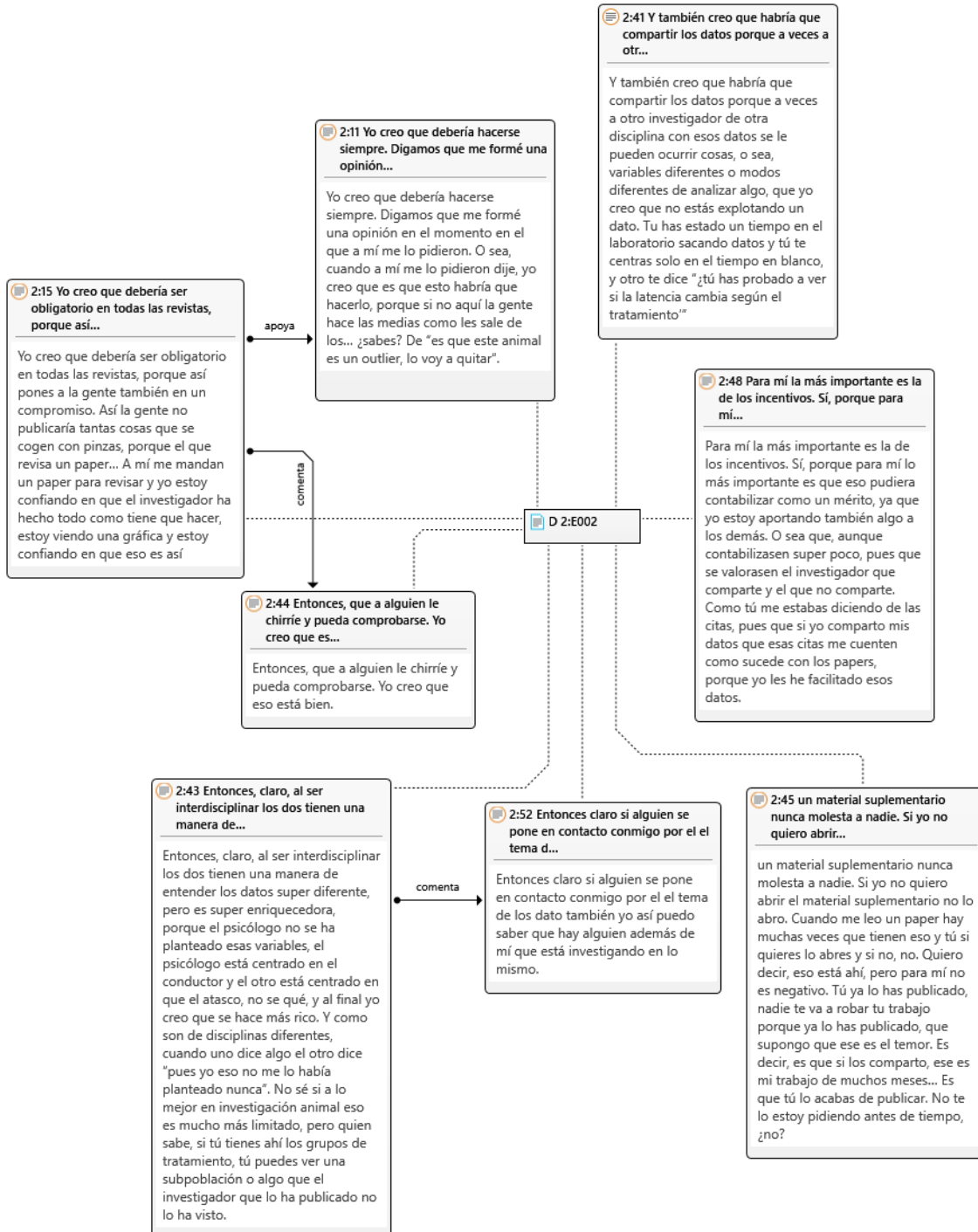


CREENCIAS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADA 002 (E002)

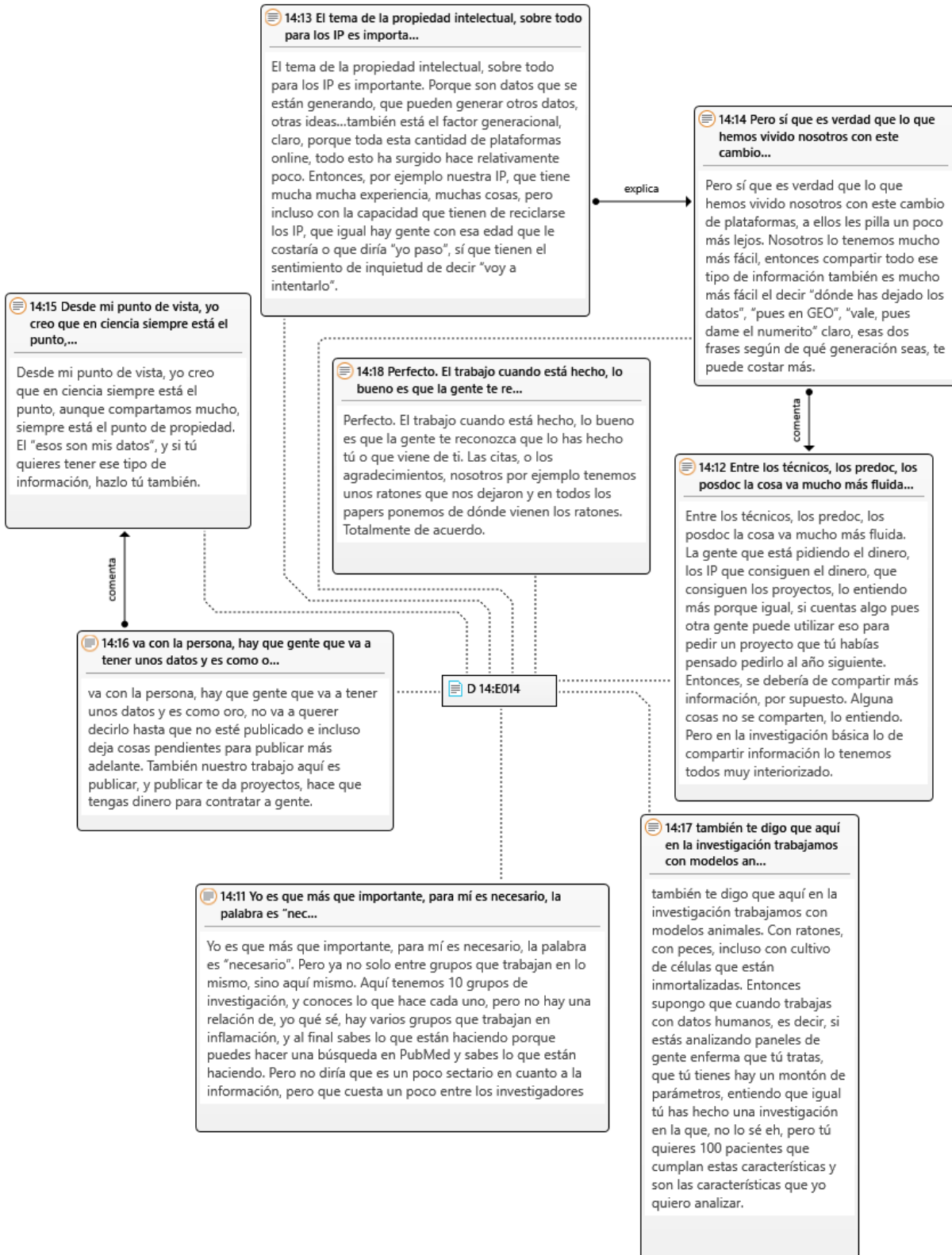


CREENCIAS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADA 014 (E014)

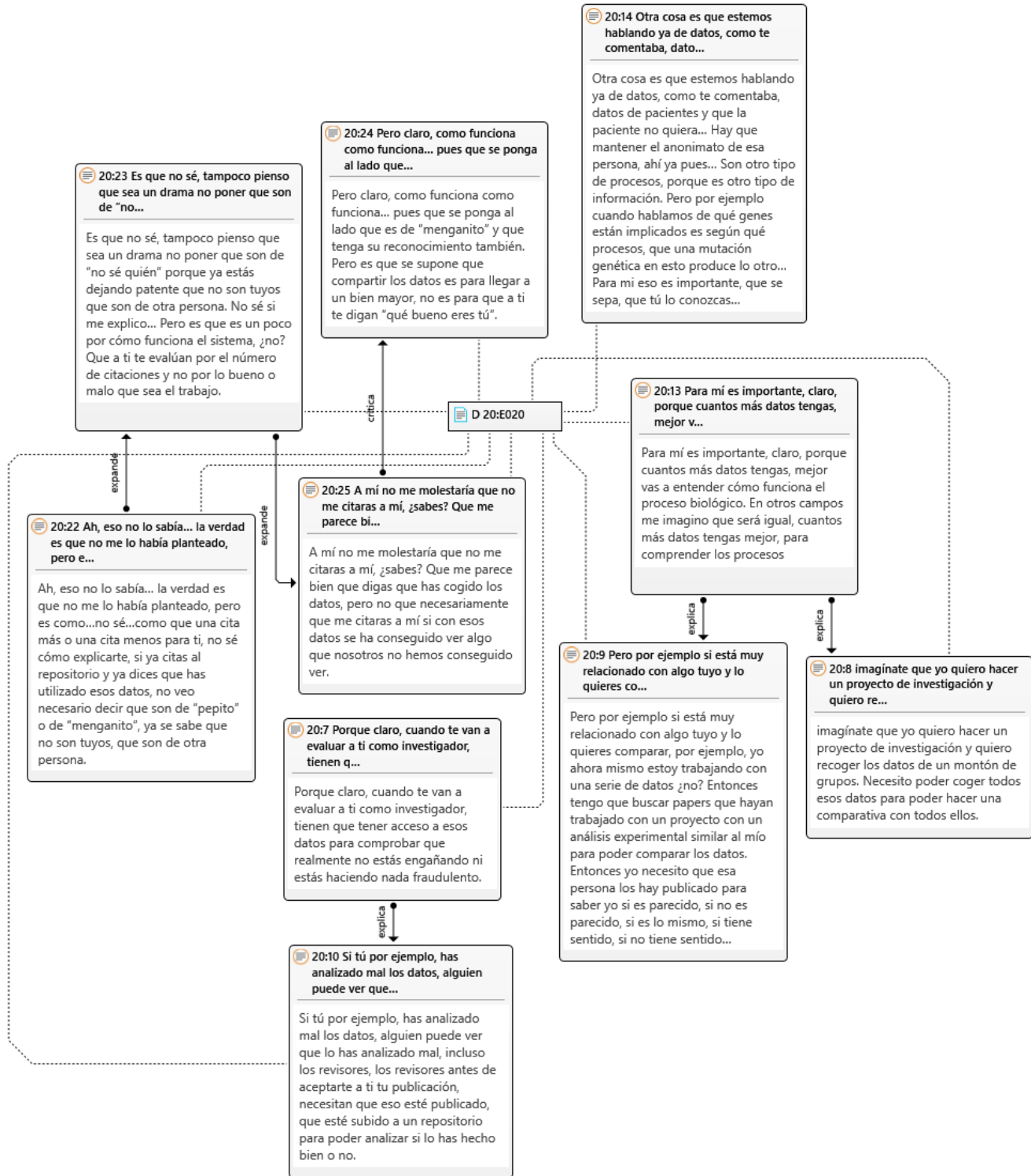


CREENCIAS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADA 020 (E020)

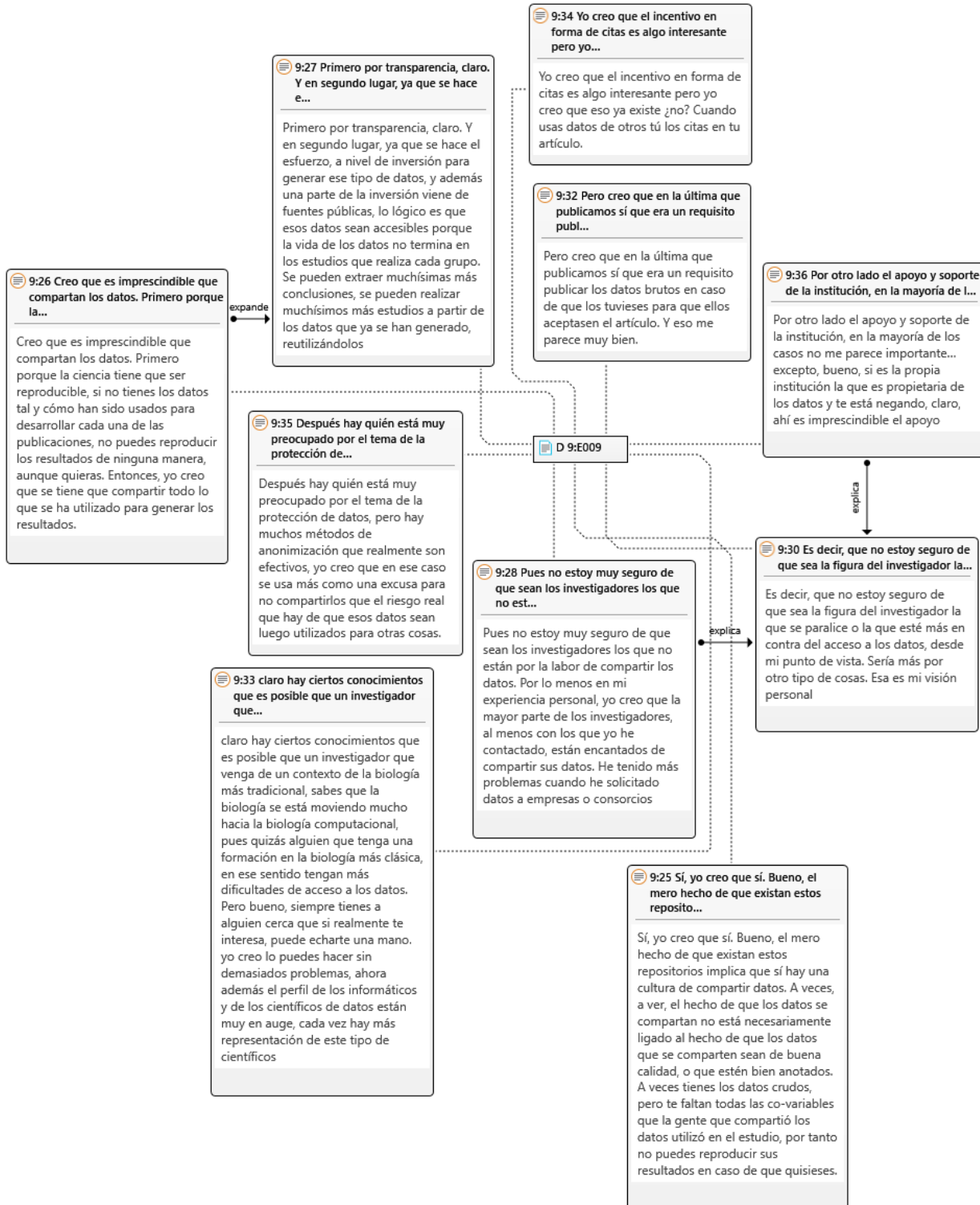


CREENCIAS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADA 009 (E009)

