

Sobre la gestión de datos de investigación científica con financiación pública y su optimización mediante *blockchain* e inteligencia artificial

* * * *

Juan Pablo Romano

Fundación Instituto Leloir

jpromano@leloir.org.ar

<https://orcid.org/0009-0000-3180-6638>

Recibido: 14 de octubre de 2024

Aceptado: 29 de noviembre de 2024

Resumen

En un contexto tradicional de gestión de datos provenientes de la investigación científica financiada con fondos públicos, en la República Argentina surgen innumerables cuestiones que pueden dar como resultado, en varios casos, una ineficiente e ineficaz gestión. Tales particularidades convergen en repetitivas e innecesarias erogaciones al tener que realizar varias investigaciones que recaben los mismos datos, sumando el temor de los investigadores a una alteración de la propiedad y/o autoría de las series de datos o bases de datos de las investigaciones de las cuales forman parte. Se suma a ello el excesivo tiempo que se tarda en tener los datos disponibles para la comunidad científica o público, luego de que hayan sido recabados. Con base en los avances tecnológicos y a la aparición de nuevas tecnologías enfocadas en gestión de información, como el *blockchain* y la inteligencia artificial, es necesario realizar una propuesta de optimización de la gestión de datos, para que, de esta manera, se produzca un ahorro y redirección de financiamiento, a la vez de un mejoramiento y la consecuente optimización de los resultados de las investigaciones científicas mientras se consolida la protección de la propiedad intelectual en la inmensidad de datos que conforman el big data, integrada por todos los datos resultantes de las investigaciones científicas financiadas con fondos públicos.

Palabras clave: gestión de datos, big data, *blockchain*, inteligencia artificial, propiedad intelectual, derecho de autor.

On the Management of Publicly Funded Scientific Research Data and its Optimization through Blockchain and Artificial Intelligence

Abstract

In the traditional context of data management from publicly funded scientific research in Argentina, countless issues arise that, in many cases, result in inefficient and ineffective management. These issues converge in repetitive costs due to the need to conduct multiple investigations that collect the same data; the researchers' fear of alteration of the ownership and/or authorship of the data sets or databases of the investigations they are part of; and the excessive time it takes to make the data available to the scientific community or the public after it has been collected. Based on technological advances and the emergence of new technologies focused on information management, such as blockchain and artificial intelligence, it is necessary to propose an optimization of data management, in order to generate cost savings and redirect funding, while improving and consequently optimizing the results of scientific research, and at the same time consolidating the protection of intellectual property in the vast amount of data that constitutes the big data generated from publicly funded scientific research.

Key words: data management, big data, blockchain, artificial intelligence, intellectual property, copyright.

Sobre a gestão de dados de pesquisa científica com financiamento público e sua otimização por meio de blockchain e inteligência artificial

Resumo

Em um contexto tradicional de gestão de dados provenientes de pesquisas científicas financiadas com recursos públicos, na República Argentina, surgem inúmeras questões que podem resultar, em muitos casos, em uma gestão ineficiente e ineficaz. Tais particularidades levam a gastos repetitivos devido à necessidade de realizar várias pesquisas que recolhem os mesmos dados, além do receio dos pesquisadores quanto à alteração da propriedade e/ou autoria das séries de dados ou bancos de dados das pesquisas de que fazem parte, e o tempo excessivo que leva para que os dados estejam disponíveis para a comunidade científica ou o público, após serem coletados.

Com base nos avanços tecnológicos e no surgimento de novas tecnologias focadas na gestão da informação, como o blockchain e a inteligência artificial, é necessário propor uma otimização da gestão de dados, de modo que ocorra uma economia e re-direcionamento de financiamento, além da melhoria e consequente otimização dos resultados das pesquisas científicas, enquanto se consolida a proteção da propriedade intelectual na imensidão de dados que compõem o big data integrado por todos os dados resultantes de pesquisas científicas financiadas com recursos públicos.

Palavras-chave: gestão de dados, big data, blockchain, inteligência artificial, propriedade intelectual, direito autoral.

1. Introducción

En la República Argentina, jurisdicción en la que se encuadra el caso que analizo en este documento, el Estado nacional, a través de determinados entes de su estructura, lleva a cabo diversos tipos de investigaciones científicas con el fin de implementar sus políticas, planes y programas. A dicha estructura se suman las investigaciones realizadas por universidades, organismos no gubernamentales, empresas y otras entidades que, de una u otra manera, están vinculadas administrativa, jurídica y/o presupuestariamente al Estado nacional, o que, en determinados casos, reciben financiamiento de este.

No abordaré investigaciones que contengan datos personales, clínicos o sensibles que requieran algún tipo de confidencialidad ni aquellas que exijan la aplicación de regulaciones sobre datos personales, ya que esto sería objeto de un análisis específico fuera del alcance del presente documento.

A los efectos del presente análisis, y bajo la normativa de Argentina, es menester presentar el Sistema Nacional de Ciencia, Tecnología e Innovación (SNCTI), creado por la Ley 25467, que lo conceptualiza como

un marco general que structure, impulse y promueva las actividades de ciencia, tecnología e innovación, a fin de contribuir a incrementar el patrimonio cultural, educativo, social y económico de la Nación, propendiendo al bien común, al fortalecimiento de la identidad nacional, a la generación de trabajos y a la sustentabilidad del medio ambiente.¹

1 Artículo 1, Ley 25467, disponible en: <https://servicios.infoleg.gob.ar/info->

El Sistema está compuesto por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), la Comisión Nacional de Energía Atómica (CNEA), el Instituto Nacional de Tecnología Agropecuaria (INTA), el Instituto Nacional de Tecnología Industrial (INTI), la Comisión Nacional de Actividades Espaciales (CONAE), el Servicio Geológico Minero Argentino (SEGEMAR), el Instituto Nacional de Desarrollo Pesquero (INIDEP), el Instituto Nacional del Agua (INA), el Centro de Investigación Tecnológica de las Fuerzas Armadas (CITEFA), la Administración Nacional de Laboratorios e Institutos de Salud (ANLIS) y un rector de universidad nacional de cada región del país a propuesta del Consejo Interuniversitario Nacional. Por imperio del artículo 4 de la citada ley, también integran el Sistema

... las universidades, el conjunto de los demás organismos, entidades e instituciones del sector público nacional, provincial, municipal y de la Ciudad Autónoma de Buenos Aires y del sector privado que adhieren a esta norma, que realicen actividades sustantivas vinculadas al desarrollo científico, tecnológico, innovador, de vinculación, financiamiento, formación y perfeccionamiento de recursos humanos, así como sus políticas activas, estrategias y acciones.²

Cada uno de los organismos integrantes del Sistema realiza distintos tipos de investigaciones científicas y/o desarrollos tecnológicos de acuerdo con sus competencias jurisdiccionales. A modo de ejemplo, citaré algunas de ellas para conceptualizar el ámbito de estudio del presente documento.

El INTA³ tiene por misión "... impulsar, vigorizar y coordinar el desarrollo de la investigación y extensión agropecuaria...", y para concreción de ésta, "... desarrollará y estimulará la investigación, experimentación y extensión agraria, como aspectos fundamentales..."⁴

legInternet/anexos/65000-69999/69045/texact.htm.

2 Artículo 4, Ley 25467, disponible en <https://servicios.infoleg.gob.ar/infolegInternet/anexos/65000-69999/69045/texact.htm>.

3 Creado por el Decreto-Ley 21680 del 4 de diciembre de 1956.

4 *Ibidem*, arts. 1 y 2.

El INIDEP⁵ tiene como finalidad “... la formulación y ejecución de programas de investigación pura y aplicada, relativos a los recursos pesqueros, de las algas y de la caza marítima, y a su explotación racional en todo el territorio Nacional...”.

La CONAE⁶ tiene entre sus funciones, según lo expresado por el art. 3, “... realizar tareas de investigación conducentes a la formación de grupos, que posean disciplinas y técnicas necesarias para el acceso a la tecnología espacial y sus aplicaciones”.

El CONICET⁷ tiene entre sus funciones

... Fomentar y subvencionar la investigación científica y tecnológica, y las actividades de apoyo a las mismas, tanto en el sector público como privado, que apunten al avance científico y tecnológico en el país, al desarrollo de la economía nacional y al mejoramiento de la calidad de vida, respetando los lineamientos establecidos por el Gobierno Nacional.

La descripción anterior de las funciones y/o misiones de algunas instituciones del Estado tiene como objetivo demostrar que, dentro de su ámbito, coexisten diversos entes que realizan investigación científica y desarrollos tecnológicos en las disciplinas de su competencia. Estos entes requieren información y datos que otras áreas o estamentos ya han relevado, están relevando o tienen proyectado relevar. En este contexto, la coordinación entre los distintos entes es de crucial importancia para optimizar los recursos presupuestarios, materiales y humanos, así como para reducir el tiempo necesario para que los datos estén disponibles o para obtener resultados de las investigaciones.

Con el propósito de fomentar esta optimización y/o coordinación entre los distintos entes, presentaré una serie de propuestas que incluyen la aplicación de nuevos desarrollos tecnológicos e inno-

5 Creado por la Ley 21673.

6 Creada por el Decreto 995/91.

7 Creado por el Decreto-Ley 1291/58; el Decreto-Ley 1291/58 fue abrogado por el art. 54 del Decreto 1661/1996. La función descrita corresponde al art. 3, inc. a) de este último.

vadores, como el big data, el *blockchain* y la inteligencia artificial, tendientes a una gestión más efectiva y eficaz de los datos y la información.

2. Los datos de investigación

Inicialmente, formularé una definición de “dato de investigación”, considerando que no existe una única definición y que esta puede variar según las disciplinas o comunidades que buscan definirlo.

En un primer acercamiento, vale mencionar que la Real Academia Española (RAE) (s.f.-a) define “dato” como la “información sobre algo concreto que permite su conocimiento exacto o sirve para deducir las consecuencias derivadas de un hecho”. Por su parte, la Comisión Económica para América Latina y el Caribe (CEPAL), más específicamente, manifiesta que “... los datos de investigación son datos que son recolectados, observados o creados para ser analizados y producir resultados de investigación originales” (CEPAL Bibliología, 2020).

En nuestro ordenamiento jurídico, de acuerdo con lo expresado por el art. 3 de la Ley 26899 de Repositorios Digitales Institucionales de Acceso Abierto,⁸ se entiende por “dato de investigación o dato primario” a “... todo dato en bruto sobre los que se basa cualquier investigación y que puede o no ser publicado cuando se comunica un avance científico pero que son los que fundamentan un nuevo conocimiento”. Esta definición, establecida por la normativa vigente, será la base conceptual adoptada en el presente documento para referirnos a los datos de investigación.

Existen otras aproximaciones en la conceptualización, como la definición del Portal de Información de Ciencia y Tecnología Argentino,⁹ que describe la “ciencia abierta” como

los esfuerzos de investigadores, gobiernos, agencias de promoción de la investigación y el desarrollo de la misma comunidad científica,

8 Ley de Repositorios digitales institucionales de acceso abierto.

9 Portal del Ministerio de Ciencia, Tecnología e Innovación: <https://datos.mincyt.gob.ar/#/>.

para hacer accesibles públicamente, en formato digital, sin o con la mínima restricción, los resultados de investigaciones financiadas por el sector público —publicaciones y datos de investigación— como un medio de contribuir al avance de la ciencia y la tecnología.

No obstante, para el propósito de este trabajo, se adoptará exclusivamente la definición de “datos de investigación” establecida por la Ley 26899, ya que esta ofrece un marco normativo claro y adecuado al contexto del análisis. Según dicha ley, se entienden como “datos en bruto sobre los que se basa cualquier investigación y que fundamentan un nuevo conocimiento” (artículo 3).

Sumado a lo anterior, y como concepto lingüístico oficial, el término “dato”, según la RAE en su primera acepción, significa: “Información sobre algo concreto que permite su conocimiento exacto o sirve para deducir las consecuencias derivadas de un hecho”.

En esta primera parte, solo resta una mención a los distintos tipos de datos distinguidos por su origen (National Science Board, 2005):

- Datos observacionales, primarios o sin procesar: son aquellos datos que han sido recolectados en algún proyecto, programa o iniciativa de investigación, pero aún no han sido procesados o analizados.
- Datos procesados o computacionales: son datos que han sido digitalizados, traducidos, transcritos, limpiados, validados, verificados y/o anonimizados.
- Datos analizados o derivados: son los datos transformados en modelos, gráficos, tablas, textos u otros, que han sido creados a partir de los datos primarios y procesados, cuyo objetivo es el descubrimiento de información útil, la presentación de conclusiones y la toma de decisiones.

2.1 La obtención de los datos

Habiendo abordado una conceptualización de los datos de investigación y sus distintos tipos, es necesario describir el modo en que se obtienen los datos, destacando la importancia de que los investigadores planifiquen la recopilación, conservación, descripción y difusión de sus datos desde el inicio de su actividad investigadora.

La mejor manera de registrar los datos es mediante un Plan de Gestión de Datos, que proporciona una estructura adecuada para la administración de los datos de investigación, facilitando su posterior procesamiento y/o análisis.

Los datos se pueden obtener mediante los siguientes métodos:

- Mediciones directas.
- Observación activa y pasiva.
- Estudios de casos.
- Encuestas.

En la actualidad, los métodos descritos pueden ser llevados a cabo por el investigador o, en algunos casos, mediante sistemas autónomos, estén o no conectados a un controlador humano.

La investigación marina y espacial son ejemplos en los que la obtención de datos puede realizarse mediante sistemas autónomos, como satélites, ROV submarinos¹⁰ y *Wave Gliders*,¹¹ entre otros. Mientras un investigador enmarcado en un proyecto de investigación realiza mediciones u observaciones en un momento específico del día, mes, o año, un sistema autónomo puede realizar mediciones de manera continua, siempre que su sistema de energía o de propulsión lo permita. En algunos casos, existen sistemas autónomos cuya propulsión no depende de una fuente externa, lo que les permite recolectar datos de forma indefinida. Un caso de estos últimos es el sistema autónomo ya mencionado *Wave Glider*, de medición de parámetros oceanográficos, inventado por la empresa Liquid Robo-

10 Remotely Operated Vehicle es un robot submarino no tripulado conectado a un barco en la superficie por medio de un cable largo.

11 Los *Wave Gliders* son robots oceánicos de larga duración que ayudan a científicos, empresas y Gobiernos a obtener nuevos conocimientos y mejorar la toma de decisiones. Operando en la superficie, estos vehículos autónomos proporcionan el vínculo esencial entre el mar, el aire y el espacio. Amplían la forma en que se recopilan datos meteorológicos y marinos, reduciendo una brecha de datos y acceso que antes eran inaccesibles; además, pueden transformar los sensores submarinos en fuentes de información en tiempo real. Diseñado y fabricado por Liquid Robotics, es el vehículo de superficie autónomo más probado del mundo. Ver en: <https://www.boeing.com/defense/autonomous-systems/wave-glider/>.

tics,¹² que toma datos las 24 horas del día los 7 días de la semana debido a que se abastece de energía solar y de las olas debido a su diseño y realiza el procesamiento de los datos *in situ*, a bordo. También es importante recordar un ejemplo emblemático de sistemas autónomos, autonomía y toma de datos: la sonda espacial robótica Voyager 1, lanzada el 5 de septiembre de 1977, que aún sigue operativa. Actualmente, la sonda, ubicada a más de 14,000 millones de millas de la Tierra, sigue enviando información, esta vez desde el espacio interestelar, ya que abandonó nuestro sistema solar alrededor del año 2013.¹³

2.2 La cantidad de datos

A modo de introducción, cito las palabras de Bill Gerhardt, Kate Griffin, y Roland Klemann,¹⁴ quienes en el año 2012 manifestaron que el problema en la actualidad es que el aluvión cada vez mayor de información (de terabytes a petabytes y exabytes) amenaza con sumergirnos en una marea de información sin filtrar, sin procesar, no estructurada y aparentemente inmanejable. Tal y como están las cosas en este momento, el ecosistema de datos está muy fragmentado. Entre los agentes que crean los datos y los que potencialmente podrían extraer valor de ellos, se encuentra un laberinto lleno de complejidad, desigualdad y falta de comunicación. Si el análisis tiene que ser la nueva “refinería”,¹⁵ deberá abordarse parte de esta fragmentación con mayor conectividad, confiabilidad y eficiencia (Gerhardt et al., 2012). Con respecto a la desigualdad y una solución a esta,

serán cruciales las infraestructuras de transporte y de almacenamiento de los Big Data. Los agentes que puedan ver el panorama en toda su extensión (a través de las redes y los Data Centers, recopilando y controlando los datos en movimiento), estarán en la mejor

12 Ver en: www.liquid-robotics.com/wave-glider/overview/.

13 Ver: www.voyager.jpl.nasa.gov.

14 Miembros del Grupo de Soluciones Empresariales para Internet de CISCO.

15 El término “refinería” alude a que, a los datos en el texto citado, se los considera el petróleo del siglo XXI.

posición para conectar a los diferentes agentes de los Big Data y crear un todo que funcione de manera sincronizada. (Gerhardt et al, 2012, p. 3)

A modo de ilustrar la magnitud de la cantidad de datos, y llevando el ejemplo a un extremo para dimensionarlo, valen las palabras de Nick Skytland (2012), del Johnson Space Center, quien afirmó:

En el tiempo que le tomó leer esta oración, la NASA¹⁶ reunió aproximadamente 1,73 gigabytes de datos de nuestras casi 100 misiones actualmente activas. Hacemos esto cada hora, todos los días, todos los años, y la tasa de recolección está creciendo exponencialmente.¹⁷

Como podemos observar, la primera y más compleja situación que surge es la cantidad de datos generados a partir de las distintas actividades de investigación, la cual se ha incrementado a medida que los sistemas se perfeccionan tecnológicamente y, como se describió anteriormente, se dotan de autonomía, almacenamiento, transmisión y autosuficiencia.

Para dimensionar la generación y obtención de datos en investigaciones científicas, resulta interesante analizar la Ley de Moore en su actual “declive”, en pos de su superación conceptual y fáctica por otras leyes. Esta ley proviene de una predicción realizada alrededor del año 1965 por Gordon Moore, cofundador de Intel Corp., quien sostenía que la cantidad de transistores en un chip se duplica aproximadamente cada dos años. Esto también implicaba que el rendimiento de esos chips, y de las computadoras que alimentaban, aumentaría de manera sustancial en el mismo lapso. En otras palabras, la computación experimentaría un crecimiento exponencial en términos de potencia, mientras que los costos relativos disminuirían drásticamente (INTEL, 2017). Actualmente, la Ley de Moore se en-

16 Siglas en inglés para Administración Nacional de Aeronáutica y el Espacio de los Estados Unidos.

17 Ver en: <https://skytland.medium.com/what-is-nasa-doing-with-big-data-a6c748588588>.

cuentra en decadencia, e incluso algunos sostienen que ha llegado a su fin, dado que los fabricantes de chips han alcanzado los límites de la fabricación de circuitos a escala atómica y las restricciones impuestas por la física de los electrones. Sin embargo, ha surgido una nueva ley, conocida como Ley de Huang, en honor al director ejecutivo y cofundador de Nvidia Corp., Jensen Huang, quien describe cómo los chips de silicio que impulsan la inteligencia artificial duplican su rendimiento más de una vez cada dos años. Si bien este incremento se puede atribuir tanto al *hardware* como al *software*, su progreso constante lo convierte en un habilitador clave para tecnologías que van desde automóviles, camiones y barcos autónomos hasta sistemas de reconocimiento facial, de voz y de objetos en nuestros dispositivos personales (Mims, 2020). Ante la generación masiva de datos, es necesario introducir un término íntimamente ligado a la obtención de grandes flujos de información: el big data. Este concepto permite luego abordar el avance en el procesamiento de dichos datos, que, en términos conceptuales, se refleja en la transición de la Ley de Moore a la Ley de Huang.

No abordaré un tema que requeriría de un análisis individual y pormenorizado, pero que revolucionará aún más la temática tratada en el presente documento: la computación cuántica.¹⁸ Aunque aún se encuentra en una etapa incipiente de desarrollo, representa un nivel superlativo de procesamiento de datos que permitirá resolver problemas para los cuales todavía no se ha encontrado solución.

2.3 El big data

En línea con los conceptos mencionados anteriormente, no existen definiciones consensuadas para muchos de los nuevos términos tecnológicos, por lo que formularé una conceptualización propia.

En 1998, John Mashey, informático teórico estadounidense, publicó un artículo titulado “Big Data and the Next Wave of Infrestress”, en el cual introdujo el término “big data”. Este concepto se aplica a conjuntos de datos cuyo tamaño excede la capacidad de las herramientas de *software* habitualmente utilizadas para capturar,

18 Ver Microsoft (s.f.) y Habla el mercado (2021).

administrar y procesar dicha información en un período de tiempo razonable. Posteriormente, Joe Hellerstein (2008)¹⁹ señaló que nos encontrábamos frente a una “revolución industrial de los datos”, caracterizada por un incremento exponencial en la cantidad y diversidad de datos digitales en tiempo real, producto del mayor uso de equipos tecnológicos con una creciente capacidad para obtener y procesar datos e información (DEFonline, 2018).

Más recientemente, el Banco Interamericano de Desarrollo (BID) ha señalado que el término “big data” proviene originalmente del ámbito de las ciencias de la computación y se ha utilizado típicamente para referirse a conjuntos de datos cuyo tamaño excede la capacidad del *software* y *hardware* estándar disponible para capturar, almacenar y analizar dicha información.

Continuando con el desarrollo conceptual presentado por el BID, es importante introducir el concepto de las “tres V” como las características que definen qué es el big data:

- V1: volumen, referido a la enorme cantidad de datos existentes. El volumen se asocia a los recursos requeridos por los datos, tanto de recolección y almacenamiento como de capacidad de cómputo.
- V2: velocidad en la que estos datos se producen y son analizados. Hace referencia al ritmo en que estos se crean, procesan, analizan y almacenan. Vale recordar lo mencionado al inicio del documento sobre los sistemas autónomos, plataformas y dispositivos que facilitan la creación y/o el intercambio de información, generando grandes volúmenes de datos que deben ser almacenados en tiempo real.
- V3: variedad de fuentes y tipos de datos. Estos últimos corresponden a la estructura de los datos, vale decir, estructurados, semiestructurados e inestructurados (Banco Interamericano de Desarrollo, 2017).

A las 3V se les podrían sumar dos más, de acuerdo con las nuevas tendencias:

19 Director de Estrategia de Trifecta Wrangler y profesor de la Universidad de Berkeley, Estados Unidos.

- V4: veracidad, que hace referencia a fuentes de información confiables para su análisis.
- V5: valor, en el sentido de creación de nuevas oportunidades de desarrollo.