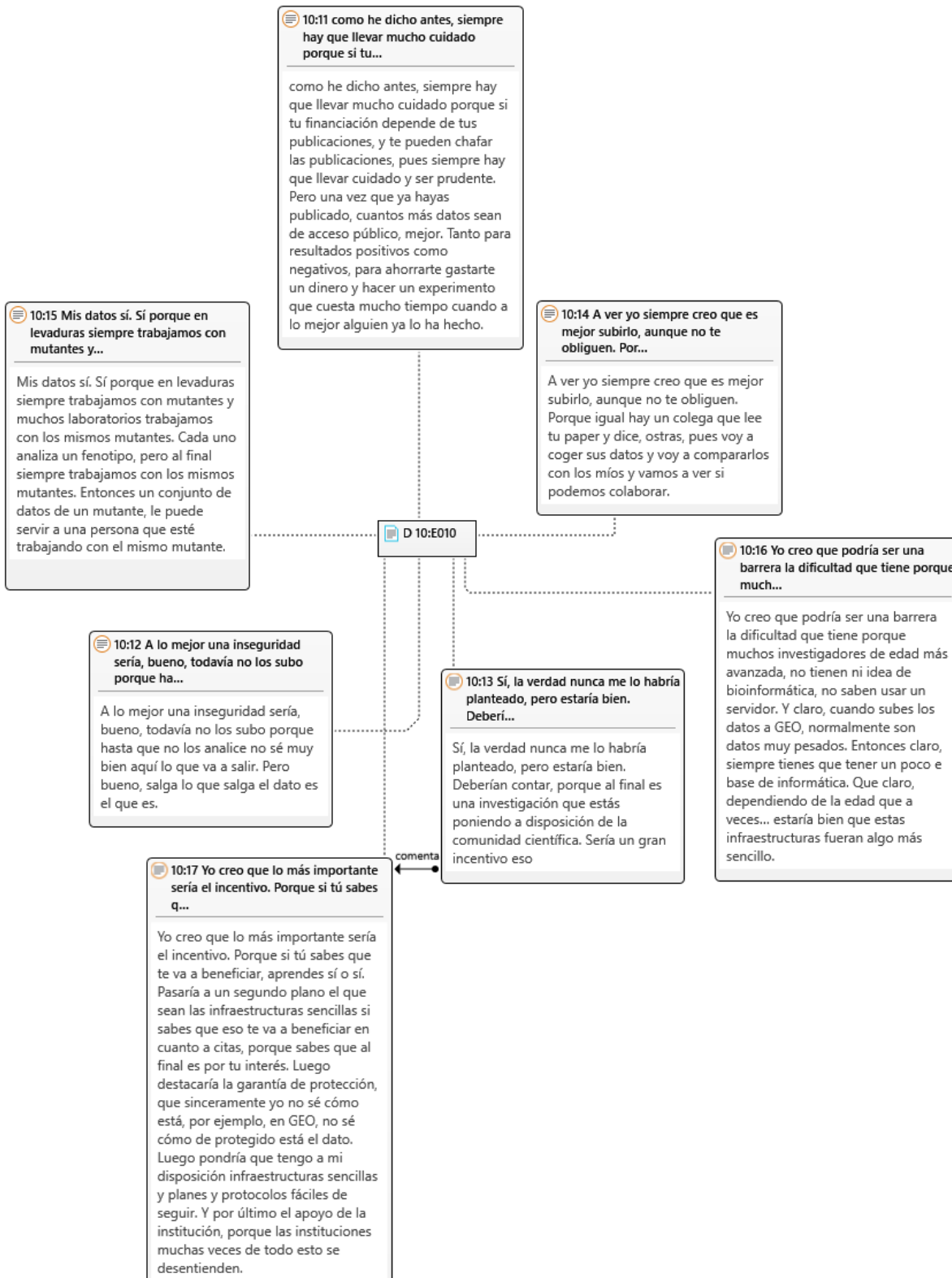


CREENCIAS

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADA 010 (E010)

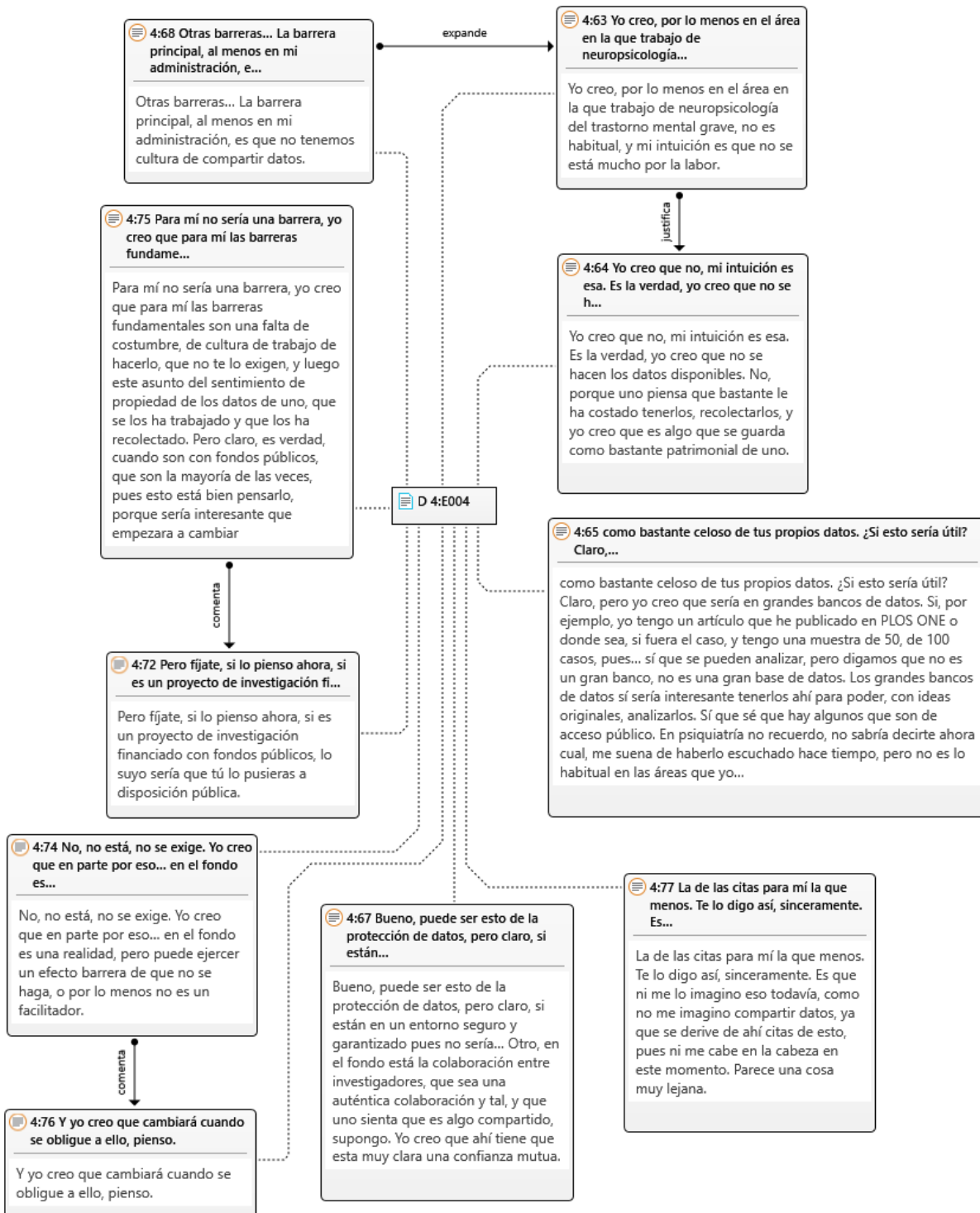


CREENCIAS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 004 (E004)

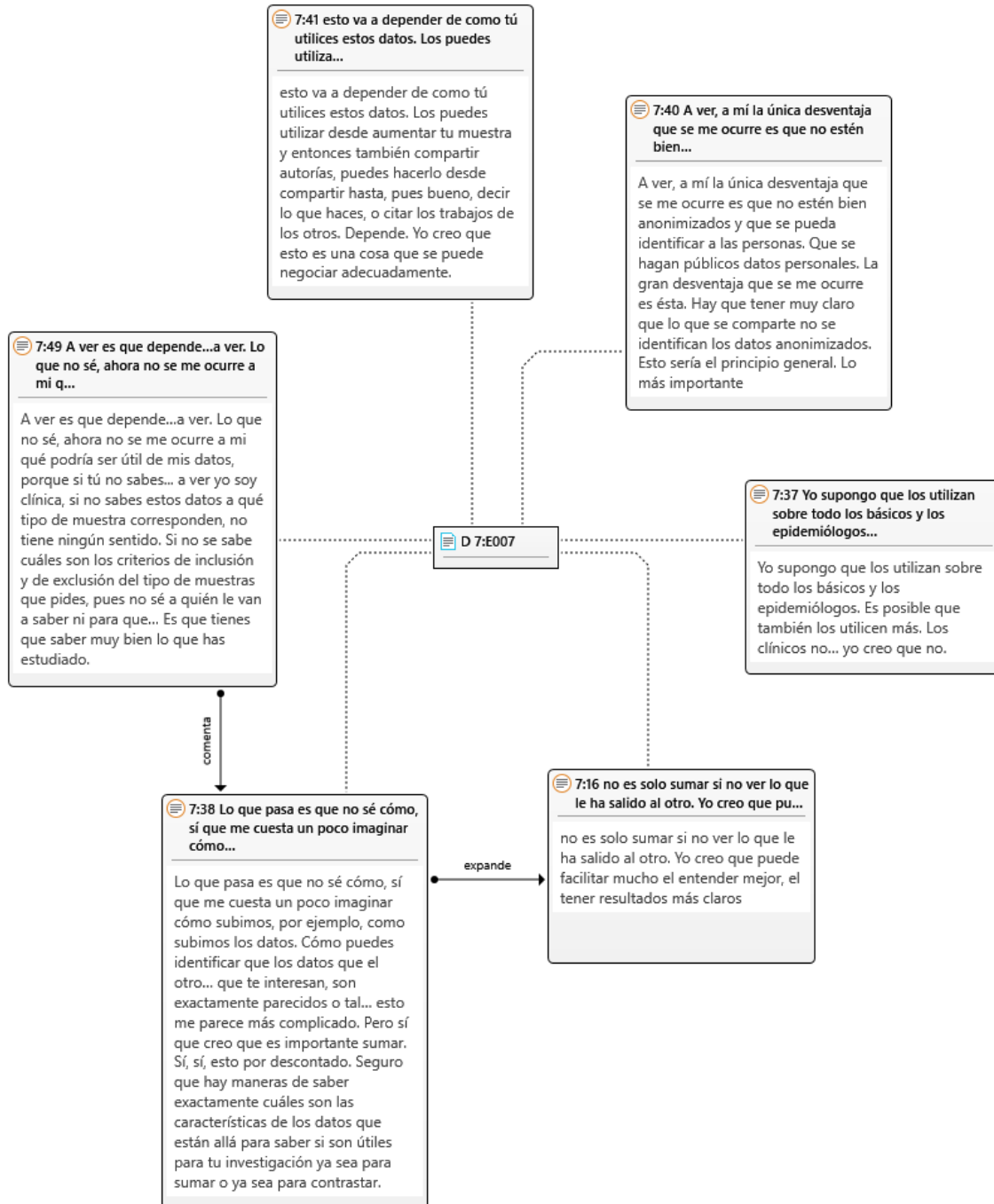


CREENCIAS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 007 (E007)

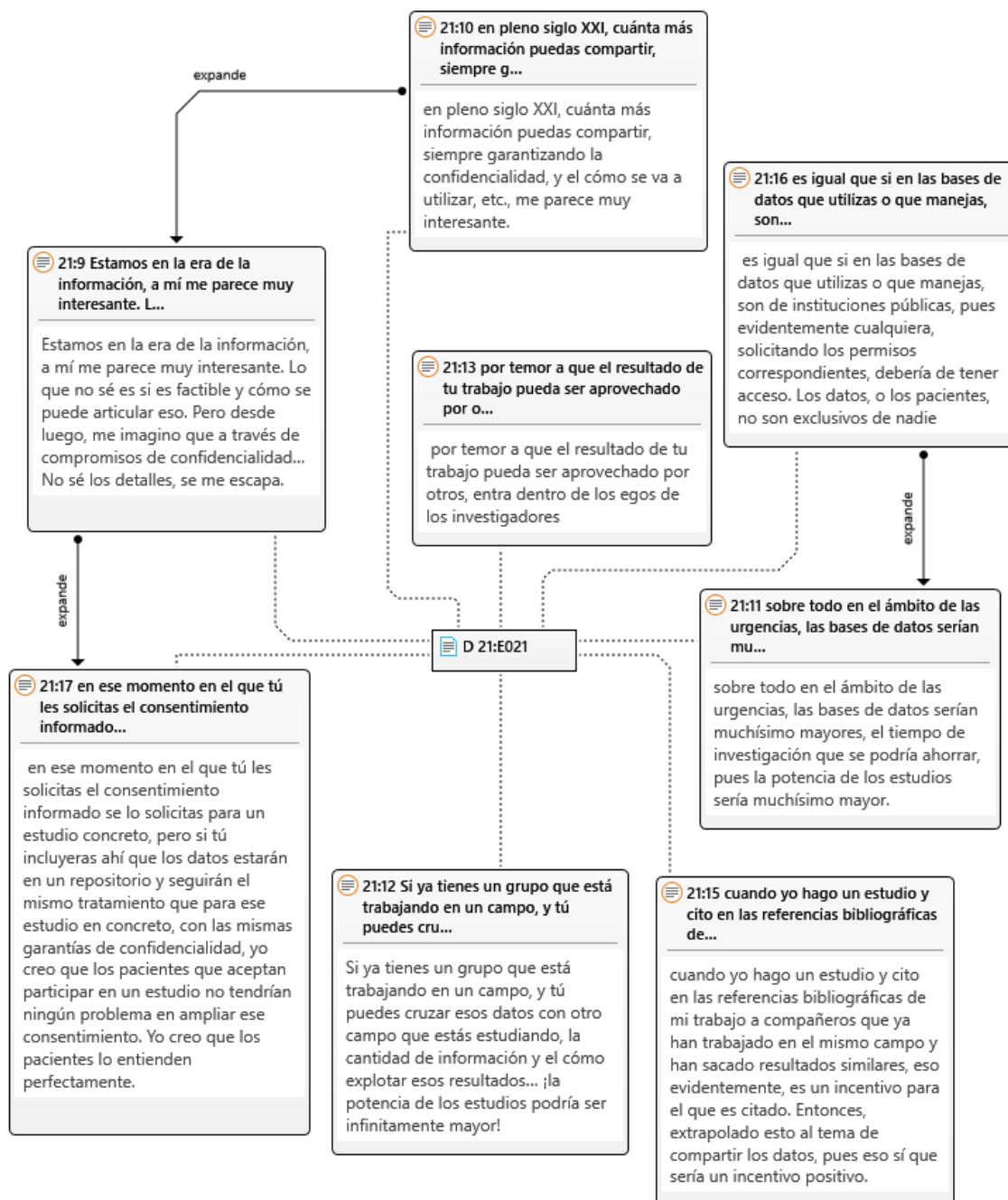


CREENCIAS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 021 (E021)

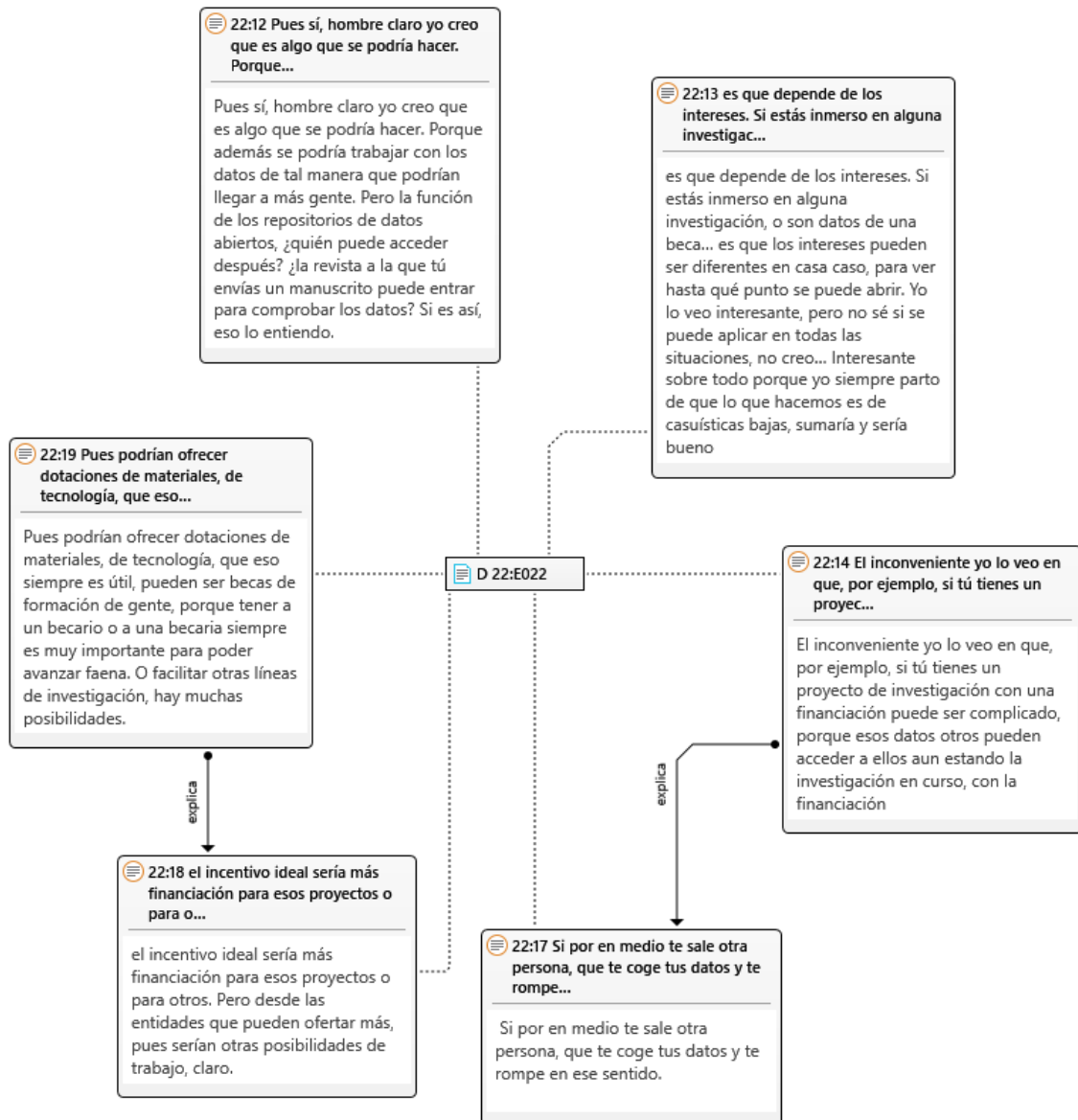


CREENCIAS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 022 (E022)