Por los conceptos expuestos y para los fines del presente documento, el big data se refiere a todos los datos que se recaban de manera continua en alguna de las múltiples investigaciones científicas llevadas a cabo por instituciones, organismos y/o investigadores cuyos proyectos, programas o iniciativas cuentan con financiamiento público.

2.4 El acceso abierto a los datos

Habiendo descrito la gran cantidad de datos que, en la actualidad, puede generar un proyecto, programa o iniciativa de investigación financiada con fondos públicos, es importante destacar la importancia del acceso abierto (AA) a los datos de investigación a los que nos estamos refiriendo.

La Organización para la Cooperación y el Desarrollo Económicos (OECD, por sus siglas en inglés) en la reunión de París de 2004, mediante la *Declaration on Access to Research Data from Public Funding*, se pronunció en favor del acceso abierto de los datos de investigación provenientes de investigaciones financiadas con fondos públicos, y se comprometió a trabajar en un documento común de recomendaciones o normativas acordadas entre los diversos miembros, lo cual se concretó en el documento *Principles and Guidelines for Access to Research Data from Public Funding*, que se hizo público en 2007 (Organización para la Cooperación y el Desarrollo Económicos, 2007). Esa serie de principios y recomendaciones persiguen el fin de facilitar el acceso eficiente y compartido a los conjuntos de datos provenientes de las actividades investigativas financiadas por los países miembros (Angelozi, 2020).

En el mismo sentido, en Argentina, la Ley 26899 de Creación de Repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos establece que los proyectos financiados con fondos públicos que generen datos primarios deberán contener un plan de

gestión de datos, de modo de asegurar su accesibilidad a través de los repositorios destinados a su conservación. En estos repositorios, los datos deberán ser depositados en un plazo no mayor a cinco años a partir de su recolección.

La Ley citada precedentemente es reglamentada por la Resolución 753/16 del entonces Ministerio de Ciencia, Tecnología e Innovación. La norma implementa un reglamento operativo, cuyo objetivo es establecer los lineamientos básicos que deben cumplir los organismos e instituciones públicas que componen SNCTI. En el reglamento, además, se pautaron los plazos de adecuación a la ley para las diferentes categorías de instituciones que conforman el sistema público de ciencia y tecnología del país.

Para las instituciones involucradas, uno de los principales desafíos consiste en generar y mantener archivos donde los datos de investigación puedan ser correctamente almacenados, identificados y preservados para que estén plenamente accesibles. Otro punto crucial es el involucramiento de los científicos y su cooperación en sistematizar sus planes de gestión de datos, volcarlos posteriormente en los repositorios y otorgar el consentimiento para compartirlos (Angelozzi, 2020).

Resumiendo, y con la definición que nos brinda la reglamentación en la Resolución 753/16, podemos señalar que se denomina “acceso abierto” a los datos, la producción científico-tecnológica y los textos completos de los artículos científicos, los cuales se pueden, de forma gratuita y accesible, leer, descargar, copiar, distribuir, imprimir, buscar y usar con propósitos legítimos ligados a la investigación científica, la educación, la gestión de políticas públicas, al desarrollo tecnológico y la innovación, sin otras barreras económicas, legales o técnicas que las que suponga internet en sí misma. La única condición que plantea este modelo para la reproducción y distribución de las obras puestas a disposición es la obligación de otorgarles a los autores el control sobre la integridad de su trabajo, así como el derecho a ser adecuadamente reconocidos y citados. Es decir que el AA, como modelo, pone el acento en la cuestión de la accesibilidad a la producción científico-tecnológica para fines determinados, tales como la investigación o la educación, lo cual no implica necesariamente su uso libre o indiscriminado (Ministerio de Ciencia y Tecnología, 2024).

Actualmente, la Ley 26899 y su reglamentación tienen un aca-

tamiento muy bajo, por no decir nulo, lo cual nos debe llevar a reflexionar sobre la necesidad de implementar medidas que fomenten el cumplimiento de la normativa. Este documento busca realizar un acercamiento con propuestas de implementación, utilizando tecnologías actuales para tal fin.

Para finalizar este capítulo, y luego de las distintas conceptualizaciones tanto doctrinarias como normativas, es necesario señalar que los datos, por sí mismos, no constituyen una obra en los términos en los que ésta es reconocida por la legislación. Según Cabanellas, el concepto de obra implica una creación, una adición al acervo intelectual (Cabanellas de las Cuevas, 2023, tomo I, p. 585). En este sentido, y para los fines del presente documento, tomo como “obra” al conjunto de datos a los cuales el investigador les ha dado un sentido, desarrollando una investigación que recaba información aislada y la transforma en series de datos significativos. Estos datos reflejan la intelectualidad que el investigador les atribuye con el objetivo de que constituyan un insumo relevante para su investigación. En la recopilación de información y datos para convertirlos en series de datos, o incluso en una base de datos, independientemente de su magnitud, considero que reside el aspecto clave que el investigador toma en cuenta al momento de exteriorizar y mostrar aquello que está estudiando. Los datos aportan la consistencia y las pruebas necesarias para sustentar lo que el investigador afirma o, a través de las mediciones, permiten determinar el comportamiento del objeto de estudio. Esa exteriorización, esa expresión, al decir de Cabanellas de las Cuevas (2023),

implica no solamente exteriorización —o sea una relación externa respecto del fuero interno del creador—, sino también una posibilidad comunicativa, una potencialidad para que personas distintas al autor tomen conocimiento y perciban la obra en cuestión, o sea la expresión en que consiste. (p. 586)

Con respecto al acatamiento de la Ley 26899, y de acuerdo con la información oficial publicada en la página web del Sistema Nacional de Repositorios Digitales,²⁰ existen 530668 publicaciones deposita-

20 Ver en: <https://repositoriosdigitales.mincyt.gov.ar/vufind/>.

das en 48 repositorios pertenecientes a 48 instituciones. Estos números abarcan las más diversas temáticas de investigación.

Para efectuar una mensura sobre los números descritos, resulta interesante lo expresado en el portal del Repositorio Institucional de la Universidad Nacional de La Plata. En dicha web, se muestra en el cuadro de “Buscar material”: “Busque entre los 160.403 recursos disponibles en el repositorio”. Si consideramos que la web del Sistema Nacional indica la existencia de 530668 publicaciones —que incluyen artículos, tesis doctorales, documentos de conferencias, tesis de grado, capítulos de libros, tesis de maestría, libros, informes técnicos, reseñas de artículos, conjuntos de datos, documentos de trabajo, entre otros— y lo comparamos con el repositorio de la Universidad Nacional de La Plata (UNLP),²¹ que es uno de los repositorios del Sistema Nacional y contiene ítems de búsqueda similares, obtenemos la siguiente ecuación: el repositorio de la UNLP representa el 30% de las publicaciones del Sistema Nacional. Indudablemente, los números no son consistentes, lo que permite inferir que, si la inclusión en el Sistema Nacional es obligatoria debido a que el articulado de la Ley 26899 indica “deberán depositar”, esta norma no se estaría cumpliendo debidamente. Es decir que, si el contenido de publicaciones y recursos de todas las demás instituciones que conforman el Sistema Nacional de Ciencia y Tecnología estuviera incluido en el Sistema Nacional de Repositorios, el número total debería ser mucho mayor. Asimismo, el de la Universidad Nacional de La Plata (que, por ser su repositorio propio, también infero que incluye la totalidad de las publicaciones de la UNLP) debería representar una proporción significativamente menor al 30%.

Es evidente que la norma que obliga a depositar no se está cumpliendo, o que, de cumplirse, el sistema no está reflejando la totalidad de la producción científico-tecnológica resultante del trabajo, la formación y/o los proyectos financiados, total o parcialmente, con fondos públicos de sus investigadores, tecnólogos, docentes, becarios de posdoctorado y estudiantes de maestría y doctorado pertenecientes a las instituciones del Sistema Nacional de Ciencia y Tecnología.

21 Ver en: <https://sedici.unlp.edu.ar/>.

2.5 Breve recapitulación del apartado 2

Lo manifestado en este apartado pone de manifiesto la relevancia de los datos de investigación como pilares fundamentales en la construcción de nuevo conocimiento y su impacto en la innovación y el desarrollo tecnológico. La transición tecnológica, ilustrada por el paso de la Ley de Moore a la Ley de Huang, junto con la capacidad de los sistemas autónomos para generar grandes volúmenes de datos, plantea desafíos en cuanto a su gestión, almacenamiento y análisis. En este contexto, el concepto de big data se presenta como una herramienta esencial para transformar estos datos masivos en insumos significativos para la investigación.

Por otro lado, la Ley 26899 y su reglamentación subrayan la necesidad de garantizar el acceso abierto a los datos financiados con fondos públicos. No obstante, el bajo nivel de cumplimiento evidencia una brecha entre el marco normativo y su aplicación efectiva, lo que subraya la importancia de desarrollar repositorios accesibles y fomentar el uso de planes de gestión de datos. Estas medidas son fundamentales para promover la transparencia, maximizar el impacto de los datos científicos y avanzar hacia un modelo de ciencia abierta que beneficie a la sociedad en su conjunto.

3. El procesamiento y gestión

Si bien la gestión de datos, tal como se describe a continuación, es una planificación que debe llevarse a cabo al inicio de la investigación, consideré necesario realizar previamente una descripción de las clases de datos, su gran volumen y su calidad de accesibilidad. Ahora, corresponde continuar con la gestión de los datos mencionados.

De manera enunciativa, resulta importante describir el ciclo de los datos científicos (Universidad de Chile. (s.f.):

1. Planificación; el investigador sienta las bases de la investigación y de la gestión de los datos que ha planificado obtener.
2. La recolección y posterior procesamiento de los datos.²²

22 En esta instancia es donde el investigador, o el instrumental autónomo, toma el dato; por ej., una boya hidrográfica que releva la temperatura y salinidad del agua o que mide las corrientes.

3. El almacenamiento y preservación de los datos, que, tal como regula la Ley 26899, deberá realizarse en los repositorios de las instituciones del SNCTI.
4. La reutilización. Este punto será retomado al final, ya que la reutilización requiere cumplir una serie de etapas, que serán parte de la propuesta y conclusión final.

Como he manifestado, en la actualidad, frente a cualquier proyecto, programa o iniciativa con carácter de investigación científica, se recopilan cantidades gigantescas de datos de investigación, considerados la nueva divisa de la era digital. Desde sonetos hasta datos estadísticos, y desde genes hasta datos geoespaciales, la cantidad de material que se crea y almacena crece de manera exponencial (Leaders Activating Research Networks, 2017).

La cantidad de datos de investigación es recopilada por investigadores de diversas disciplinas y distintas instituciones, lo que añade un elemento de complejidad a la búsqueda de una gestión eficiente²³ y eficaz²⁴ de estos datos.

Ya se ha mencionado que la gran mayoría de las instituciones de investigación pública están dentro del SNCTI, las cuales, dado que forman parte de los distintos programas, proyectos e iniciativas, generan datos.²⁵ Aquí comienza a cobrar mayor relevancia el marco legal de Argentina, en particular la Ley 26899, mencionada anteriormente. Dicha ley establece en su artículo 2º que los organismos e instituciones públicos comprendidos en su artículo 1º deberán implementar políticas institucionales para la gestión y preservación a largo plazo de los datos que obtengan.

23 “Eficiencia: Capacidad de disponer de alguien o de algo para conseguir un efecto determinado” (Real Academia Española, s.f.-b).