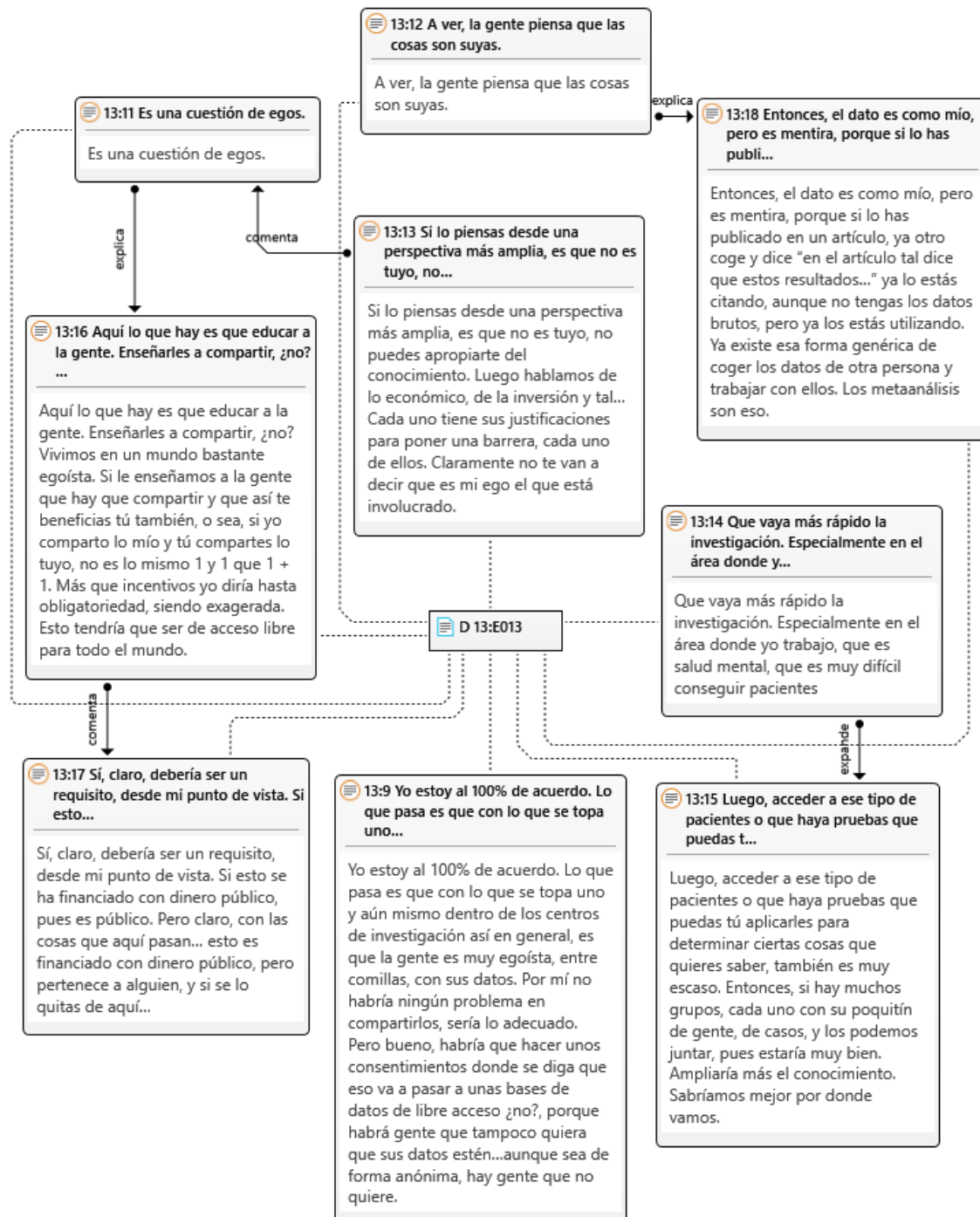


CREENCIAS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADA 004 (E004)

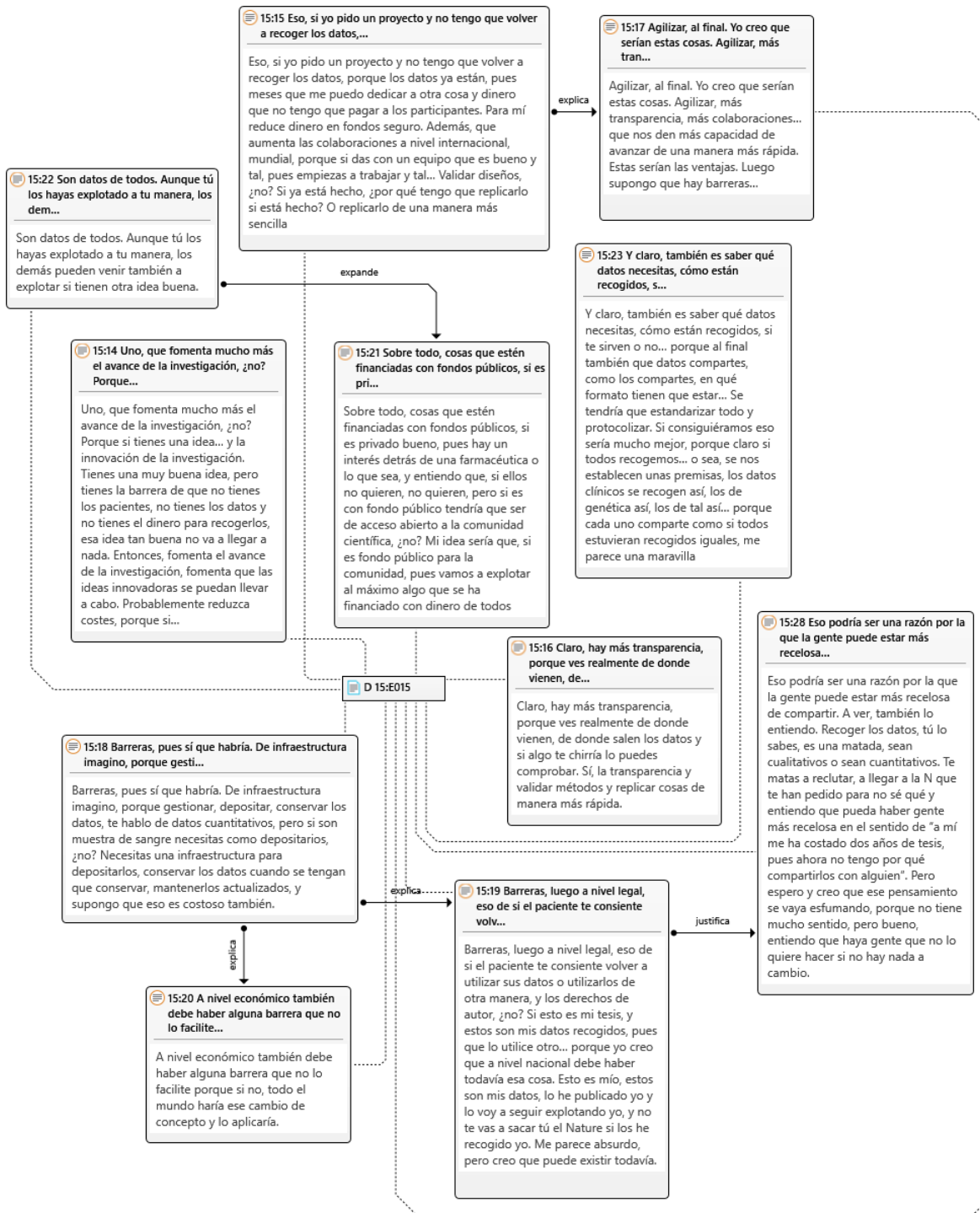


CREENCIAS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

ENTREVISTADA 015 (E015)

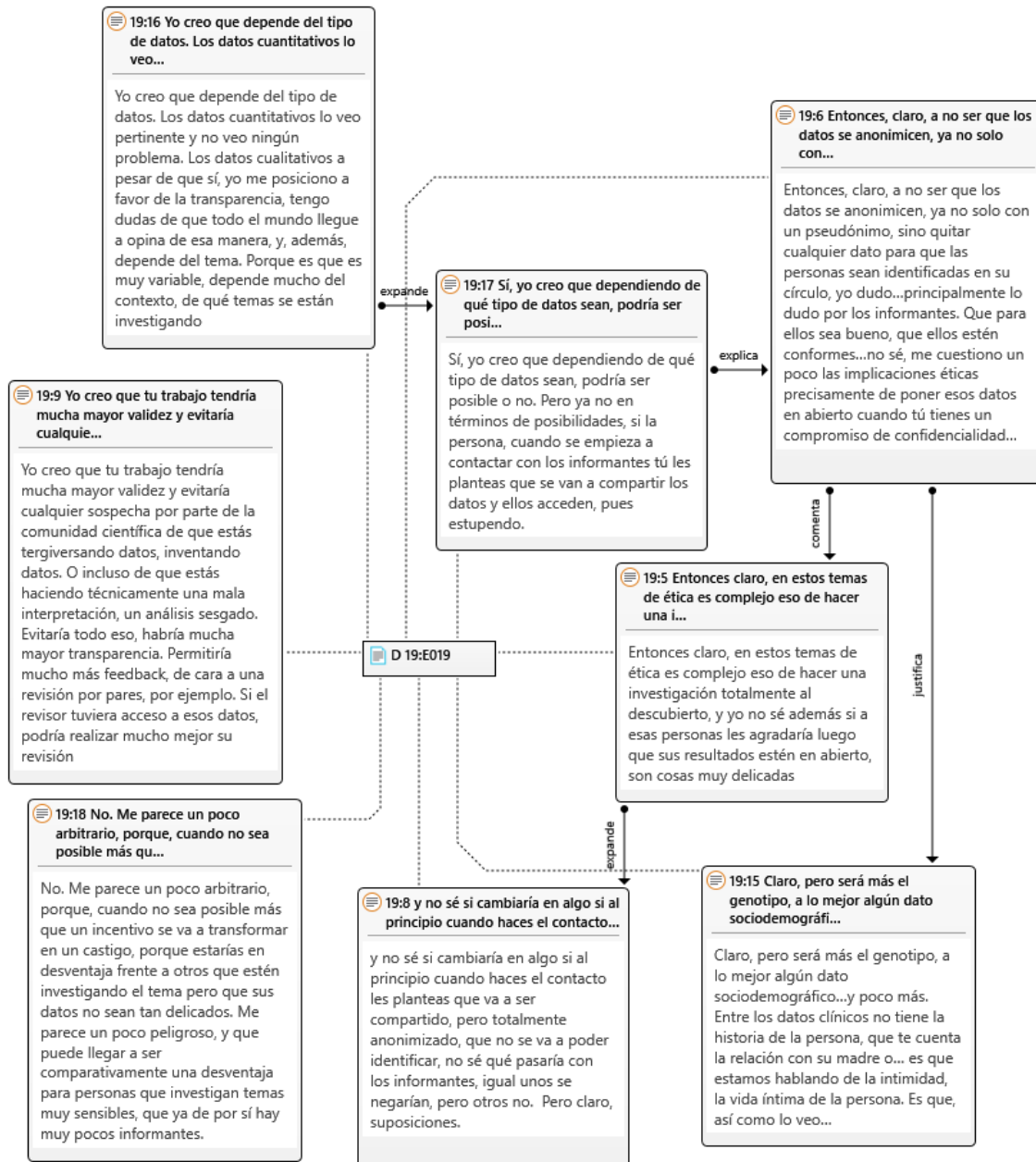


CREENCIAS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADA 019 (E019)

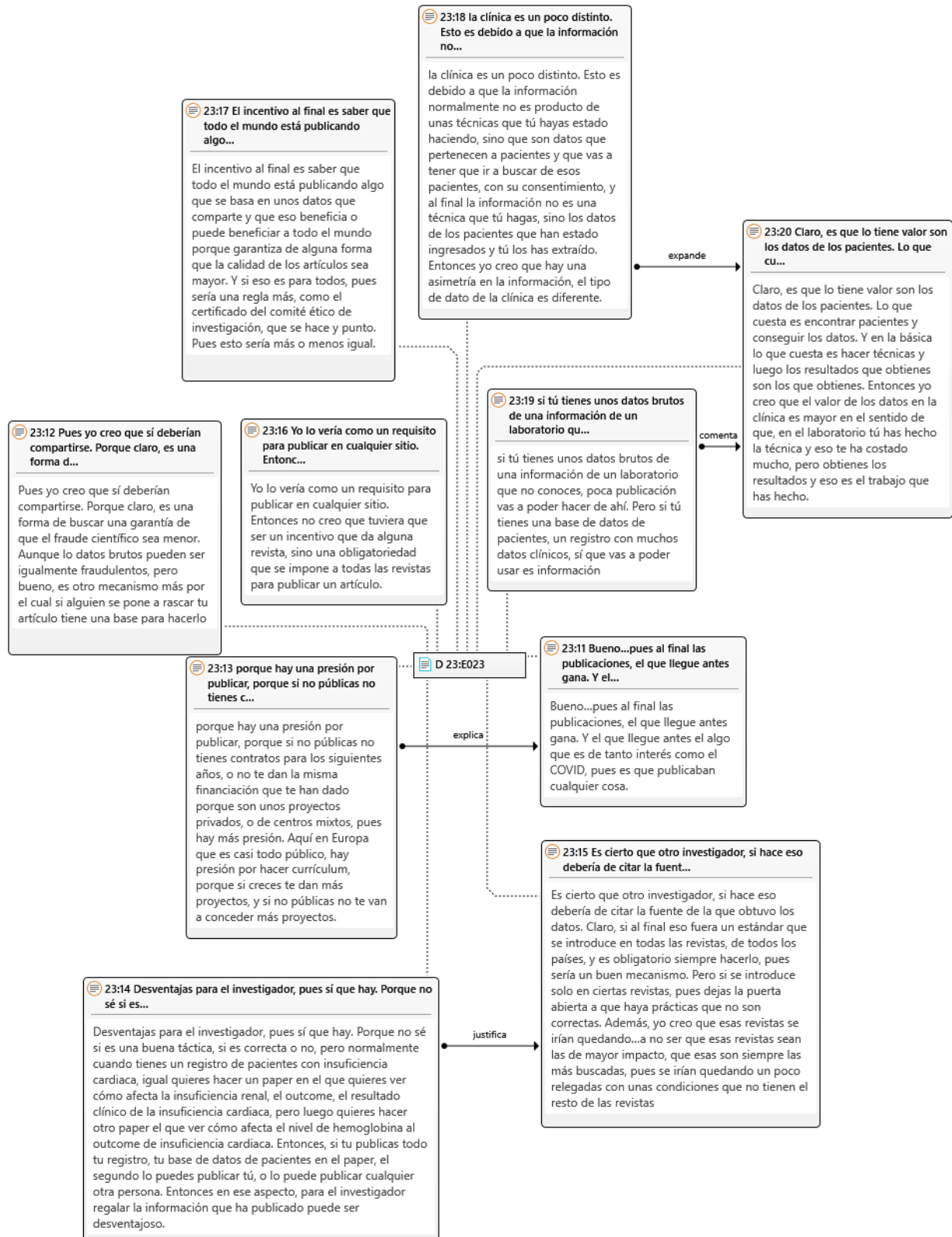


CREENCIAS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADA 023 (E023)