24 “Eficacia: Capacidad de lograr el efecto que se desea o se espera” (Real Academia Española, s.f.-c).

25 Como mencioné anteriormente, los datos pueden ser generados por los propios investigadores o por sistemas autónomos de relevamiento, no significando diferencia a los efectos del presente trabajo.

3.1 Plan de gestión de datos

Un plan de gestión de datos (PGD) es un documento formal elaborado por el investigador o grupo de investigación al inicio de un proyecto de investigación. Este documento describe todos los aspectos relacionados con la gestión de los datos, es decir, qué se hará con ellos durante y después del proyecto.

El PGD tiene como propósito proporcionar un análisis de los principales elementos de una política de gestión de datos y detallar el tratamiento que recibirán los conjuntos de datos recopilados o generados en el curso del proyecto. Su finalidad es servir como una herramienta de apoyo durante el ciclo de vida de los datos que se recopilen, procesen o generen durante el proyecto (CEPAL Bibliología, 2020).

En el marco del proyecto Leaders Activating Research Networks (LEARN) (2017), se elaboró un documento modelo de políticas de gestión de datos donde los ítems a considerar son:

- a. Jurisdicción de aplicación.
- b. Gestión de derechos de propiedad intelectual de los investigadores, institución de filiación, agencias de financiamiento.;
- c. Cuestiones operativas relativas al archivo de los datos y gestión de los repositorios (integridad, identificación, trazabilidad, accesibilidad, interoperabilidad, tiempo de resguardo, normativas de descripción y citación).
- d. Responsabilidades, tareas y derechos de los investigadores: respeto por las políticas institucionales, elaboración del plan de gestión de datos, posibilitar la reutilización de los datos.
- e. Responsabilidades, tareas y derechos de la institución: entre otras, proveer capacitación e infraestructura para archivo y acceso a los datos y soporte permanente.

Los ítems mencionados deben ser coordinados para lograr una gestión eficiente de los datos. De lo contrario, estos no estarán disponibles en condiciones óptimas para ser utilizados por quienes requieran esa información, o, aunque estén disponibles, carecerán de alguna de las características previamente mencionadas.

3.2 Breve recapitulación del apartado 3

El procesamiento y la gestión de datos científicos representan elementos fundamentales en la investigación moderna, especialmente ante el crecimiento exponencial en la cantidad y diversidad de datos generados. La implementación de un ciclo de vida de los datos que abarque desde su planificación, recolección y procesamiento hasta su almacenamiento, preservación y reutilización resulta esencial para garantizar su accesibilidad y utilidad. En este marco, el cumplimiento de normativas como la Ley 26899 en Argentina, que exige políticas institucionales para la gestión de datos en las instituciones del SNCTI, adquiere una relevancia crucial.

Asimismo, la elaboración de Planes de Gestión de Datos (PGD) bien estructurados permite abordar aspectos clave, como los derechos de propiedad intelectual, la interoperabilidad de los repositorios y las responsabilidades de los investigadores e instituciones. Sin embargo, para enfrentar los desafíos actuales, es imprescindible integrar tecnologías emergentes como el *blockchain* y la inteligencia artificial, que ofrecen soluciones innovadoras para mejorar la trazabilidad, integridad y accesibilidad de los datos. Solo mediante la consolidación de estas prácticas y herramientas será posible maximizar el impacto de los datos científicos en el avance del conocimiento y el desarrollo tecnológico.

4. La financiación pública

Como se ha venido desarrollando, los datos a los que me refero, y que forman parte de la problemática abordada, son aquellos provenientes de investigaciones científicas realizadas con fondos públicos. Estas investigaciones pueden ser llevadas a cabo por instituciones estatales, por instituciones con participación estatal o por entes privados que, en el marco de un programa, proyecto u otra iniciativa, hayan recibido fondos públicos para investigar.

El uso de fondos públicos implica que los datos resultantes de la investigación financiada poseen ciertas características distintivas en comparación con aquellos financiados con recursos privados. Pero ¿a qué nos referimos con “fondos públicos”? Básicamente, se trata de los recursos financieros que un organismo del Estado destina a un fin determinado, que en este caso es la investigación científica.

Es evidente que, al existir un financiamiento público, este condiciona la investigación en cuestión, otorgándole ciertas particularidades, como las establecidas por la Ley 26899:

Los organismos e instituciones públicas que componen el SNCTI, conforme lo prevé la ley 25.467, y que reciben financiamiento del Estado Nacional, deberán desarrollar repositorios digitales institucionales de acceso abierto, propios o compartidos, en los que se depositará la producción científico-tecnológica resultante del trabajo, formación y/o proyectos, financiados total o parcialmente con fondos públicos, de sus investigadores, tecnólogos, docentes, becarios de posdoctorado y estudiantes de maestría y doctorado.

A los efectos de especificar el concepto de financiamiento público, resulta pertinente citar la Resolución 753 E/16,²⁶ la cual en la sección correspondiente a “Definiciones Generales” establece que, en el marco de la Ley 26899, se entiende por “financiamiento total o parcial con fondos públicos” la inversión que el Estado realiza, ya sea de forma directa (como el pago de salarios, incentivos, subsidios, etc.) o indirecta (como el acceso a bibliografía científica adquirida con fondos públicos, la utilización de la infraestructura de las instituciones y organismos, el uso de insumos y equipamientos, el financiamiento total o parcial de viajes, etc.) para el desarrollo de la actividad científica, tecnológica y de innovación.

Aquí se identifican dos características que, a través de la normativa, se transforman en obligaciones tanto para determinadas instituciones como para quienes realizan investigaciones financiadas con fondos públicos:

- a. Las instituciones del SNCTI que reciben fondos públicos²⁷ deben desarrollar repositorios digitales. Se consideran repositorios digitales aquellas colecciones de la producción científ

26 La Resolución 753 E/16 emana de la Secretaría de Articulación Científico Tecnológica del Ministerio de Ciencia, Tecnología e Innovación, que reglamenta la Ley 26899 (10 de noviembre de 2016).

27 La totalidad de las instituciones que forman parte del SNCTI reciben de una u otra manera fondos públicos, sean instituciones autárquicas o autónomas. El presupuesto proviene del Estado nacional, entre otras fuentes de financiamiento.

fico-tecnológica de una institución que permiten la búsqueda y recuperación de información para su posterior uso, tanto a nivel nacional como internacional. Incorporan mecanismos para importar, identificar, almacenar, preservar, recuperar y exportar un conjunto de objetos digitales, normalmente accesibles desde un portal web y que están descritos mediante metadatos que facilitan su recuperación. Además, son abiertos e interactivos, ya que cumplen con protocolos internacionales que garantizan la interoperabilidad entre ellos.

- b. Los investigadores deben depositar en los repositorios digitales la producción resultante de sus investigaciones.

Agrego una tercera característica que, aunque no se desprende directamente de la normativa, existe en la actualidad y tendrá una mayor incidencia en el futuro cercano: la toma o generación de datos mediante instrumentos autónomos, ya definidos previamente, tales como los satélites y los ROV submarinos, los UAV,²⁸ boyas autónomas, estaciones meteorológicas, etc., que si bien en algunos casos requieren el manejo de humanos, la toma la realizan mediante sensores de manera autónoma. Este tipo de instrumentos están en un ascendente desarrollo, volviéndose cada vez más usuales y, me atrevería a decir, cada vez más necesarios.

Podría darse el caso de que los puntos a) y b) mencionados anteriormente no se cumplan, ya sea porque la institución aún no ha constituido su repositorio digital o porque el investigador no ha depositado el resultado de su investigación. Esto, sumado a la tercera característica, genera situaciones en las que se produce una inmensa cantidad de datos que no están ni estarán disponibles de manera óptima para quienes los requieran, tema sobre el cual profundizaré más adelante.

Recapitulando, nos encontramos a esta altura con una gran cantidad de datos provenientes de investigaciones llevadas a cabo por distintos organismos del Estado y financiadas con fondos públicos. Estos datos deberían estar alojados en repositorios digitales y, en conjunto, podrían calificarse como una big data.

28 UAV (Unmanned Aerial Vehicle, por sus siglas en inglés): vehículo aéreo no tripulado (VANT).

4.1 Breve recapitulación del apartado 4

Sobre este apartado, puedo concluir que la financiación pública de investigaciones científicas genera un compromiso intrínseco con la transparencia y accesibilidad de los datos obtenidos. La Ley 26899 establece la obligatoriedad de desarrollar repositorios digitales que garanticen el acceso abierto a los datos producidos con fondos públicos, fortaleciendo así la ciencia abierta y promoviendo el uso óptimo de estos recursos por parte de la comunidad científica y la sociedad.

No obstante, el cumplimiento parcial de estas disposiciones, ya sea por la falta de repositorios adecuados o por el incumplimiento en el depósito de datos por los investigadores, evidencia la necesidad de reforzar políticas y mecanismos de implementación. Adicionalmente, el auge de instrumentos autónomos para la generación de datos plantea nuevos desafíos para su gestión, almacenamiento e interoperabilidad. En este contexto, consolidar un sistema eficiente de repositorios digitales resulta clave para aprovechar plenamente el potencial de los datos como insumo estratégico en el desarrollo científico y tecnológico financiado con recursos públicos.

5. La propiedad intelectual en el contexto

En este apartado abordaré la situación en el ámbito de la propiedad intelectual vinculada a los datos y bases de datos obtenidos en el marco de las investigaciones financiadas con fondos públicos, en el camino hacia una propuesta de optimización de la gestión de datos. Excluyo del análisis, como manifesté anteriormente, las investigaciones que contengan datos personales, clínicos o sensibles que requieran algún tipo de confidencialidad, o cualquier tipo de datos que requieran la aplicación de regulaciones sobre datos personales, ya que esto debería ser objeto de un análisis específico, fuera del alcance del presente documento.

El sistema de propiedad intelectual, representado en Argentina en el caso que nos ocupa por la Ley 11723 de propiedad intelectual, es un sistema de incentivos para la producción y publicación de obras. Busca estimular la creación al ofrecer beneficios y resguardo legal a los autores, otorgando protección, control y beneficios económicos, lo que motiva la producción continua de obras y su distribución

pública en beneficio tanto de los creadores como de la sociedad en general o del área en el cual se fomenta dicha creación, como puede ser el campo científico.

En el ámbito científico, los sistemas de incentivos a la producción funcionan de un modo totalmente diferente, ya que los investigadores y docentes del sistema público cuentan con fondos del presupuesto nacional para realizar estas investigaciones, y el propio sistema los condiciona para su publicación. Los científicos no viven del sistema de derechos de autor, sino del salario y de los fondos que instituciones de investigación, universidades, etc. destinan a su trabajo. Incluso con un sistema de propiedad intelectual estricto, la mayoría de los investigadores no reciben pago alguno por la publicación de sus trabajos (Busaniche, 2013).

El Dr. Alejandro Ceccatto,²⁹ al referirse a la Ley 26899, señaló: “El objetivo es que la producción científica financiada por la sociedad sea accesible. Es inaceptable que, si el Estado nacional financia la investigación de una persona, después no pueda la sociedad toda acceder a ese conocimiento”.³⁰

En la región, solo Perú contaba con una legislación de este estilo: la Ley 30035.³¹ Otros países como Brasil han venido discutiendo desde hace tiempo, pero sin mayores avances en materia legislativa. En los Estados Unidos, de acuerdo con el *Nelson Memo*³² del 25 de agosto de 2022, las investigaciones médicas financiadas con fondos del Gobierno federal son de acceso abierto, y todo lo producido por las agencias federales es de dominio público (un ejemplo interesante es la producción científica de la NASA).