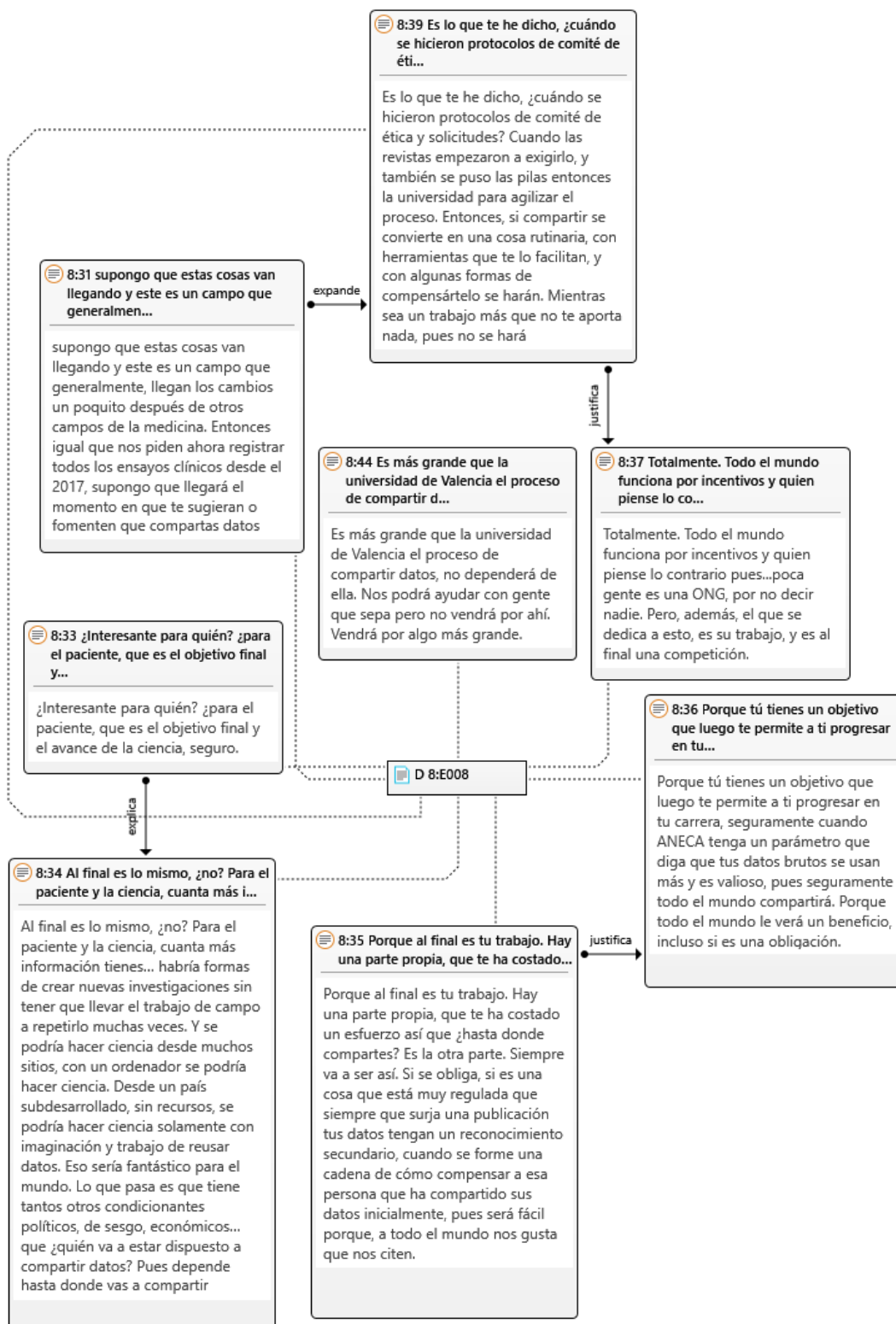


CREENCIAS

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADA 008 (E008)

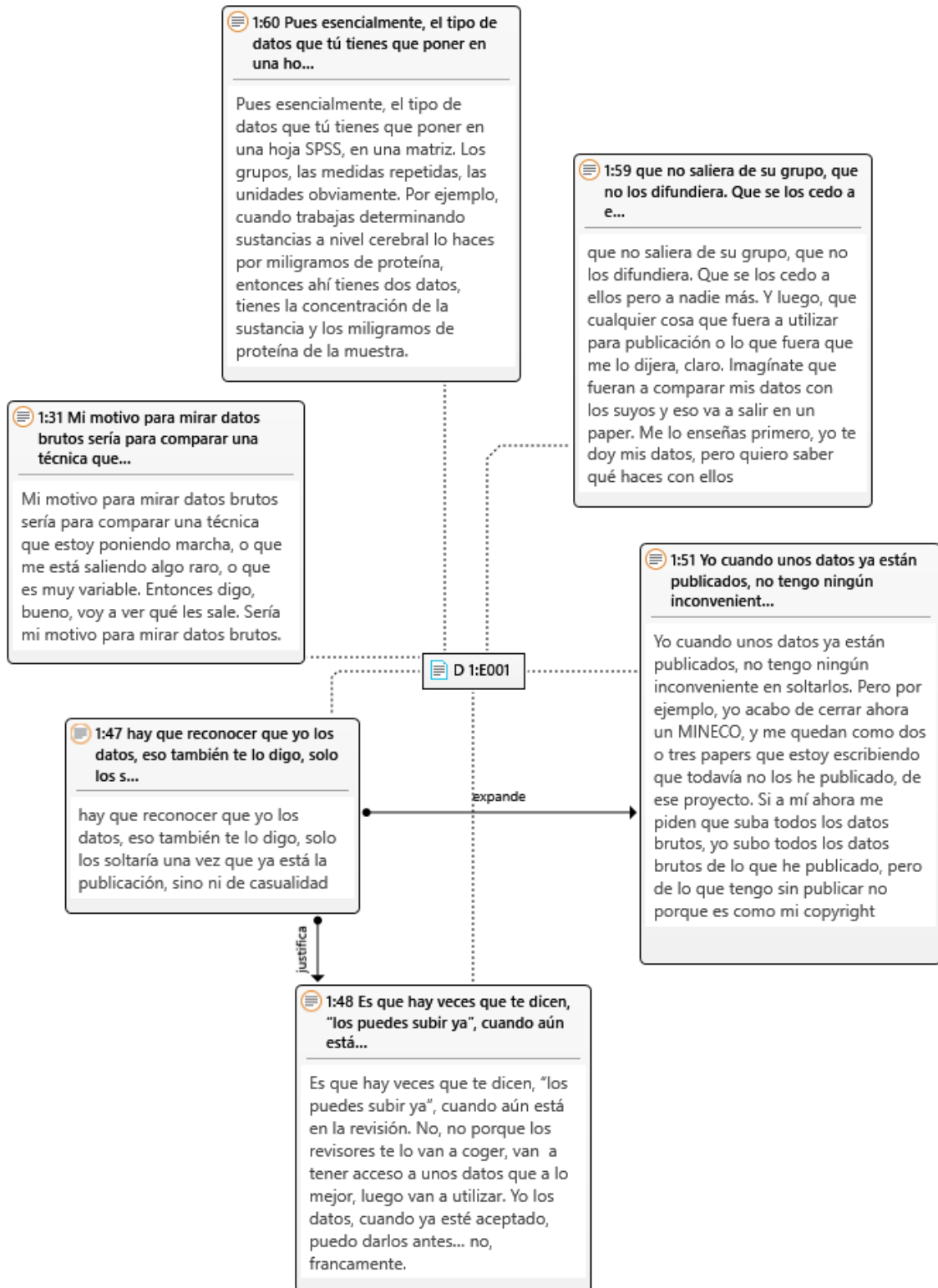


ACTITUDES

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 001 (E001)

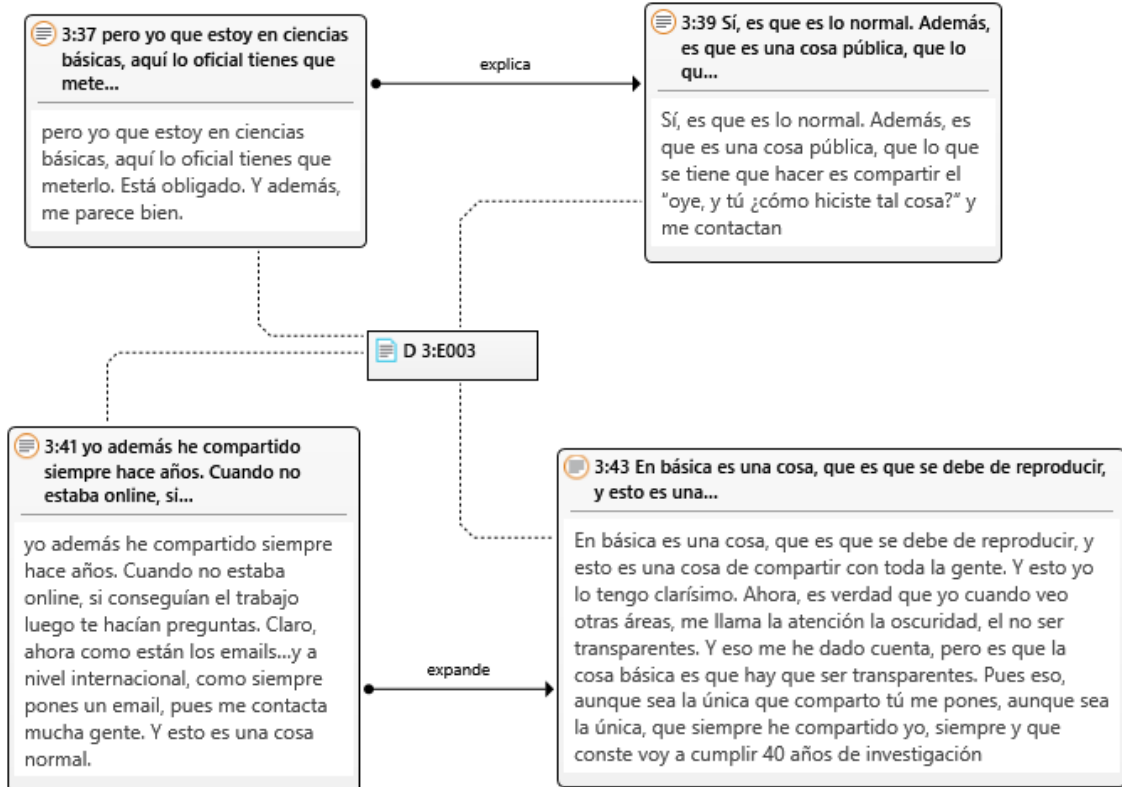


ACTITUDES

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

ENTREVISTADA 003 (E003)

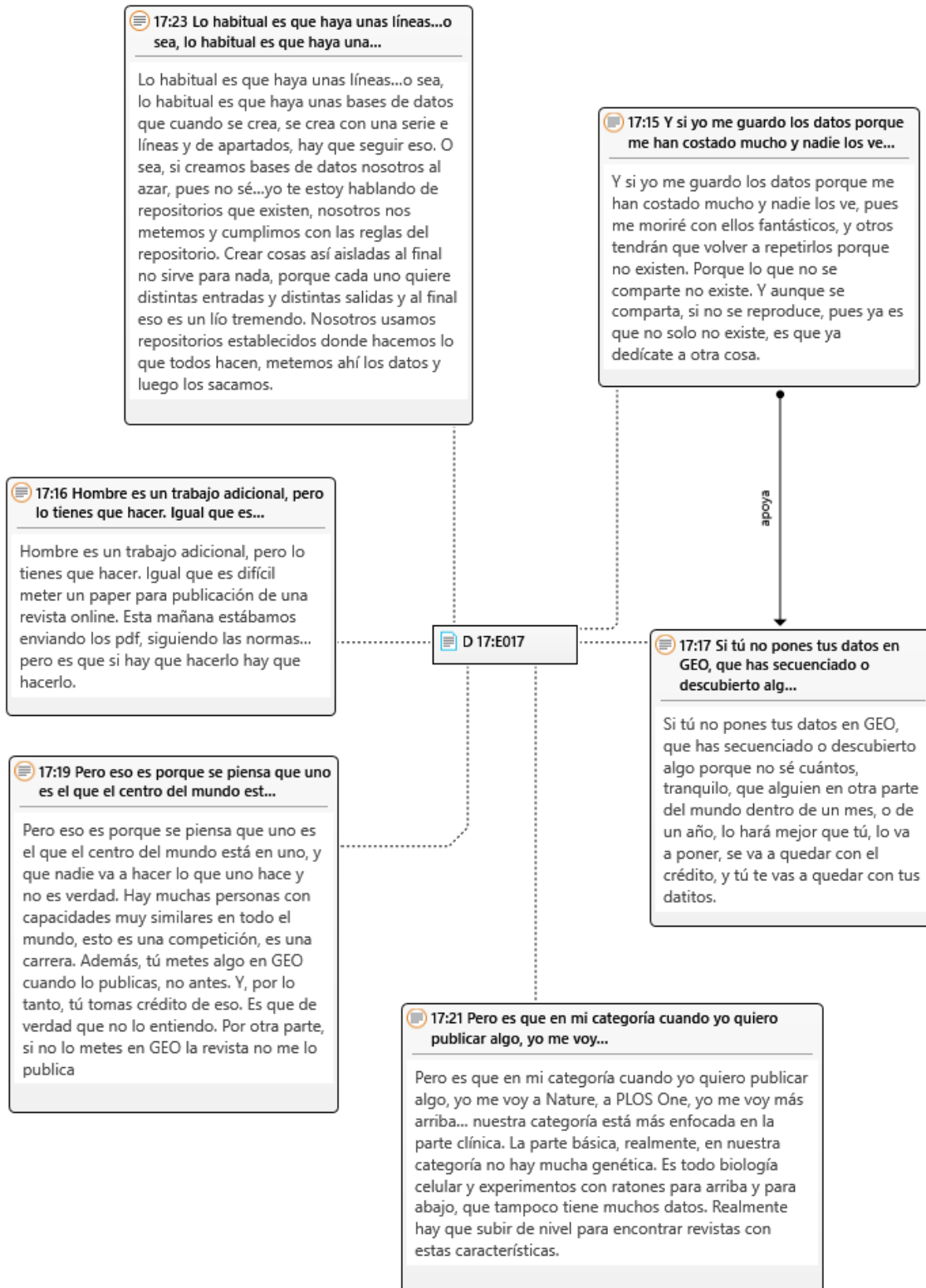


ACTITUDES

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADO 017 (E017)

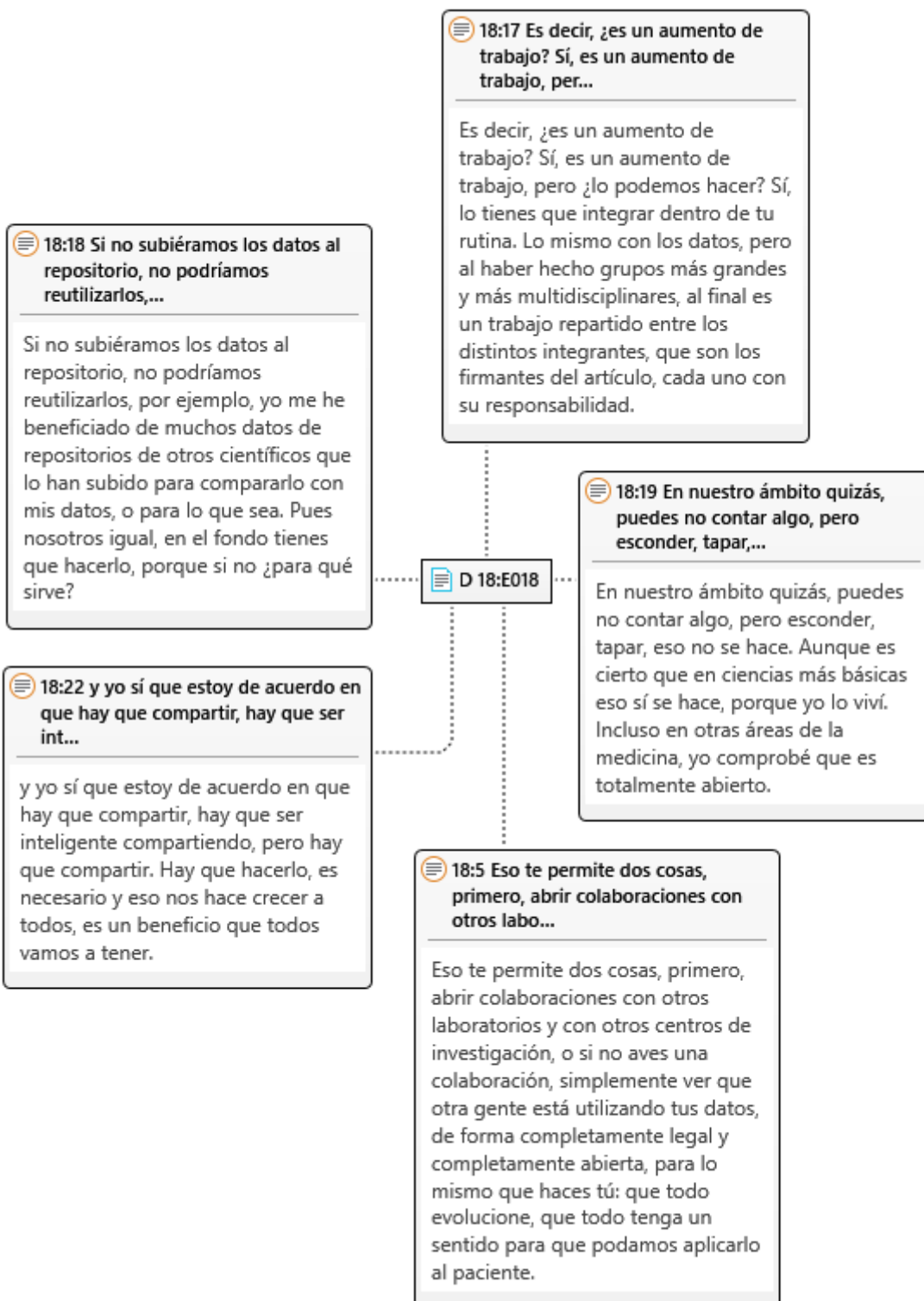


ACTITUDES

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADO 018 (E018)

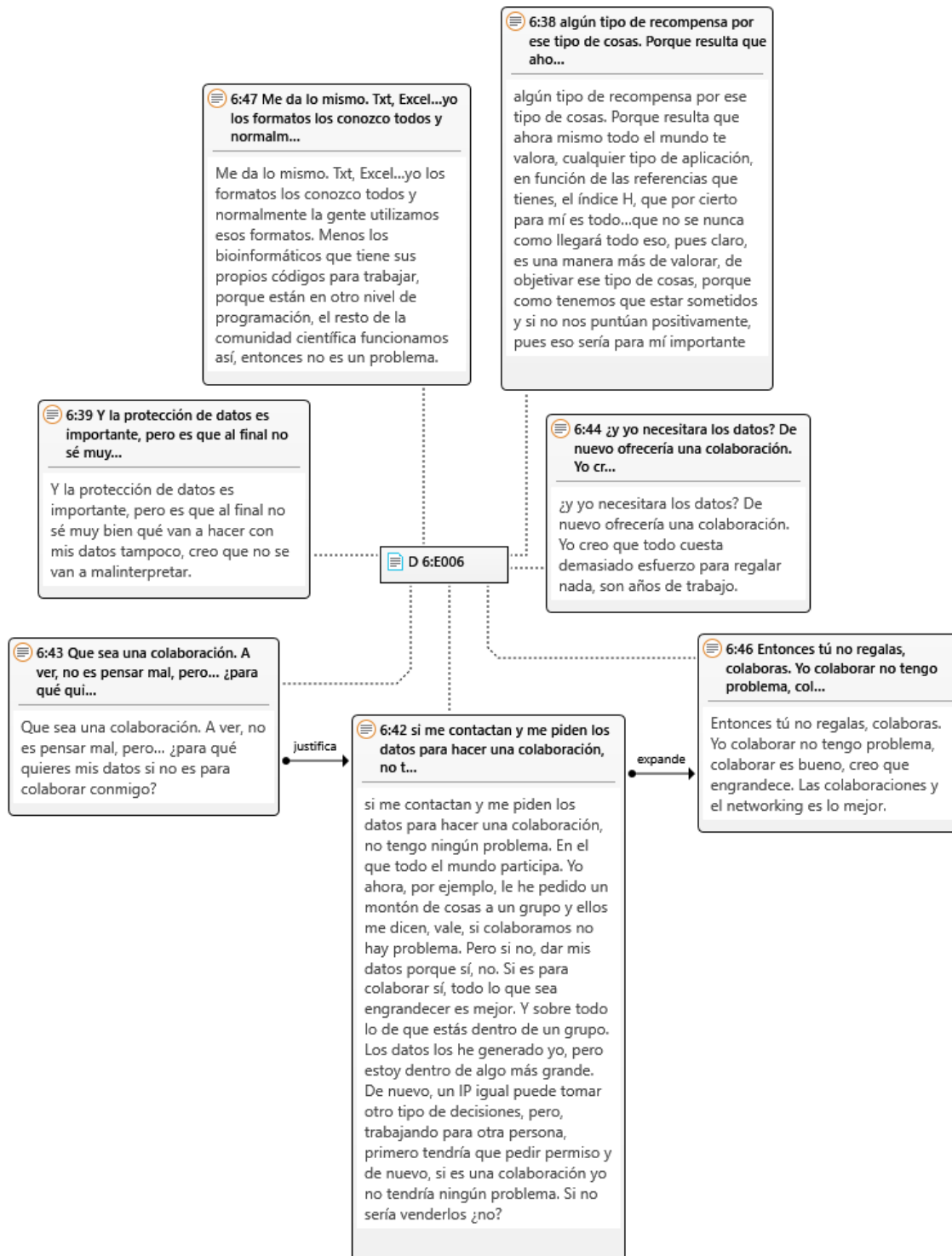


ACTITUDES

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 006 (E006)

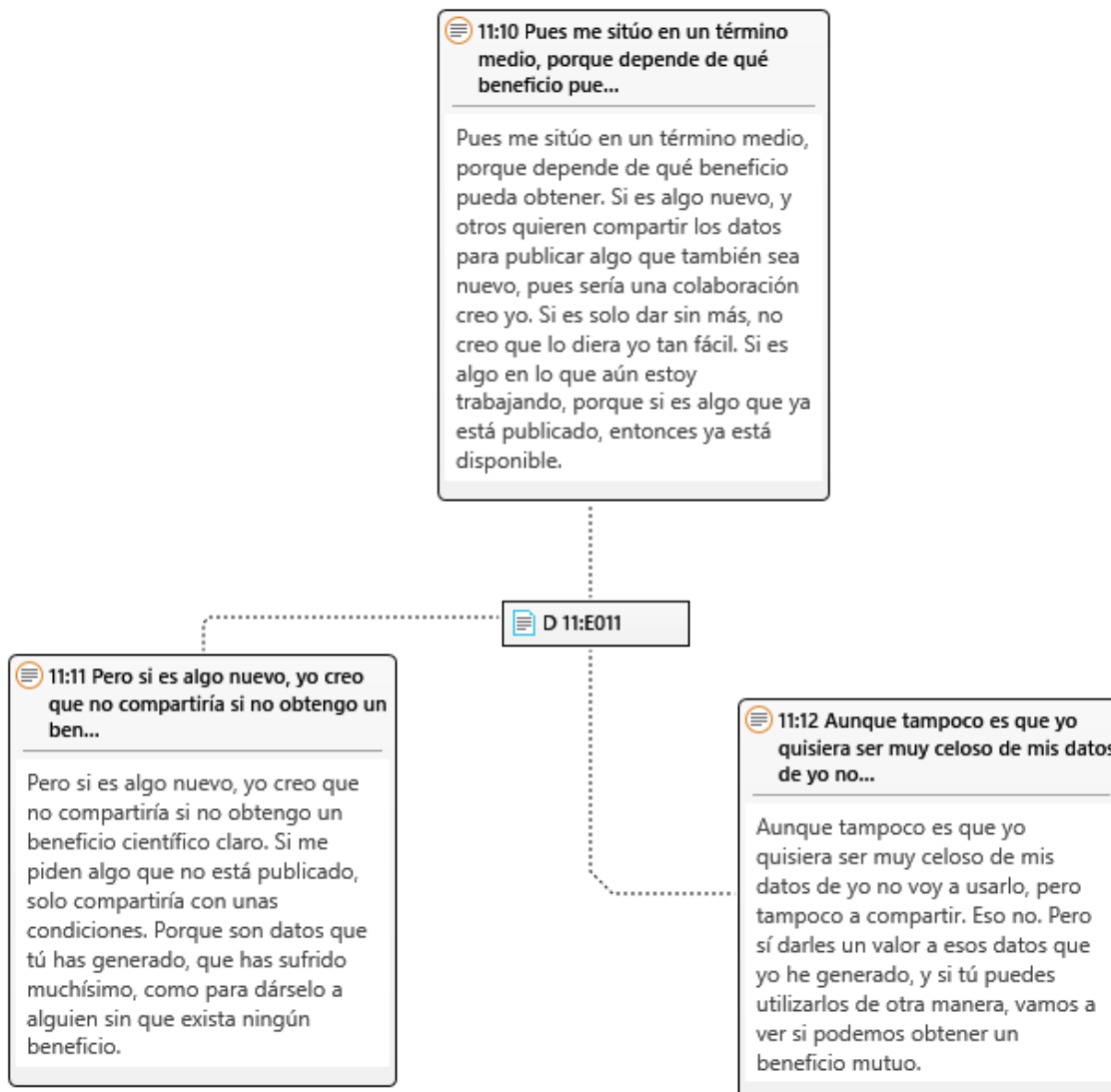


ACTITUDES

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADA 011 (E011)

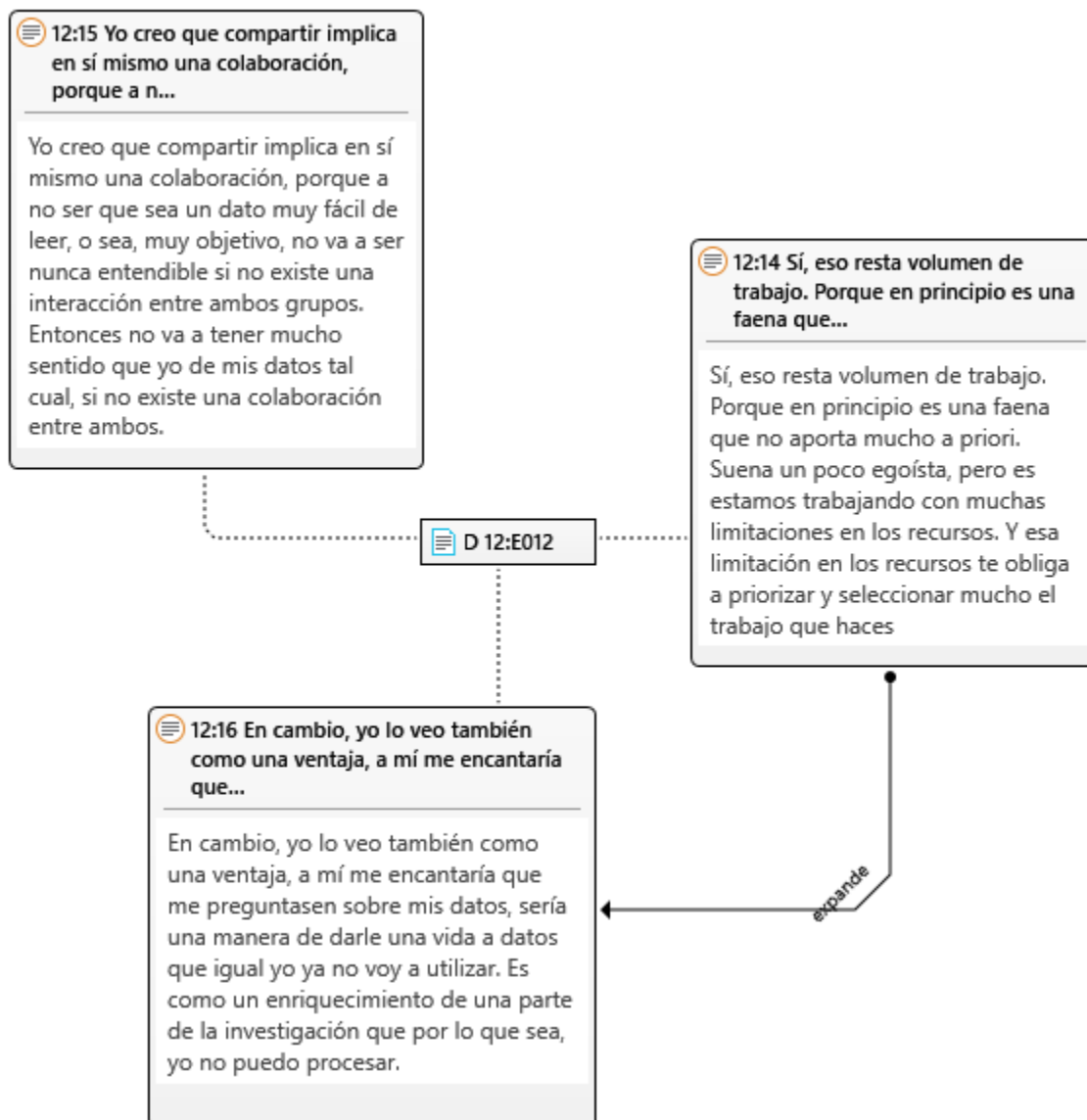


ACTITUDES

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADO 012 (E012)

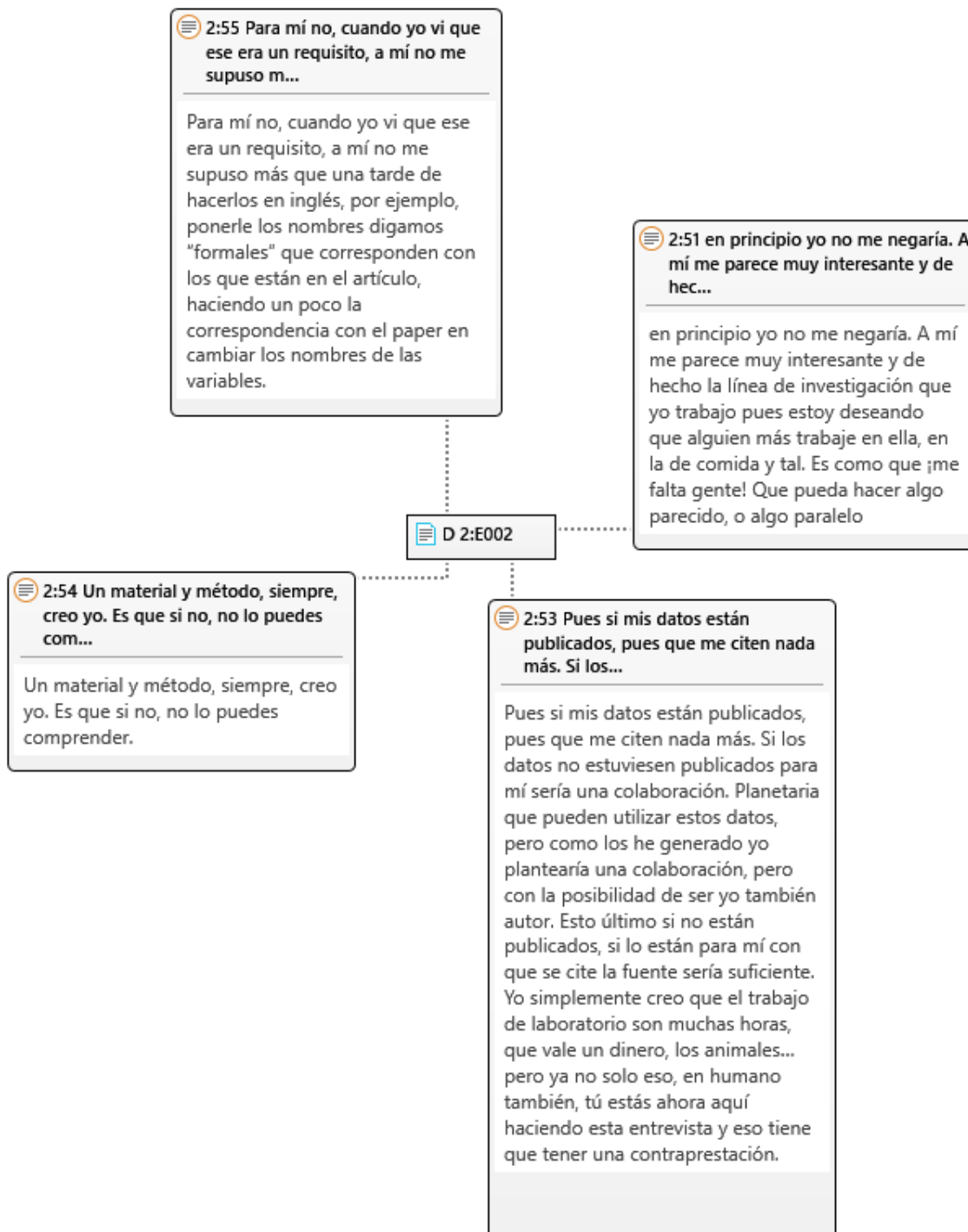


ACTITUDES

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 002 (E002)

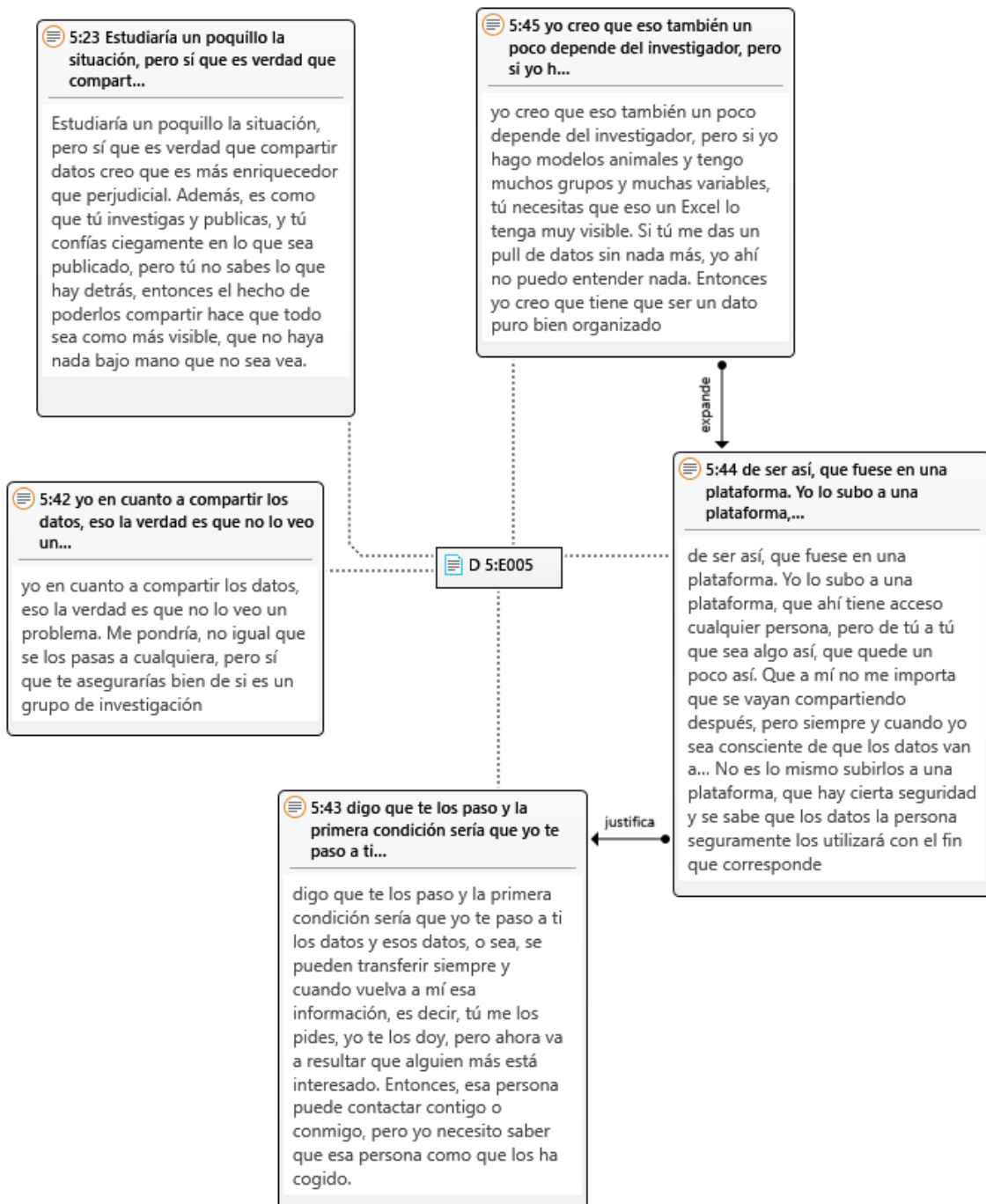


ACTITUDES

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 005 (E005)