La Unión Europea ha avanzado con políticas aisladas, aunque su discurso público es de promoción del acceso abierto: “Los estados miembros de la UE tienen cierta libertad para definir sus propias políticas de Acceso Abierto” (Oficina de Conocimiento Abierto, 2018).

29 Doctor en física, expresidente del CONICET, exsecretario de Articulación Científico Tecnológica de la Nación Argentina.

30 Ver en: <https://santafe.conicet.gov.ar/es-ley-el-acceso-libre-a-la-informacion-cientifica/>.

31 Ley 30035, disponible en: https://www2.congreso.gob.pe/Sicr/TraDocEst-Proc/Expvirt_2011.nsf/Repexpvirt?OpenForm&Db=201101188&View.

32 Nelson Memo, disponible en: <https://www.whitehouse.gov/wp-content/up->

En el Reino Unido,³³ Alemania o los Países Bajos, donde la publicación académica es una industria importante, los Gobiernos están negociando con los editores para avanzar hacia el acceso abierto inmediato después de la publicación. Por lo tanto, estos países del norte de Europa han asignado un presupuesto adicional para sostener la transición hacia un modelo Gold OA. Éste se refiere al acceso gratuito en línea a los resultados de la investigación, como artículos de revistas y libros. El contenido OA (*open access*) está abierto a todos, sin cuotas de acceso. Existen dos vías principales para que los resultados de la investigación sean accesibles en acceso abierto. Una consiste en publicar artículos o libros por la vía del acceso abierto en la plataforma de un editor (lo que se conoce como “acceso abierto dorado” o por sus siglas en inglés Gold OA) (Springer Nature, 2024). La otra, que se evidencia en el sur de Europa (Italia, España), prefiere los modelos Green OA y no asigna fondos de investigación adicionales ni ofrece variaciones en las duraciones del período de bloqueo (Oficina de Conocimiento Abierto, 2018). El Green OA consiste en archivar una versión del manuscrito en un repositorio de acceso abierto (“acceso abierto verde” o Green OA, por sus siglas en inglés) (Springer Nature, 2024).

En Francia, muchas agencias de financiamiento, como la Agencia Nacional para la Investigación (ANR), alientan el acceso abierto, pero aún no existe una obligación formal. La mayoría de los institutos de investigación y universidades francesas llegaron a un acuerdo en 2013:³⁴ promover el uso de archivos abiertos, como la plataforma HAL (Hyper Articles en Ligne) desarrollada por el (Centro Nacional Francés para la Investigación Científica (CNRS, por sus siglas en francés). Además, en virtud del artículo 30 de la Ley 2016-1321 de Francia, titulada Loi pour une République Numérique, los investigadores pueden negociar el derecho de publicación de OA con los editores tan pronto como el 50% de su trabajo sea financiado con fondos públicos.³⁵

loads/2022/08/08-2022-OSTP-Public-access-Memo.pdf.

33 Reino Unido, al momento del artículo citado, formaba parte de la Unión Europea.

34 Ver en: <https://hal.science/>.

35 Ver en: <https://oca.unc.edu.ar/2018/10/23/acceso-abierto-en-europa/>.

Argentina se posiciona como uno de los primeros países en establecer una política pública generalizada para toda la investigación financiada con fondos públicos, cumpliendo así con la vocación de servicio público de la ciencia financiada por el Estado nacional en beneficio de la sociedad en su conjunto (Busaniche, 2013).

Enfocándome en el objeto de estudio del presente trabajo, consideraré como insumo los datos, en cualquier formato, que se obtengan y el procesamiento de dichos datos hasta el momento en que sean volcados en una publicación científica por un investigador. En ese punto, la autoría y pertenencia ya son manifiestas y los datos han sido procesados en el marco del proyecto de estudio.

En esta instancia vale la cita de la ley 11723, Régimen Legal de la Propiedad Intelectual, que sigue vigente hoy pese a los avances tecnológicos, porque desde su origen en 1933 fue diseñada para adaptarse a nuevos formatos y necesidades en el ámbito de la propiedad intelectual. Su marco inicial, que respondía a innovaciones como la industria cinematográfica y la reproducción técnica de obras, estableció principios amplios que han permitido su actualización continua, como la inclusión de bases de datos y obras digitales.³⁶

Dicha ley manifiesta en su art. 1 que

A los efectos de la presente Ley, las obras científicas, literarias y artísticas comprenden los escritos de toda naturaleza y extensión, entre ellos los programas de computación fuente y objeto; las compilaciones de datos o de otros materiales; las obras dramáticas, composiciones musicales, dramático-musicales; las cinematográficas, coreográficas y pantomímicas; las obras de dibujo, pintura, escultura, arquitectura; modelos y obras de arte o ciencia aplicadas al comercio o a la industria; los impresos, planos y mapas; los plásticos, fotografías, grabados y fonogramas, en fin, toda producción científica, literaria, artística o didáctica sea cual fuere el procedimiento de reproducción (...).

De lo expresado por la norma citada, podemos resaltar los siguientes términos: “obras científicas”, “compilaciones de datos o de

36 Cfr. Lacquaniti (2017).

otros materiales”, “modelos”, “planos y mapas”, “fotografías, grabados y fonogramas” y, por último, lo que engloba a las demás, “toda producción científica...”. Los datos provenientes de la investigación científica a los que me refiero estarían comprendidos, al menos, en uno de los términos destacados en la norma.

Con respecto a los datos obtenidos, en cualquier estado, llegará un momento en que el investigador podría experimentar dudas y desconfianza ante la posibilidad real de una disociación entre los datos y el investigador, o entre el proyecto de investigación y los datos. También podría temer que los datos, que tanto esfuerzo le costó adquirir, carezcan de su adecuada identificación como el generador de dicha información, lo que podría redundar en una retención excesiva o, incluso, en una apropiación indefinida de estos. A ello se suma la desconfianza respecto a la veracidad o exactitud de determinados datos o mediciones, especialmente en referencia a otros investigadores que requieran utilizar los datos obtenidos por terceros.

Para analizar lo expresado anteriormente, y como enlace con el siguiente apartado, tomaremos una investigación de parámetros oceanográficos como caso de estudio para describir las distintas situaciones planteadas en este documento.

Supongamos que, en el marco de un proyecto sobre la modelización de la corriente circumpolar antártica y la medición de sus parámetros condicionantes del clima, se planea realizar distintas mediciones en el ámbito geográfico de dicha corriente. Este proyecto es elaborado por un investigador en el contexto de un programa o proyecto mayor o como parte de alguna línea de investigación perteneciente a una institución, que podría ser, a modo de ejemplo, el CADIC,³⁷ el IAA³⁸ (ambas instituciones integrantes del SNCTI) alguna otra institución que esté investigando esta temática.

El proyecto incluirá, entre otras cosas, un plan de trabajo orientado a obtener los datos necesarios para llevar a cabo la modelización

37 Centro Austral de Investigaciones Científicas del Consejo Nacional de Investigaciones Científicas y Técnicas, el cual se encuentra en la ciudad de Ushuaia, provincia de Tierra del Fuego, Antártida e Islas del Atlántico Sur.

38 Instituto Antártico Argentino, dependiente de la Dirección Nacional del Antártico del Ministerio de Relaciones Exteriores, Comercio Internacional, y Culto de la Nación.

de la corriente y determinar los parámetros que condicionan el clima. Para dichas mediciones, será necesario embarcarse en una campaña oceanográfica³⁹ o utilizar algún instrumento capaz de realizar dichas mediciones para recopilar los datos requeridos.

No siendo el objeto principal de este trabajo, es importante señalar, al menos en el caso que nos ocupa, la complejidad y el alto costo que implica llevar a cabo una campaña oceanográfica. Por esta razón, cuando se realiza una, se suelen incluir distintos proyectos relacionados con el área geográfica o la temática en cuestión.

Usualmente, no existe vinculación entre los distintos proyectos o investigadores, salvo que se trate de una campaña enfocada específicamente en un tema particular. En consecuencia, se medirán diversos parámetros y se recopilarán infinidad de datos de manera independiente entre los distintos investigadores.

En el caso del ejemplo, una vez que el buque haya zarpado y llegado a la zona de medición, se procederá a la recopilación de datos, mediciones y/o muestras, utilizando distintos instrumentos o los sensores del buque. Por ejemplo, los datos podrían ser recolectados mediante un CTD.⁴⁰ Dicho instrumental tomará, por ejemplo, la temperatura del agua, lo que significa que, al finalizar la campaña, se dispondrá de una gran cantidad de datos de temperatura obtenidos a distintas profundidades y en diversas zonas geográficas. Estos datos estarán en poder del investigador responsable, pero es posible que otro investigador, ubicado en el otro extremo del buque, en otro lugar o incluso en otra institución esté realizando, proyecte realizar o necesite realizar mediciones similares o simplemente encuentre útiles los datos recopilados. Sin embargo, ese investigador nunca sabrá de la existencia de estos datos, al menos hasta que sean publicados o depositados en repositorios, una situación que, en muchos casos,

39 Las campañas oceanográficas son la base de la investigación científica en el mar. Su objetivo es recabar datos y obtener muestras que, luego de ser procesadas en los laboratorios en tierra, serán clasificadas y utilizadas para elaborar distintos tipos de modelos. Ver en: <https://www.pampazul.gob.ar/investigacion-y-desarrollo/campanas-de-investigacion/>.

40 Instrumento de oceanografía que se utiliza para medir la conductividad, la temperatura y la presión del agua de mar (la “D” significa *deep*, “profundidad”, que está estrechamente relacionada con la presión).

no sucede en el tiempo óptimo o incluso podría no ocurrir nunca.

Imaginemos el caso anterior, pero multiplicado por los investigadores y las instituciones que integran el SINCYT. Para ejemplificarlo gráficamente, me centraré únicamente en lo concerniente al CONICET y su red de instituciones, lo que adquiere una relevancia especial, ya que, a través de organismos como este, la Agencia I+D+i,⁴¹ universidades nacionales, centros de investigación públicos y privados y laboratorios, el sistema moviliza una notable infraestructura científica y recursos humanos altamente especializados.

Argentina cuenta con más de cien institutos de investigación y miles de investigadores y tecnólogos distribuidos a lo largo de todo el territorio nacional. Esta red permite abordar un amplio espectro de áreas de conocimiento, desde las ciencias básicas hasta la innovación tecnológica.

Para finalizar con el ejemplo citado, el investigador utilizará esos datos para su proyecto y/o su publicación. Sin embargo, hasta que dicha publicación no sea realizada, nadie podrá acceder a esos datos, a pesar de haber sido recopilados con financiamiento público.⁴²

En cada una de las instituciones a lo largo y ancho de la República Argentina, y solo por mencionar aquellas vinculadas al CONICET (Figura 1), probablemente esté ocurriendo que investigadores de distintas disciplinas recopilen datos relevantes para sus áreas, utilizando instrumental, herramientas e infraestructura del Estado nacional, además de recibir financiamiento por parte de este.

41 Organismo nacional descentralizado con autarquía administrativa y funcional, dependiente de la Secretaría de Innovación, Ciencia y Tecnología. Disponible en: <https://www.argentina.gob.ar/ciencia/agencia/la-agencia>.

42 En el caso que describo, tanto el buque como el instrumental, la institución y los salarios son abonados por el Estado nacional.