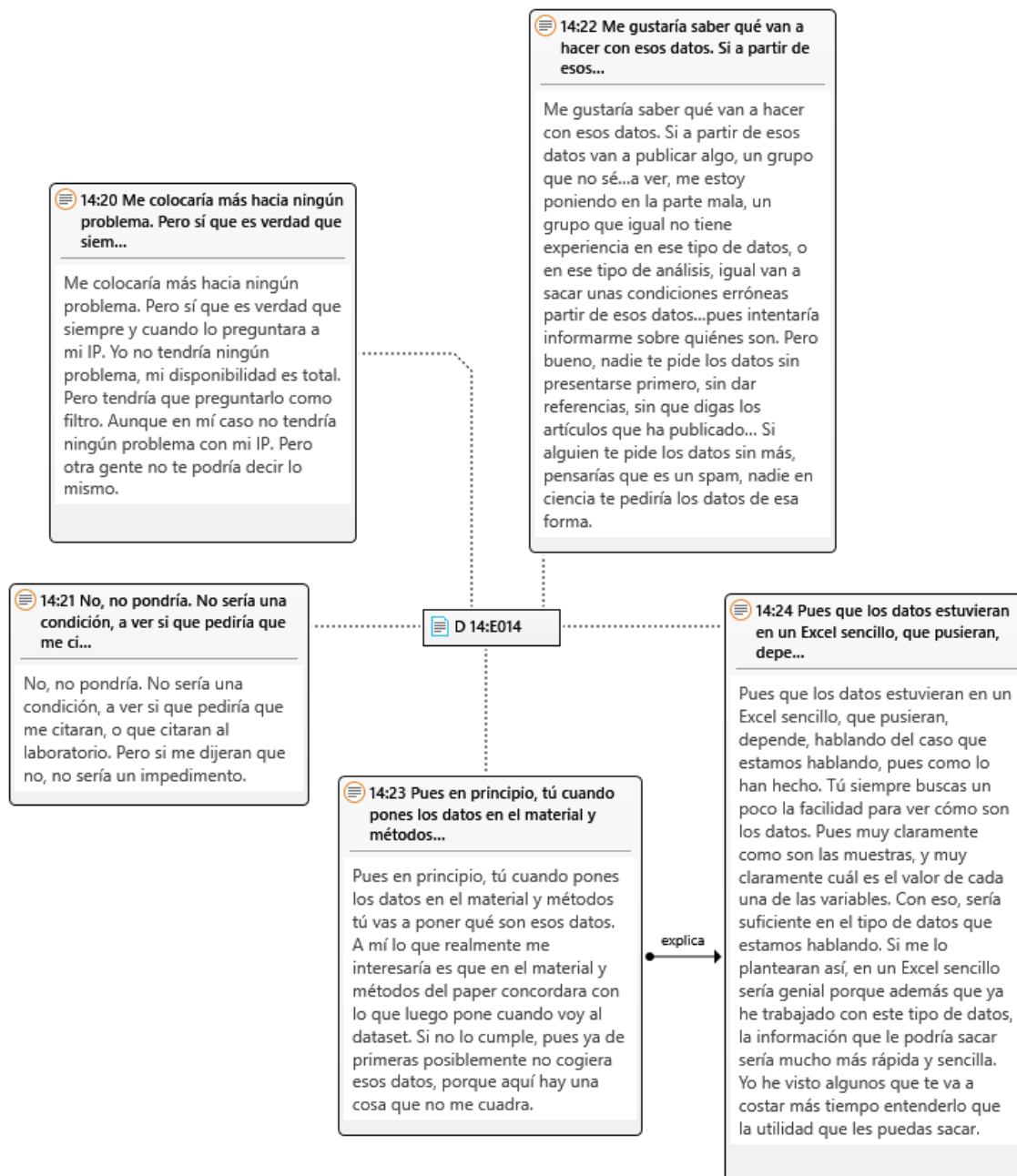


ACTITUDES

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 014 (E014)

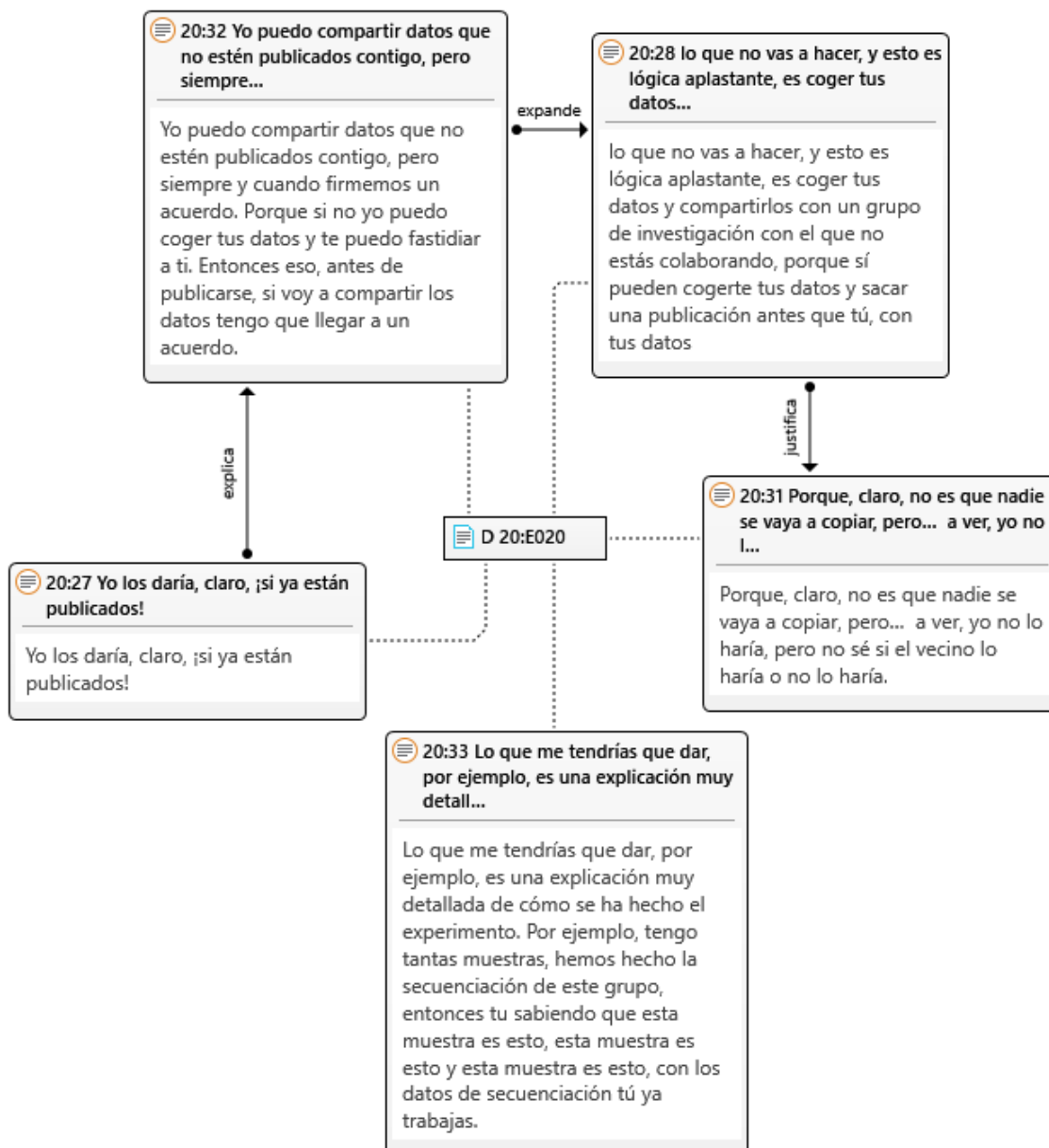


ACTITUDES

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 020 (E020)

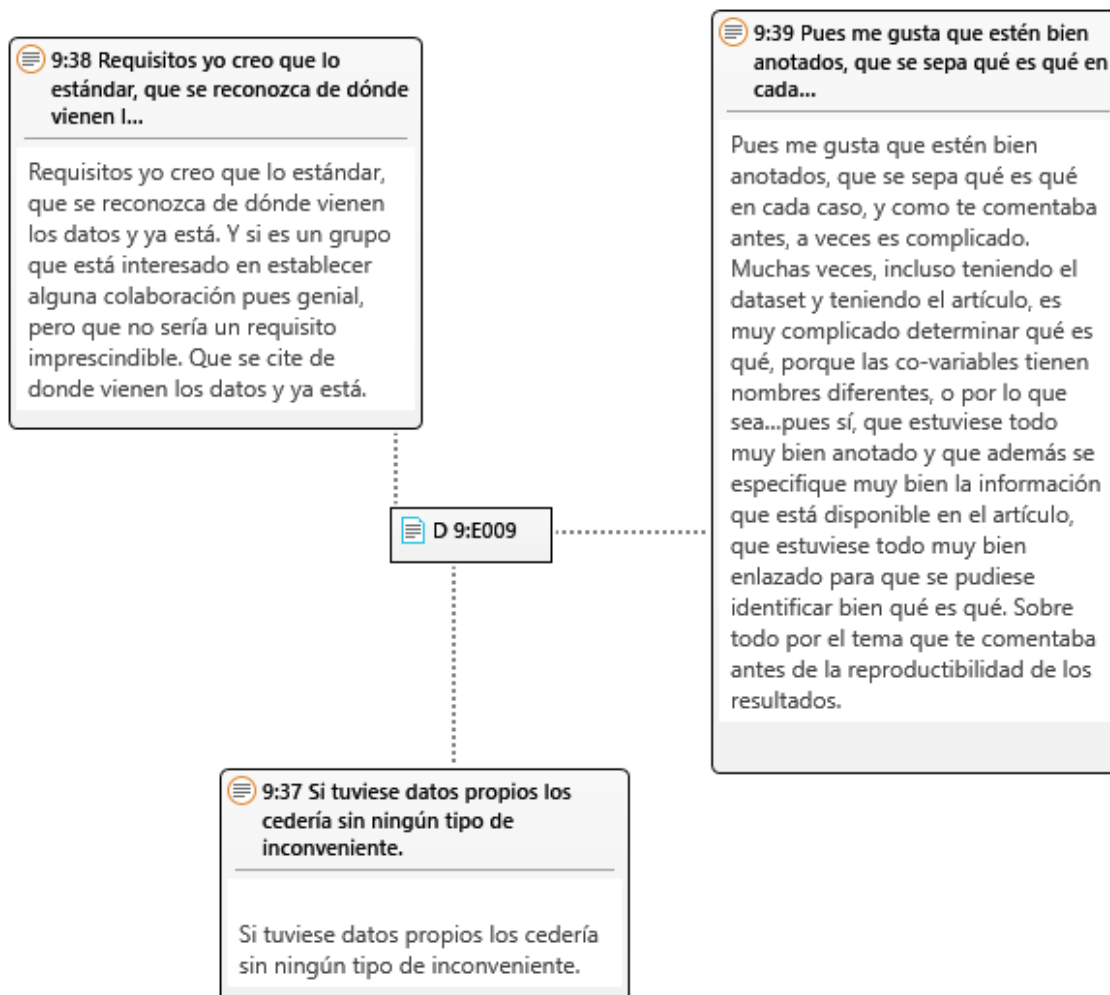


ACTITUDES

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 009 (E009)

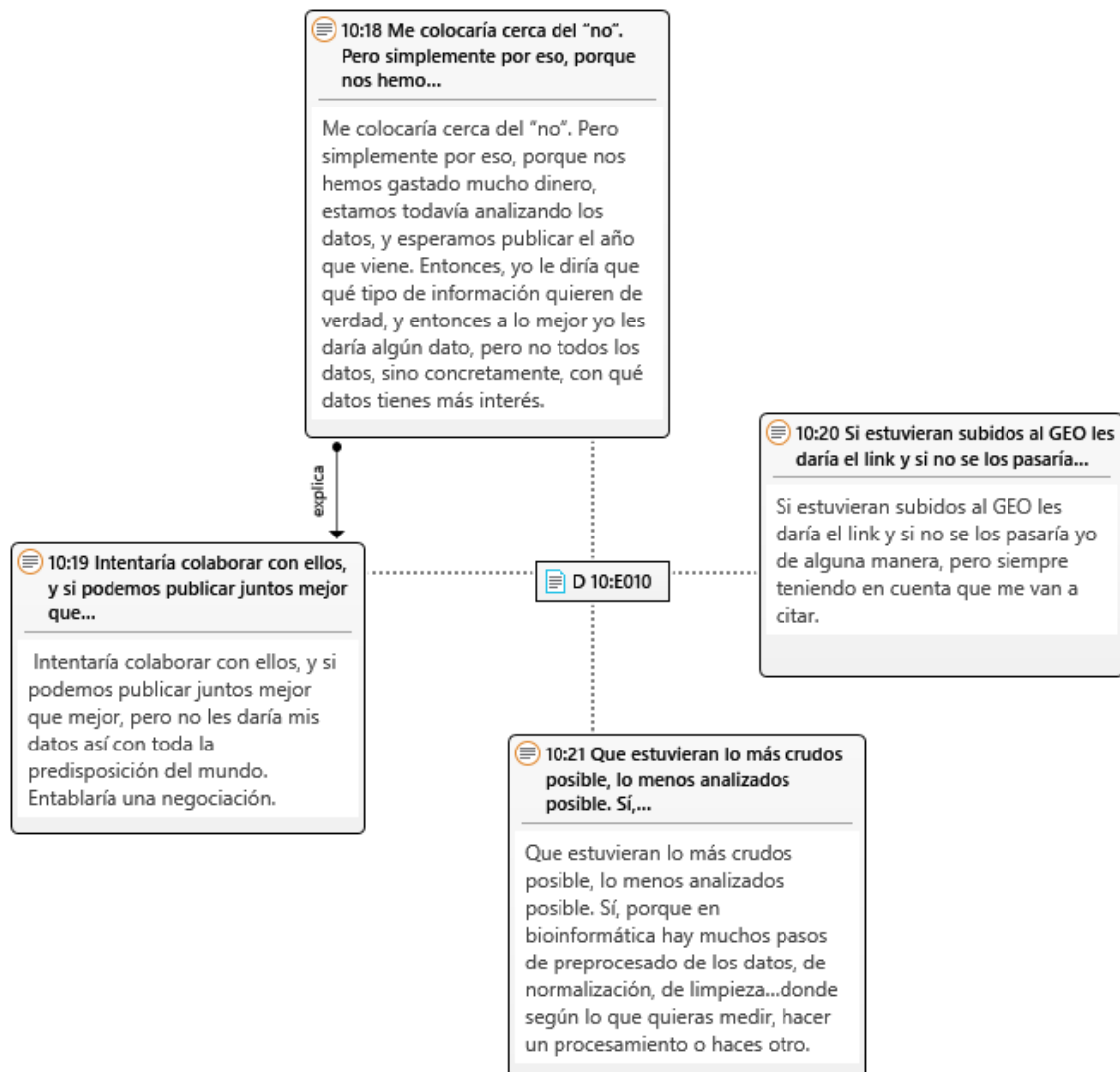


ACTITUDES

INVESTIGACIÓN BÁSICA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 010 (E010)

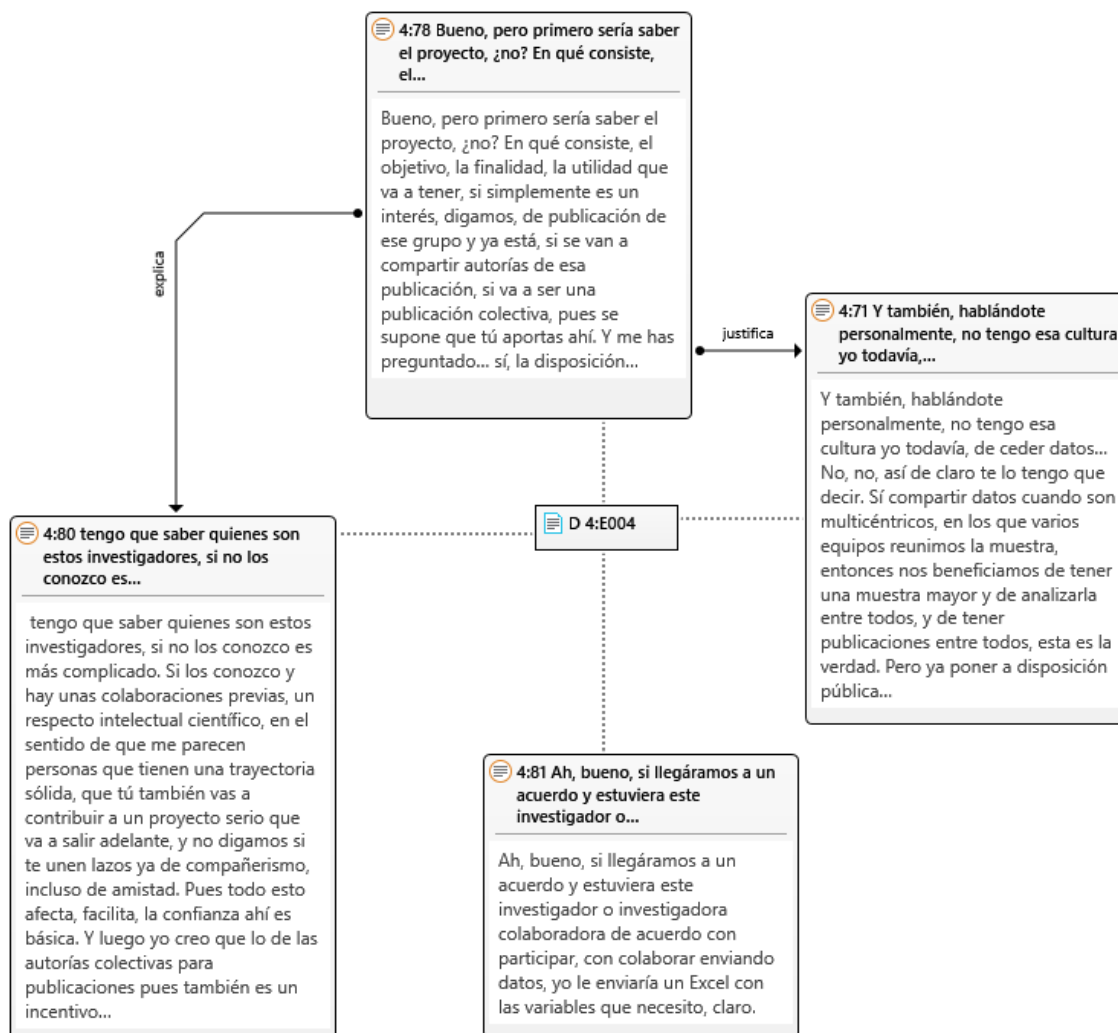


ACTITUDES

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADO 004 (E004)

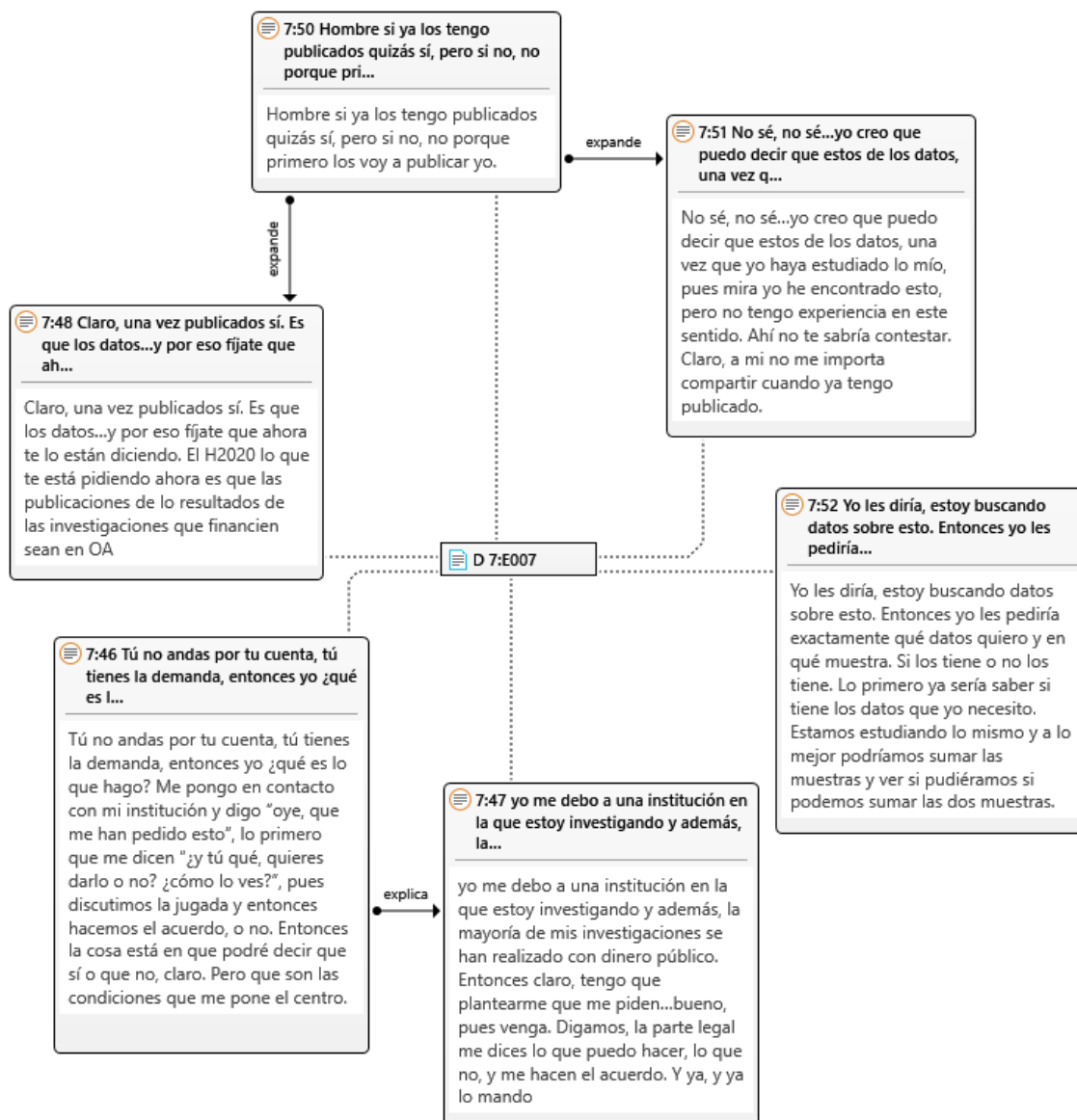


ACTITUDES

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADO 007 (E007)

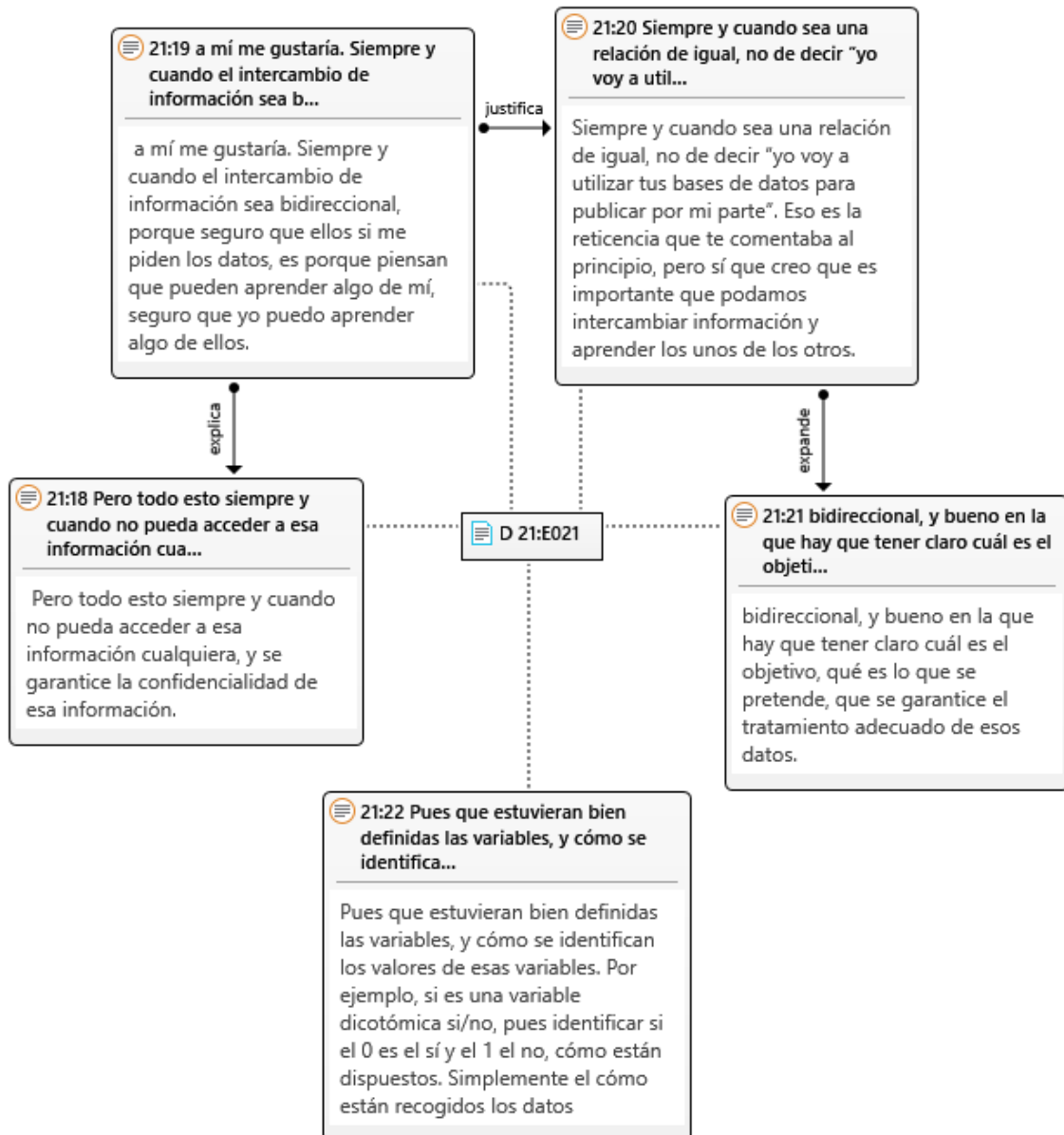


ACTITUDES

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADO 021 (E021)

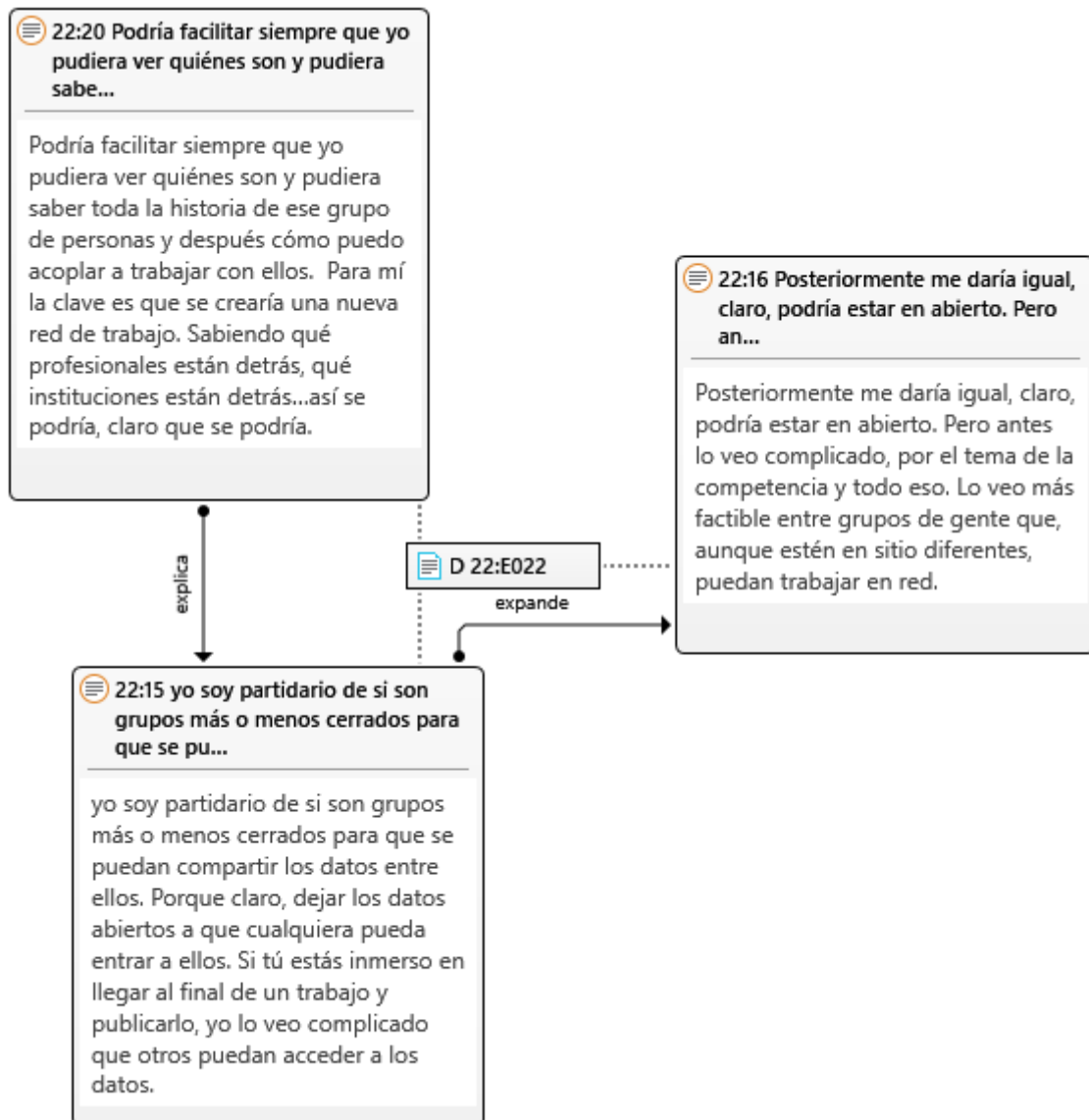


ACTITUDES

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADO 022 (E022)

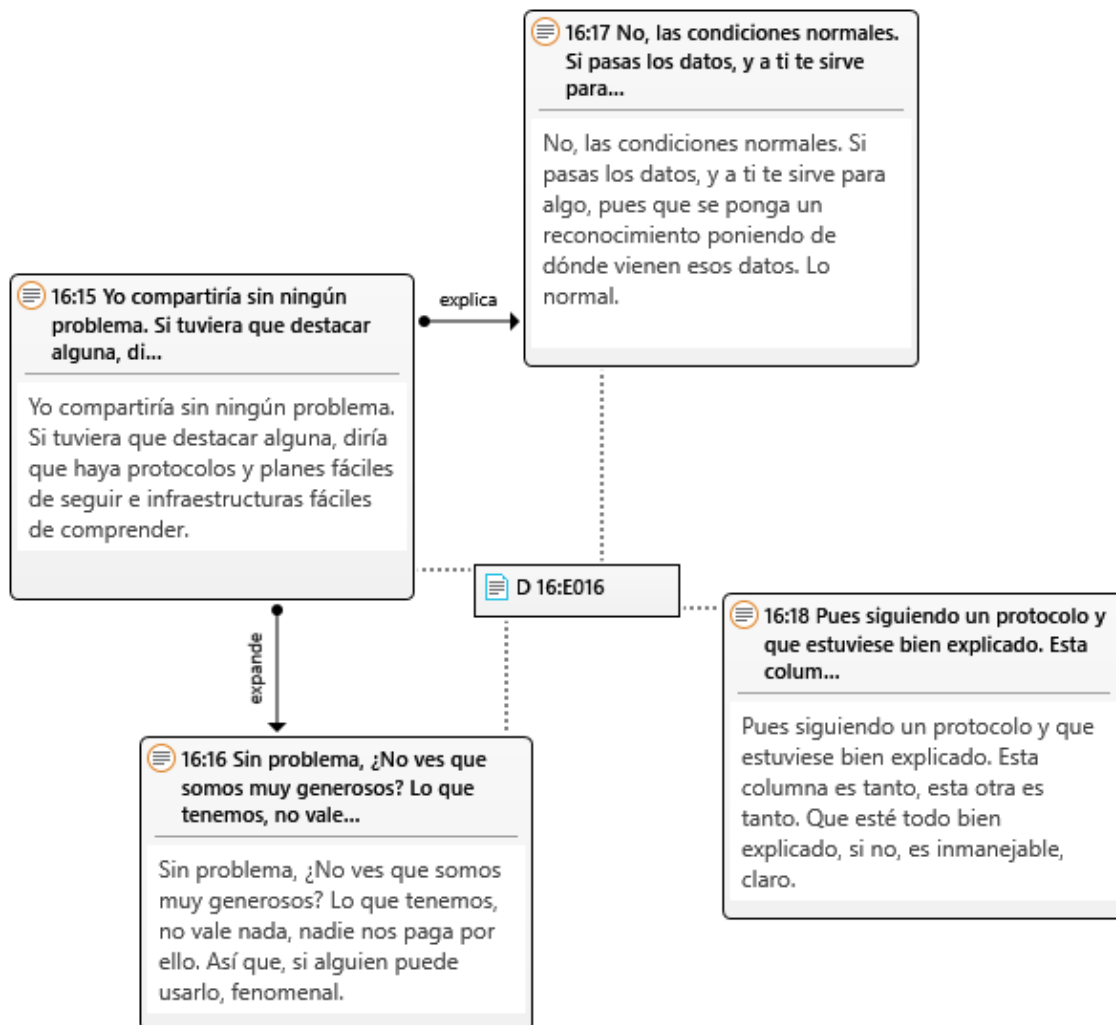


ACTITUDES

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS SÉNIOR

- ENTREVISTADO 016 (E016)