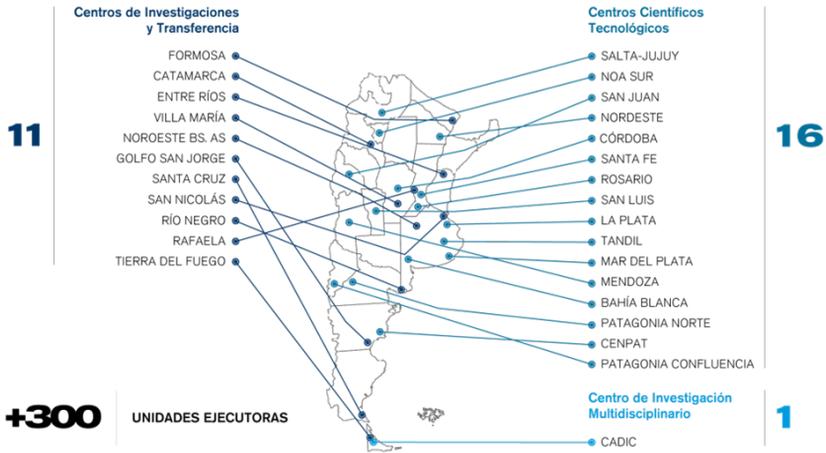


Figura 1. Red Institucional, CONICET.⁴³

En este punto, es importante destacar que la titularidad de dichos datos, de bases de datos y de publicaciones posteriores no está clara ni se pretende cuestionar. Sin embargo, es fundamental asegurar que estos datos no puedan ser modificados ni vulnerada su eventual propiedad intelectual. Aquí surge la primera propuesta en torno a este tema: la utilización de una tecnología prácticamente nueva y disruptiva: el *blockchain*.

5.1 El *blockchain*

En esencia, el *blockchain*, o cadena de bloques, es un registro de información abierto que permite anotar transacciones y realizar su seguimiento. Este registro se intercambia y se verifica en una red entre particulares. Junto con otras tecnologías de registros distribuidos, crea un sistema fiable y transparente al permitir que múltiples partes participen en la transacción y verifiquen, de antemano, los datos que se anotarán en el registro, sin que ninguna de ellas tenga la posibilidad de modificar las entradas posteriormente.

Cada transacción o “bloque” se transmite a todos los participan-

43 Ver en: <http://red.conicet.gov.ar/organizacion/>.

tes de la red y debe ser verificada por cada “nodo” participante mediante la resolución de un complejo problema matemático. Una vez verificado, el bloque se añade al registro o cadena. Aunque el concepto tradicional de la cadena de bloques consiste en una red abierta y anónima, también existen cadenas “privadas” que preseleccionan los nodos autorizados para administrar el registro.

Si bien la asociación del *blockchain* con el mundo financiero y las criptomonedas es inmediata, la tecnología se ha vuelto rápidamente atractiva y ya se utiliza para seguir la evolución de productos en las cadenas de suministro. Esto resulta particularmente interesante para muchos sectores que hacen un uso intensivo de la propiedad intelectual, como las industrias farmacéutica, automovilística, de lujo y de bienes de consumo, donde la trazabilidad de los productos es fundamental.

Estas cadenas pueden contener diferentes tipos de datos, que van desde información relativa a criptomonedas, transacciones y contratos hasta documentos de datos, fotos, videos y diseños (Clark, 2018).

Retomando el ejemplo de investigación oceanográfica descrito anteriormente, relacionado con la toma de datos para la modelización de la corriente circumpolar antártica, y en el contexto del complejo marco de recursos materiales y humanos implicados, financiados y/o pertenecientes al Estado nacional, resulta necesario plantear la primera propuesta en la línea de optimizar la gestión de datos de investigación financiados con recursos públicos.

En este marco, surge la necesidad de vincular los datos obtenidos y las bases de datos donde se alojan con la tecnología *blockchain*.

La vinculación entre los datos, las bases de datos y el *blockchain* se da en el siguiente marco: el *blockchain* permite trazar una base de datos de gran volumen y distribuirla en decenas, miles o incluso millones de nodos, creando una gran red P2P descentralizada. Se trata de un libro de registro inmutable que contiene la historia completa de todas las transacciones ejecutadas en la red (Cagnodata Consulting, 2019).

En esencia, una cadena de bloques pública es un registro público mundial, independiente de toda autoridad y control central. Este registro especial solo contiene una lista de transacciones de la crip-

tomoneda subyacente que alimenta y asegura la propia red de la cadena de bloques (Barulli, 2021).⁴⁴

El registro merece el adjetivo “público” porque cualquier persona puede obtener una copia actualizada, cualquiera puede añadir algo al registro enviando una nueva transacción a la red, pero nadie puede hacer cambios ni borrar nada del registro (Cagnodata Consulting, 2019).

En este sentido, la combinación de una cadena de bloques pública con un repositorio de datos versionados y cifrados permitirá crear una plataforma digital de trazabilidad de la propiedad intelectual con la que se podrá demostrar fácilmente la existencia, la integridad y la propiedad de documentos, composiciones artísticas, datos de investigación, diseños, programas informáticos, planes de negocio y contratos, entre otros activos. Cada objeto digital recibe un sello temporal permanente, se archiva y se vincula a su propietario de forma inmutable (Barulli, 2021).

Asimismo, los conceptos de integridad y propiedad enmarcan dos de las características más importantes que un investigador considera al momento de preservar los datos obtenidos.

Volviendo al ejemplo oceanográfico, y mediante este sistema, el investigador estaría certificando la integridad de los datos y su procedencia. Un aspecto crucial a destacar, tratándose del *blockchain*, es que el registro será inmutable, lo que podría representar una solución para la incertidumbre de los investigadores y sus instituciones a la hora de compartir los datos tal como fueron recopilados.

Una vez que el dato se haya obtenido y sometido a una cadena de bloques, contará con su veracidad, integridad y la identificación del compilador, sin posibilidad de ser modificado. A partir de ese momento, la asociación entre el investigador, los datos recopilados y el *blockchain* será indisoluble.

Solo restará determinar en qué momento de la investigación los datos serán registrados mediante una cadena de bloques. La inclusión o constitución podría darse, inicialmente, en el propio instrumento, es decir, que el instrumental científico, a través del geoposiciona-

44 Disponible en: https://www.wipo.int/wipo_magazine_digital/es/2021/article_0002.html.

miento y una firma digital tanto del aparato como del investigador usuario, inicie la cadena que contendrá la serie de datos, quedando esta disponible para cualquier otro investigador que desee utilizarla.

Lo relevante es que quien utilice esa cadena tendrá la absoluta certeza de la información que contiene, eliminando la necesidad de efectuar una certificación, validación o incluso una nueva medición, lo que implica un ahorro significativo de recursos económicos y humanos.

Cada conjunto de datos registrado en *blockchain* puede vincularse a la firma digital de la persona responsable de su captura, lo que ofrece dos beneficios: por un lado, asegura la autoría y, por otro, certifica que el origen y la validez de los datos son auténticos. La firma digital actúa como una capa adicional de seguridad, garantizando que solo las personas autorizadas puedan verificar, actualizar o certificar los datos en el *blockchain*. De ahora en más, quien quiera utilizar en un proyecto, por ejemplo, los datos de salinidad recabados por el investigador que se embarcó para realizar la medición en el mar, solo tiene que obtener el *blockchain* que contenga dicha serie de datos.

Se podr[ía resumir en lo siguiente:

1) Investigador inicial: es quien recabó el dato y procedió a certificar mediante *blockchain*:

- a. La integridad e inmutabilidad de su dato o serie de datos.
- b. Su identificación como compilador o curador de los datos.
- c. La veracidad de sus datos.
- d. Toda información concerniente al instrumento con que tomó la medición.⁴⁵

2) Investigador requirente: es quien requiere esos datos para una investigación:

- a. Ahorra tiempo y recursos financieros.
- b. No le está permitido manipular la integridad de los datos, y no puede alterar la cadena.
- c. Debe respetar la autoría de los datos que requiera.
- d. Confía en los datos que utiliza por tener toda la información de estos a su alcance.

45 Por ejemplo, la calibración, tan necesaria a la hora de determinar la calidad del dato obtenido.

- 3) Repositorio digital: lugar donde se alojan los datos:
 - a. Tienen datos certificados.
 - b. Tienen los datos con autoría y pertenencia institucional.
 - c. Pueden ponerlos a disposición sin ninguna traba, ya que la información está toda contenida y es inmutable; donde va la cadena, va la información.

En la descripción anterior, también pueden entrar las instituciones que requieran dichos datos, lo cual haría crecer el ahorro en recursos financieros y tiempo.

Hemos llegado a un punto en el cual es necesario una recapitulación para pasar al último paso de la propuesta de gestión de datos.

Como hemos visto, en el marco del SNCTI se llevan a cabo distintos proyectos, programas e iniciativas en los que constantemente se recopilan datos de diversa índole, temática, rubro o sector. Estos datos, al ser financiados con fondos públicos, deberían estar disponibles y ser de acceso público desde el inicio, tal como lo establece la Ley 26899 mencionada anteriormente.

Este sistema permite registrar y rastrear de manera inmutable el ciclo de vida completo de los datos, desde su generación inicial hasta su uso en diversas aplicaciones y trabajos posteriores (De Filippi y Wright, 2018). Mediante la tecnología *blockchain*, crearemos un registro transparente de todas las interacciones con los datos, garantizando que cada transformación, análisis o transferencia quede adecuadamente registrado. De este modo, se garantiza no solo la integridad y autenticidad de los datos, sino también la posibilidad de auditar su recorrido y utilización. Esto resulta especialmente relevante en contextos donde la colaboración interdisciplinaria o interinstitucional requiere un alto nivel de confianza en los datos compartidos (Tapscott y Tapscott, 2016). Incorporar esta trazabilidad fortalece la reproducibilidad, fomenta la transparencia y asegura que los datos sean utilizados de manera ética y eficiente, maximizando su valor en diferentes etapas del trabajo científico o técnico.

Con el *blockchain*, tal como lo he descrito y explicado, he buscado arrojar luz sobre la integridad, veracidad y compilación de los datos. Sin embargo, si un ingeniero agrónomo de un instituto en Tucumán necesita conocer la incidencia climática en el cultivo del

limón y el impacto del cambio climático en la producción de dicho fruto, ¿cómo podría saber que existe una serie de datos que evidencian la influencia de la corriente circumpolar antártica en el clima en general? Y, específicamente, ¿cómo ese cambio climático podría incidir en los cultivos?

Es aquí donde corresponde abordar el último tramo y el concepto final: la inteligencia artificial.

5.2 La inteligencia artificial

Como he desarrollado a lo largo del presente trabajo, considero necesaria una conceptualización introductoria sobre la inteligencia artificial (IA) antes de abordar su incidencia directa en la optimización de la gestión de datos en el marco de la investigación científica financiada con fondos públicos.

La IA suele referirse al aumento de la independencia, velocidad y magnitud en la adopción informática y automatizada de decisiones. No es un concepto único, sino más bien una “constelación” de procesos y tecnologías que les permiten a las computadoras complementar o reemplazar tareas específicas que, de otro modo, serían ejecutadas por seres humanos, como tomar decisiones y resolver problemas.

En el núcleo de la IA están los algoritmos: códigos informáticos diseñados y escritos por seres humanos que ejecutan instrucciones para traducir datos en conclusiones, información o productos. Desde hace tiempo, los algoritmos son fundamentales para el funcionamiento de los sistemas cotidianos de comunicación e infraestructura. El enorme volumen de datos generado por la vida moderna y la necesidad de analizarlos han impulsado el desarrollo de la IA.

El sector privado entiende los datos desde esta perspectiva: cuantos más datos estén disponibles para alimentar los algoritmos y mayor sea la calidad de esos datos, más poderosos y precisos serán los resultados que dichos algoritmos pueden generar. Los sistemas algorítmicos pueden analizar rápidamente grandes volúmenes de datos y así hacer que los programas de IA puedan ejecutar funciones de adopción de decisiones que antes eran dominio de seres humanos que actuaban sin instrumentos informáticos (Organización de las Naciones Unidas, 2018).

Retomando el ejemplo de investigación oceánica citado anteriormente, resulta interesante destacar cómo, en la actualidad, se está implementando de forma paulatina la acción conjunta y combinada entre el big data y la IA.

En las investigaciones actuales, el investigador adopta un rol cada vez más pasivo frente a los distintos sistemas de toma de datos autónomos, debido a la inmensa cantidad de datos que estos pueden captar durante periodos de tiempo mucho más prolongados. También en el análisis “las máquinas” y su evolución tecnológica cada vez más avanzada están tomando la posta.

Tom Zimmerman,⁴⁶ en el marco de sus investigaciones, nos llama a imaginarnos la capacidad de agrupar la protección del medio ambiente de origen. Un mundo en el que pudiéramos observar cómo se mueven organismos diminutos como el plancton, las diminutas criaturas que suministran dos tercios del oxígeno que respiramos. Trabajando juntos, hemos inventado un sistema de realidad microscópica que les permite a los científicos ver la vida en una escala en la que un cabello humano es tan grande como un árbol, que se puede combinar con IA para detectar el comportamiento del plancton y predecir la salud del medio ambiente.⁴⁷

Stella Loiacono, CTO⁴⁸ de IBM Argentina, manifiesta que, en cinco años, pequeños robots microscópicos con inteligencia artificial, conectados en red a la nube y desplegados en todo el mundo, serán capaces de monitorear continuamente el agua de los océanos. Esto permitiría, entre otras cosas, estudiar el comportamiento del plancton en tiempo real sin necesidad de recolectar muestras y llevarlas a laboratorio.⁴⁹

De lo expuesto, podemos afirmar que esos cinco años ya están ocurriendo. Actualmente, existe un ecosistema de instrumentos y sensores que recopilan datos de manera continua, ya sea de forma automática o autónoma (con o sin supervisión humana, respectivamente).

46 Inventor maestro en IBM Research – Almaden, que explora las fronteras de la interacción hombre-máquina y la detección ambiental.

47 Ver en: <https://research.ibm.com/blog/microscopic-reality-ai-microscopes>.

48 Chief Technology Officer.

49 Ver en: <https://www.lanacion.com.ar/tecnologia/5-innovaciones-que-nos-cambiaran-la-vida-en-el-proximo-lustro-segun-ibm-nid2118883/>.

La constitución del big data está actualmente en constante crecimiento. Solo resta que el desarrollo de algoritmos y redes neuronales converja en la creación de una IA específica para los objetivos buscados. Para abordar este último punto, es importante señalar que otra forma de afrontar el desarrollo de herramientas inteligentes es mediante la llamada IA basada en datos, también conocida como IA subsimbólica, aprendizaje automático, aprendizaje de máquina o, en inglés, *machine learning* (ML). Esta abarca las tecnologías inteligentes que aprenden con algoritmos que se entrenan con base en los datos suministrados. En su evolución, estos algoritmos adquieren la capacidad de responder frente a situaciones nuevas, de forma tal que, a partir de estas tecnologías, resulta posible descubrir patrones de comportamiento en esos datos (Corvalán, 2021).

Uno de los tipos más utilizados de algoritmos de *machine learning* es aquel que simula el comportamiento del cerebro humano. En nuestro cerebro, se activan y se disparan los reconocedores de patrones. Por ejemplo, al observar un animal, la detección de un hocico activa una neurona que “avisa” a las demás que podríamos estar viendo un perro. Si, además del hocico, observamos plumas, se inhibe el reconocimiento inicial, y el cerebro necesita procesar más información antes de decidir si realmente es un perro. Este proceso, simplificado de manera extrema, ocurre en una fracción de segundo.

Las neuronas en nuestro cerebro pueden “dispararse” o no. Cuando alcanzan un nivel específico de estímulo, conocido como “umbral de excitabilidad”, la neurona transmite información a todas las demás neuronas a las que está conectada. De esta forma, el estímulo recibido es procesado y transmitido a través de distintas células. En resumen, este proceso explica cómo reconocemos patrones de información para resolver problemas o alcanzar objetivos.

Los modelos de ML, denominados redes neuronales artificiales (RNA), se inspiran en el proceso descrito. En las últimas dos décadas, los avances de la industria 4.0, que abreva interconectividad y la producción vertiginosa de datos, ha contribuido a la expansión actual de este tipo de IA (Corvalán, 2021).

A esta altura, ya contaríamos con una posible solución para la identificación de datos o información dentro del big data obtenido: la implementación de una IA mediante la constitución de redes neu-

ronales artificiales capaces de detectar la información necesaria que un investigador o institución requiera para su utilización.

5.3 Breve recapitulación del apartado 5

En este apartado, he planteado los principales desafíos relacionados con la gestión de datos científicos obtenidos en investigaciones financiadas con fondos públicos, destacando la necesidad de garantizar la integridad, trazabilidad y acceso abierto de dichos datos. A lo largo del análisis, se han identificado limitaciones estructurales y operativas en el manejo de información científica, así como las oportunidades que surgen con la implementación de tecnologías innovadoras, como el *blockchain* y la inteligencia artificial.

En este marco, la utilización de *blockchain* ofrece una solución concreta para asegurar la inmutabilidad, autoría y disponibilidad de los datos, permitiendo una gestión más eficiente y transparente. Asimismo, el desarrollo de inteligencia artificial, particularmente mediante técnicas de *machine learning* y redes neuronales artificiales, plantea un camino hacia la optimización en la identificación y análisis de grandes volúmenes de datos.

Ambas herramientas, integradas en un modelo de gestión, permitirían no solo preservar la calidad y validez de los datos obtenidos, sino también promover un acceso más equitativo al conocimiento, alineado con los principios establecidos en la Ley 26899. En consecuencia, considero que estas propuestas representan un paso necesario hacia un sistema más eficiente, colaborativo y orientado al beneficio de la sociedad en su conjunto.

6. Hacia una propuesta. La situación actual

A partir de todo lo expresado, y en referencia al ejemplo citado anteriormente, se podría plantear la situación de la siguiente manera, con el objetivo de proponer una línea de acción orientada a optimizar la gestión de los datos provenientes de investigaciones financiadas con fondos públicos.

6.1 Esquema tradicional de la investigación y sus resultados

- a. Un investigador presenta un proyecto de investigación de la corriente circumpolar antártica para determinar, mediante la medición de sus parámetros, la incidencia en el clima.
- b. Se embarca en una campaña oceanográfica, en la cual tomará datos mediante instrumental del buque e instrumental autónomo que lanza al mar.
- c. Obtiene infinidad de mediciones de temperatura, salinidad, oxígeno disuelto, etcétera.
- d. Retiene la información para la presentación de sus resultados, publicaciones, *papers*, etcétera.
- e. En el mejor de los casos el resultado es alojado en el repositorio digital, quedando allí con la referencia de los metadatos.

Si alguien quisiera ver o utilizar los datos relevados, debería identificar el repositorio en el cual se encuentren y esperar el tiempo que le lleve al investigador publicarlos; posiblemente tendrá acceso a los datos con algún tipo de procesamiento. Este esquema ha resultado muy oneroso e ineficaz, ya que:

- i. Se abona el salario del investigador.
- ii. Se financia el proyecto.
- iii. Se financia la campaña oceanográfica.
- iv. Se financia el instrumental.
- v. Se financia la publicación (si el presupuesto lo permite).
- vi. Los datos se guardan durante un tiempo excesivo, transcurso en el cual nadie puede utilizarlos. Cualquier investigador que quiera usarlos en un proyecto, estudio o publicación debe repetir el proceso o realizar una verdadera arqueología de información con un resultado incierto y una procedencia dudosa de los datos que obtenga.

6.2 La confidencialidad de la información

Antes de abordar las propuestas y conclusiones, es importante citar los aspectos relacionados con la confidencialidad de la información. Incluyo este tema como el último apartado debido al orden lógico de los pasos en la investigación que tomo como ejemplo.

Los datos, series de datos o información accesoria que se obtienen no requieren, en principio, un tratamiento de confidencialidad, al menos en su génesis. A continuación, procederé a explicar esta afirmación.

La información obtenida, compuesta por datos, series de datos, etc., en su generación no se considerará secreta. Sin embargo, y complejizando la situación, surge una cuestión: la investigación oceanográfica del ejemplo podría estar revelando algún parámetro marino que requiera confidencialidad. Por ejemplo, la salinidad del agua en ciertas latitudes. La velocidad del sonido se ve afectada por variables oceanográficas como la temperatura, la salinidad y la presión (Campo Varela et al., 2023). Esto adquiere relevancia en el contexto del comportamiento del sonar de los submarinos, ya que, en ciertos casos y zonas, estas variables constituyen información clasificada como secreta para la defensa.

Por ello, al presentar los proyectos de investigación, deberían informarse las variables que serán relevadas. En tales casos, dichas variables podrían considerarse confidenciales y requerir una evaluación posterior para decidir su eventual publicación o disposición al público.

Asimismo, la propuesta que desarrollo a continuación sería eficiente, ya que todas las variables podrían ser identificadas incluso después de haber sido ingresadas en una base de datos, independientemente de cuál sea. Lo mismo aplicaría a otros tipos de datos que se pretenden obtener, como aquellos relacionados con la pesca industrial o cuestiones geológicas que podrían derivar en información de interés para el sector petrolero.

En todos los casos mencionados, sería necesario establecer qué datos revisten carácter confidencial y cuáles no, o si la investigación en cuestión no debería ser considerada confidencial. Esto no implica que dicha práctica sea habitual, sino que describe un ideal basado en las disposiciones legales y regulatorias.

Finalmente, en todos los casos, el reclamante de la confidencialidad será el Estado nacional como principal financiador de la investigación.

6.3 Breve recapitulación del apartado 6

A lo largo de este apartado, se ha puesto de manifiesto la necesidad de revisar y optimizar los esquemas actuales de gestión de datos provenientes de investigaciones financiadas con fondos públicos. El modelo tradicional, caracterizado por la fragmentación y el acceso limitado a los datos, no solo representa un uso ineficiente de los recursos, sino que además retrasa el progreso científico al dificultar la reutilización de información clave.

En este contexto, se ha planteado la importancia de definir criterios claros en torno a la confidencialidad de la información, particularmente cuando esta puede implicar cuestiones de seguridad nacional o interés estratégico para el país. Esto resalta la necesidad de un sistema que permita identificar y gestionar variables sensibles, sin comprometer la accesibilidad y transparencia que caracterizan a la ciencia abierta.

Por tanto, se propone avanzar hacia un modelo más eficiente y equitativo que contemple tanto la disposición de los datos como las excepciones necesarias. Un sistema que, bajo la tutela del Estado nacional como principal financiador, asegure un balance entre el libre acceso a la información científica y la protección de los intereses nacionales. Con estas bases, se podrá garantizar que los datos recolectados por los investigadores no solo contribuyan a sus fines originales, sino que también se conviertan en un recurso valioso para la comunidad científica y otros sectores estratégicos.