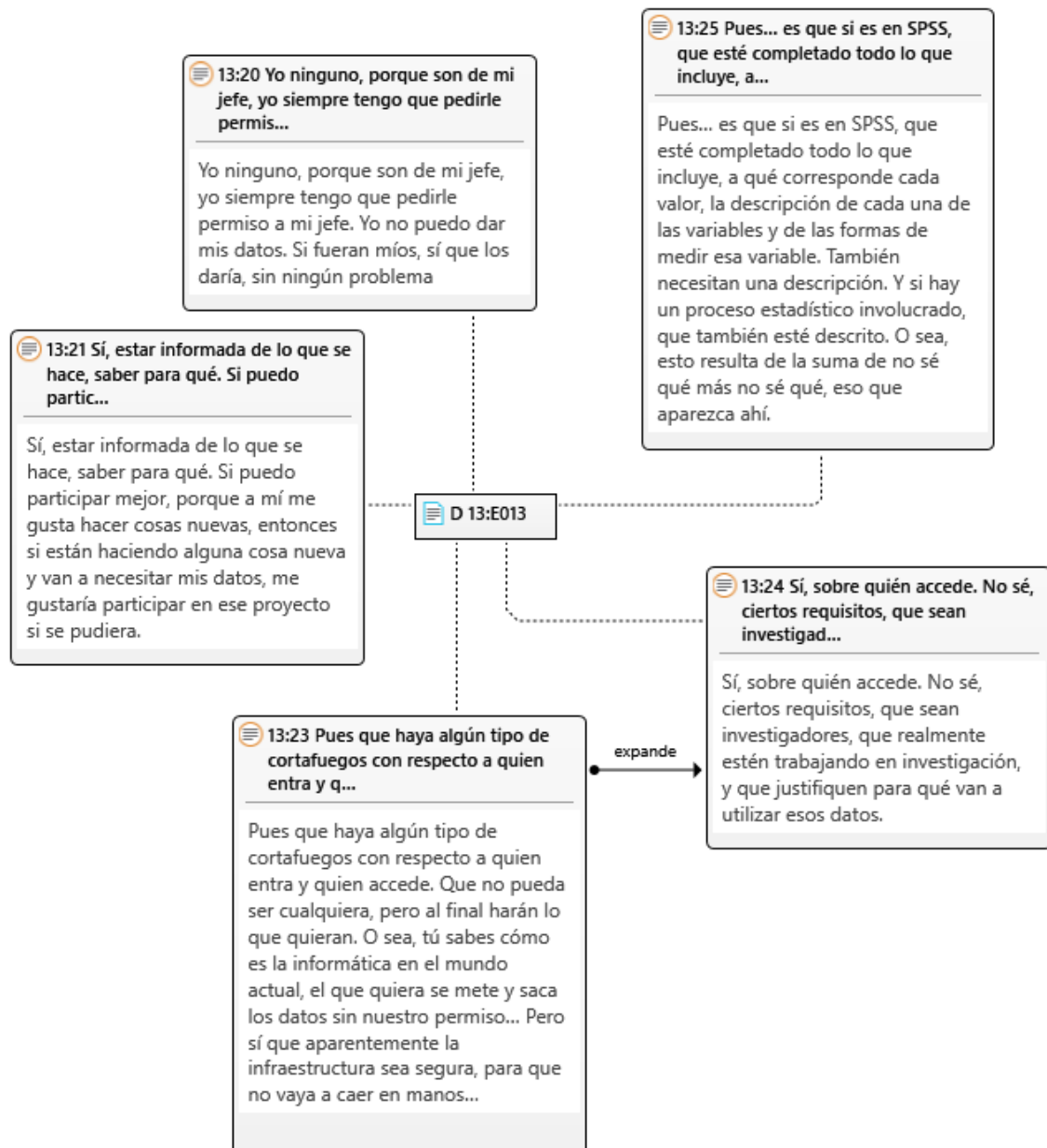


ACTITUDES

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 013 (E013)

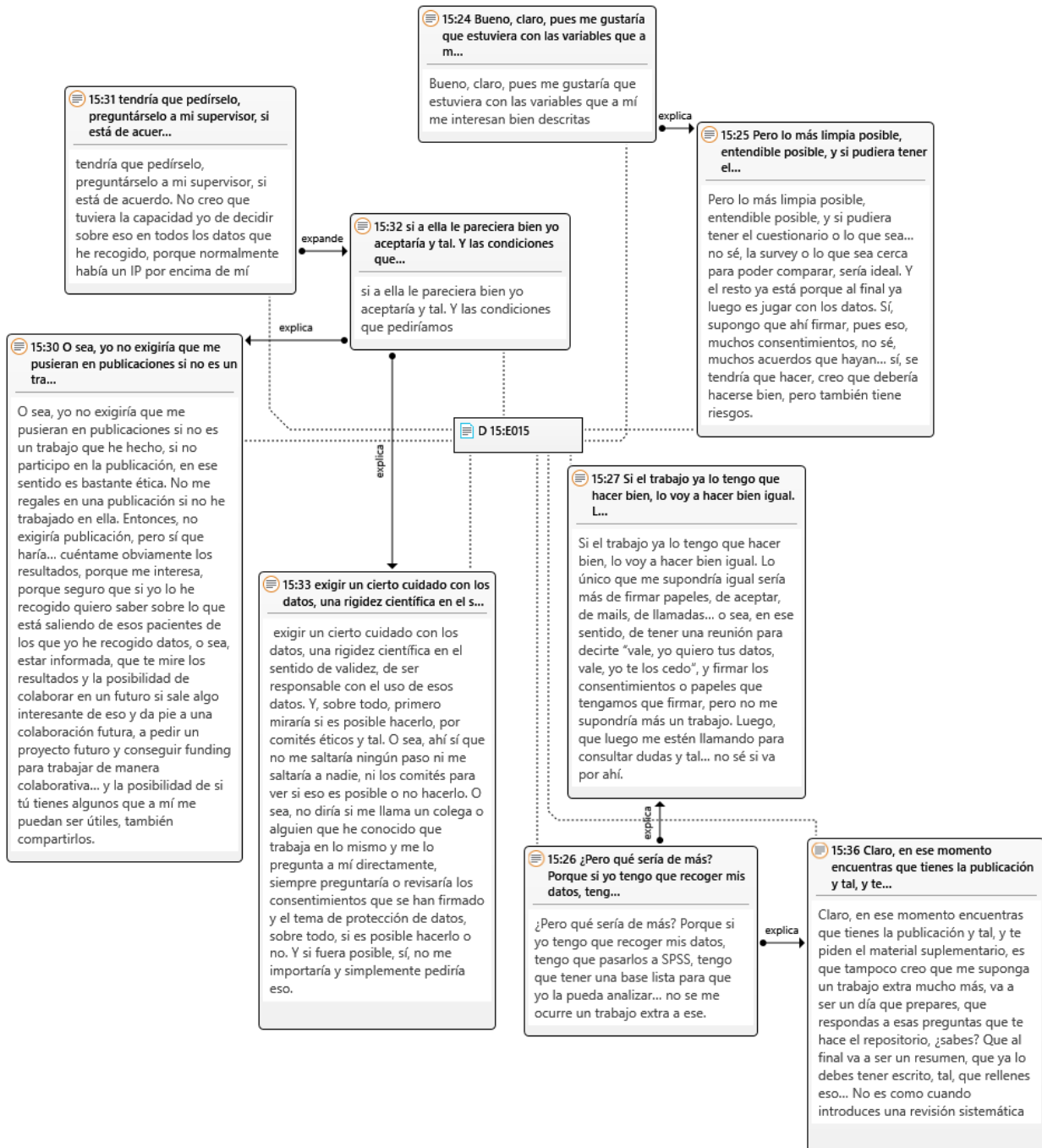


ACTITUDES

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 015 (E015)

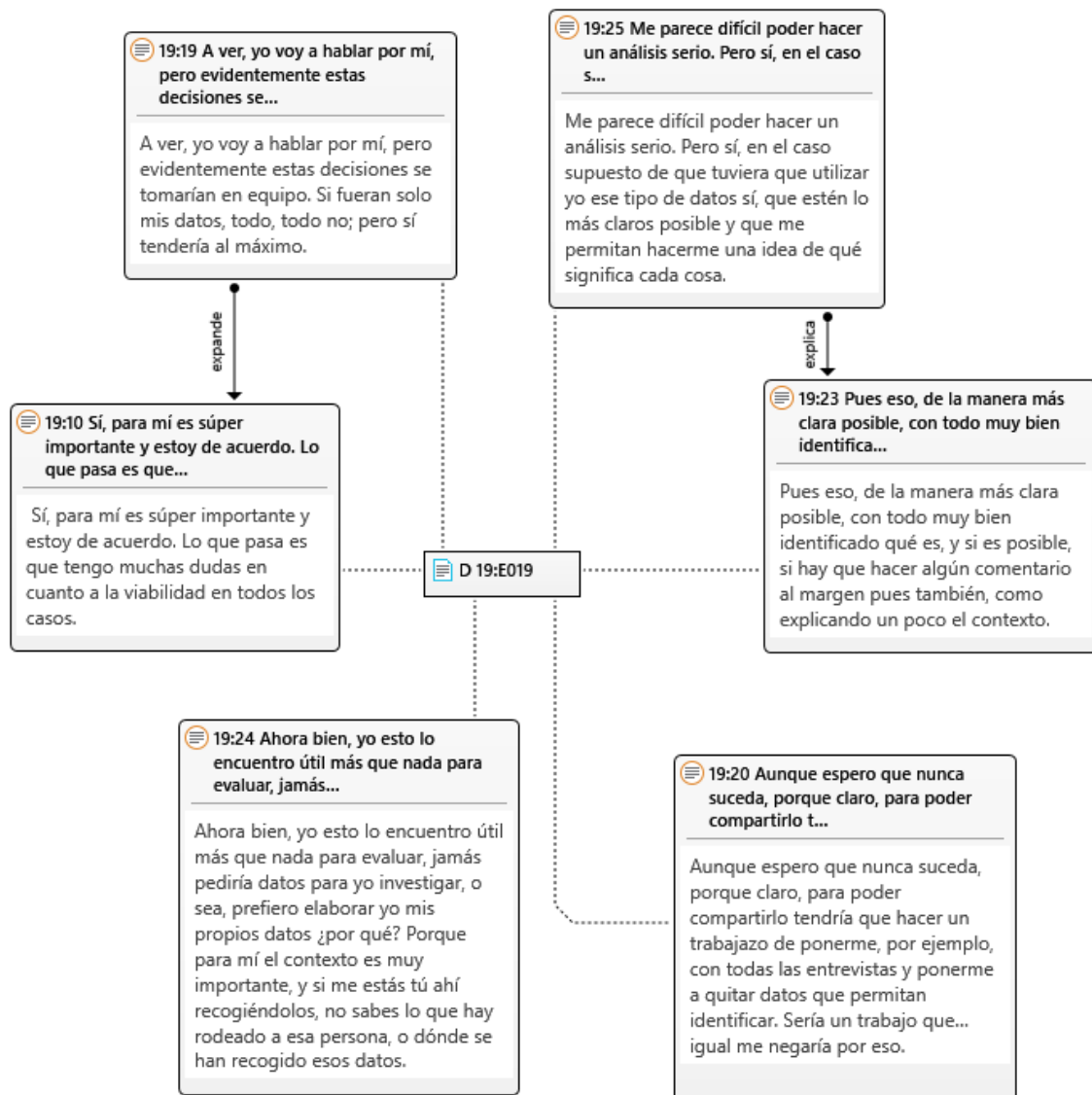


ACTITUDES

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 019 (E019)

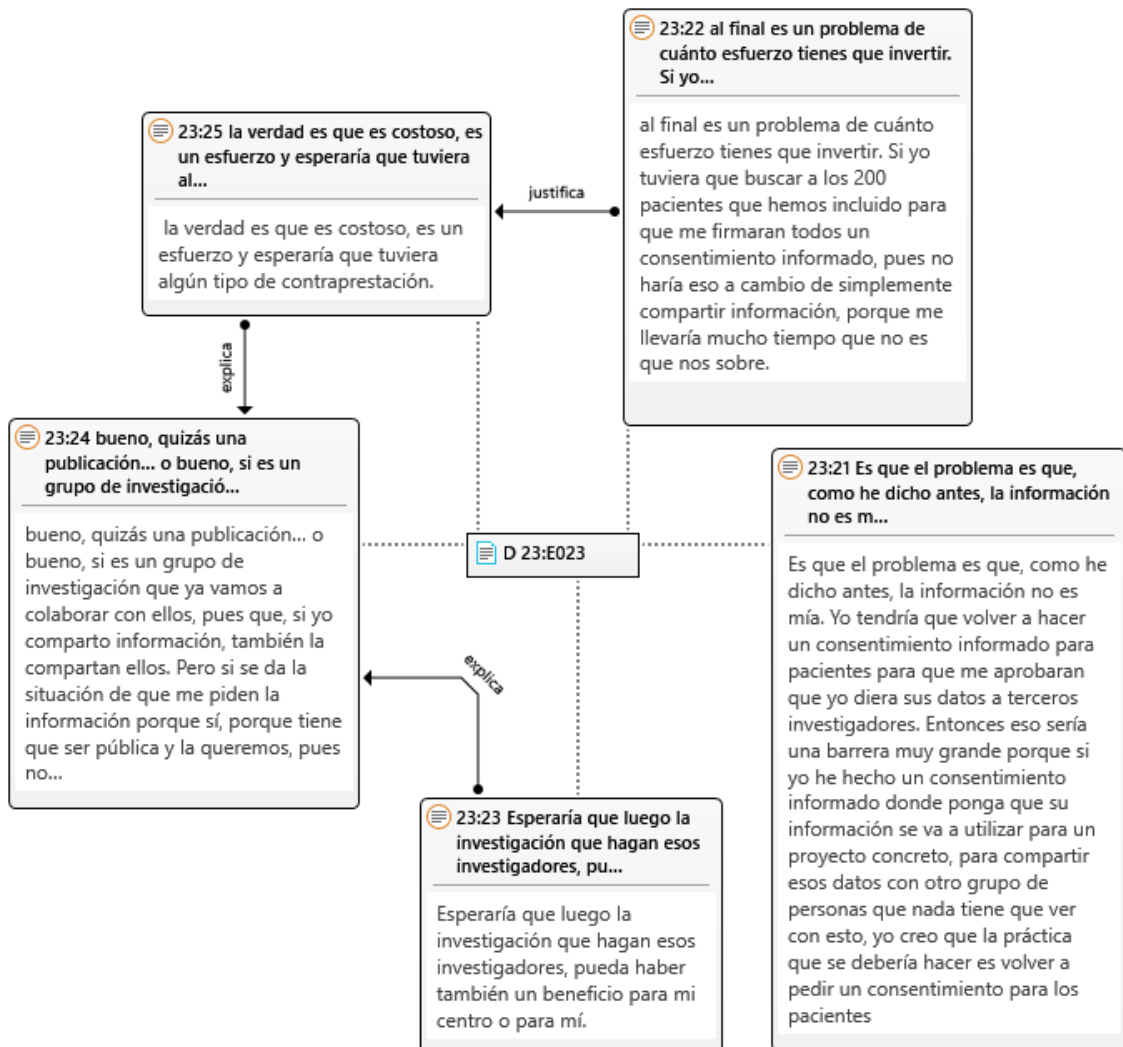


ACTITUDES

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 023 (E023)



ACTITUDES

INVESTIGACIÓN APLICADA

INVESTIGADORES E INVESTIGADORAS JÚNIOR

- ENTREVISTADO 008 (E008)