7. Propuesta de optimización

Para optimizar el proceso, es fundamental partir de datos certificados y validados, es decir, datos confiables, lo cual se ha logrado mediante la implementación de la tecnología *blockchain*. Esta tecnología permite registrar datos, series de datos o cualquier información relevante, garantizando su certificación por parte de su originador.

Es importante destacar que toda la investigación y los elementos que la componen, incluida la recolección de datos, son financiados con fondos públicos. Esto implica que, en caso de generarse propiedad intelectual, la titularidad correspondería tanto a los investigadores como a las instituciones participantes. En este contexto, la

certificación de la titularidad adquiere una relevancia esencial, especialmente si se establece desde el origen, ya que esto facilita futuras aplicaciones en investigaciones o proyectos que puedan estar sujetos a protección de derechos de propiedad intelectual.

Tomando en cuenta el mismo caso anterior, las propuestas de optimización serían las siguientes.

Parte 1. Investigadores A

Presenta un proyecto de investigación de la corriente circumpolar antártica para determinar, mediante la medición de sus parámetros, la incidencia en el clima.

Se embarca en una campaña oceanográfica, en la cual tomará datos mediante instrumental del buque e instrumental autónomo que lanzará al mar.

Obtiene infinidad de mediciones de temperatura, salinidad, oxígeno disuelto, etc.

Propuesta 1: estos datos se certifican en origen, se los firma digitalmente,⁵⁰ se los geoposiciona, se constituyen los metadatos⁵¹ y se forma un *blockchain* con los datos o series de datos. La autoría de esa serie de datos, o incluso el dato propiamente dicho que será incluido en una publicación (aún en ciernes) se consolida y se blindo mediante la aplicación de la cadena de bloques; el investigador implicado y los datos ya serán un ente inseparable de ahora en más. En este caso, el *blockchain* actúa como medida tecnológica de protección.

50 He tomado el concepto de “firma digital” de acuerdo con el art. 2 de la Ley 25506: “Se entiende por firma digital al resultado de aplicar a un documento digital un procedimiento matemático que requiere información de exclusivo conocimiento del firmante, encontrándose ésta bajo su absoluto control. La firma digital debe ser susceptible de verificación por terceras partes, tal que dicha verificación simultáneamente permita identificar al firmante y detectar cualquier alteración del documento digital posterior a su firma.”

51 Descripción estandarizada de las características de un conjunto de datos. En el contexto del documento electrónico, cualquier tipo de información en forma electrónica asociada a los documentos electrónicos, de carácter instrumental e independiente de su contenido, destinada al conocimiento inmediato y automatizable de alguna de sus características, con la finalidad de garantizar la disponibilidad, el acceso, la conservación y la interoperabilidad del propio documento. Ver Real Academia Española (s.f.-d).

Concluyendo con esta propuesta, la firma digital y la implementación de *blockchain* están estrechamente relacionadas, ya que ambas tecnologías garantizan la integridad, autenticidad y trazabilidad de los datos. Firmar digitalmente una secuencia de datos genera un *hash* único que asegura que la información permanezca inalterada desde el momento de la firma, permitiendo verificar su autoría.

Al incorporar estos datos en *blockchain*, cada bloque enlaza su contenido firmado con los bloques anteriores, creando una cadena cronológica e inmutable. Cualquier alteración rompe la secuencia, evidenciando posibles manipulaciones. En conjunto, la firma digital autentica los datos en su origen, mientras que el *blockchain* los resguarda en un registro seguro e inalterable, lo que resulta clave para procesos que requieren alta seguridad, como investigaciones o transacciones.

El investigador A ya no necesita conservar los datos de manera independiente, ya que estos están registrados en el *blockchain* inmutable, el cual resguarda la integridad de los datos originales que el investigador haya decidido incluir para garantizar su incorruptibilidad. De este modo, la atribución del origen de los datos y su posible protección mediante derechos de propiedad intelectual sobre los desarrollos derivados de la investigación, como publicaciones o avances tecnológicos, quedan aseguradas. Cualquier intento de modificar los datos registrados alteraría el *blockchain* y evidenciaría el cambio, garantizando así la autenticidad y preservación del contenido original.

El dato o serie de datos almacenados en *blockchain* se integran al big data de datos oceánicos y pueden ubicarse en cualquier servidor, siempre que todo el proceso y los servidores involucrados empleen tecnología *blockchain*. De este modo, sin importar la ubicación, los datos serán identificables en cuanto a su origen, incluyendo quién los registró, dónde y cómo se obtuvieron.

Parte 2. Investigador B o cualquier persona o institución que requiera de los datos

El investigador presenta un proyecto de investigación sobre las variables que afectan la corrosión en las construcciones. Para ello, debe incorporar a su modelo variables climáticas de diversa índole. Considera que incluir valores relacionados con la corriente circumpolar

podría ayudarle a determinar de manera más precisa la variabilidad climática y las condiciones de cambio. Esto le permitiría identificar con qué frecuencia y de qué manera cambian las condiciones que impactan en las estructuras edilicias.

Propuesta 2: con las variables a medir y los parámetros necesarios, se constituye una red neuronal diseñada para identificar, tanto en campañas pasadas como futuras, datos relevantes relacionados con cuestiones oceánicas, los parámetros relevados que inciden en el clima y su posible impacto en las condiciones climáticas.

Es posible que el investigador B no necesite repetir ciertas acciones realizadas previamente por el investigador A, ya que contará con datos fiables y exactos de los parámetros requeridos. Además, no podrá alterar la atribución de origen de estos datos, ya que están protegidos de forma inalterable. De este modo, la identificación del investigador A como responsable original de los datos permanecerá fija en cualquier uso futuro, sin importar dónde o cómo se utilicen.

El investigador B podrá evaluar cómo afecta la dinámica oceánica y demás variables que haya considerado en los cambios del clima y las estructuras edilicias y podrá constituir su red neuronal a los efectos de que la IA “vaya a buscar” e identifique los datos idóneos, a fin de poder realizar los cálculos y poder determinar si los fuertes vientos, las bajas temperaturas y la dinámica del hielo marino antárticos conducen a patrones de corrientes que influyen fuertemente el clima de la Tierra, y esta modifica de tal manera el clima donde el investigador B esté llevando adelante su proyecto.

Para concluir, es fundamental abordar el tema del espacio de almacenamiento específico, que debería considerarse en la etapa final del proceso, es decir, cuando los datos alcanzan su ubicación definitiva y “salen” del ámbito directo del investigador, independientemente de la localización territorial de los servidores. Es importante destacar que el alojamiento de la información debería realizarse en sitios oficiales pertenecientes a instituciones de investigación públicas, constituyendo un requisito indispensable vinculado al financiamiento público.

Un caso a considerar es el de la Unión Europea, donde los particulares, las empresas y las organizaciones pueden utilizar, recopilar, almacenar, transferir o gestionar datos no personales, así como uti-

lizar centros de procesamiento de datos o servicios en la nube desde cualquier lugar dentro de la UE.⁵²

Con la salvaguarda de los aspectos mencionados anteriormente sobre una eventual confidencialidad o secreto, la circulación debería ser la norma respecto a la información, los datos, las series o incluso las bases de datos generadas como resultado de investigaciones financiadas con recursos del erario público.

7.1 Breve recapitulación del apartado 7

Luego de haber planteado una propuesta de optimización, y previo a la conclusión final del documento, es importante resaltar que la implementación de tecnologías como la firma digital y el *blockchain* representa una herramienta transformadora para garantizar la integridad, autenticidad y trazabilidad de los datos generados en investigaciones científicas, particularmente aquellas financiadas con recursos públicos. Estas tecnologías no solo aseguran que los datos se mantengan inalterados desde su origen, sino que también refuerzan la atribución de su autoría, facilitando su uso en investigaciones posteriores y promoviendo la transparencia en la gestión de recursos públicos.

En este contexto, el almacenamiento seguro en *blockchain*, combinado con la interoperabilidad en redes y sistemas, no solo blindo los datos frente a manipulaciones, sino que fomenta la creación de ecosistemas colaborativos, donde el acceso a datos confiables permite optimizar esfuerzos de investigación y promover avances tecnológicos. Asimismo, asegurar que el almacenamiento y la gestión de los datos se realicen en plataformas oficiales o públicas es un paso crucial para garantizar que los resultados de investigaciones con financiamiento público estén disponibles y protegidos, en beneficio del desarrollo científico y social.

Esta propuesta de optimización pone en relieve la importancia de adoptar medidas tecnológicas innovadoras que armonicen con los

52 Libre circulación de datos no personales. Disponible en: https://europa.eu/youreurope/business/running-business/developing-business/free-flow-non-personal-data/index_es.htm.

principios de ética y sostenibilidad, asegurando que el conocimiento generado sea accesible, seguro y plenamente atribuible a quienes lo producen.

8. Conclusión final

El presente trabajo ha permitido abordar de manera integral la importancia de la gestión de datos científicos generados en investigaciones financiadas con fondos públicos, enfatizando su rol estratégico en la construcción de nuevo conocimiento y su impacto en el desarrollo científico y tecnológico. En un contexto de constante evolución tecnológica, el uso de herramientas como el *blockchain*, el big data y la inteligencia artificial no solo representa una oportunidad para optimizar procesos, sino también una necesidad para asegurar la trazabilidad, integridad y acceso abierto a la información.

Se ha puesto de manifiesto que el esquema tradicional de gestión presenta fallas estructurales, como la fragmentación de datos, los extensos plazos de accesibilidad y un bajo nivel de cumplimiento de normativas fundamentales, como la Ley 26899. Estas deficiencias, además de comprometer la transparencia y la reutilización de los datos, redundan en una ineficiente asignación de recursos públicos y en el desaprovechamiento del potencial científico de la información recopilada.

En este sentido, las propuestas tecnológicas desarrolladas en el documento plantean soluciones concretas y viables. La implementación del *blockchain* se presenta como una herramienta disruptiva para garantizar la autoría, integridad e inmutabilidad de los datos desde su origen, eliminando incertidumbres sobre su procedencia y autenticidad. A esto se suma el concepto de big data, que permite estructurar y procesar grandes volúmenes de información bajo los principios de las “cinco V” (volumen, velocidad, variedad, veracidad y valor), maximizando su utilidad para investigaciones interdisciplinarias. Por su parte, la inteligencia artificial, mediante algoritmos de aprendizaje automático y redes neuronales artificiales, habilita la identificación y análisis de patrones en los datos, facilitando su aplicación en modelos predictivos y decisiones estratégicas.

Asimismo, se ha destacado la importancia de armonizar estos

avances con el marco normativo vigente, garantizando que los repositorios digitales cumplan con su objetivo de democratizar el acceso a los datos, al tiempo que consolidan la propiedad intelectual de los investigadores y sus instituciones. Este modelo de gestión optimizada debe además contemplar excepciones vinculadas a la confidencialidad de datos sensibles o estratégicos, permitiéndole al Estado nacional, como principal financiador, equilibrar los principios de la ciencia abierta con los intereses nacionales.

En conclusión, la implementación de estas tecnologías y estrategias no solo asegura una gestión más eficiente y transparente, sino que fortalece el vínculo entre la investigación científica y la sociedad, promoviendo el uso óptimo de los recursos públicos y consolidando a la ciencia como un motor de desarrollo equitativo y sostenible. Este enfoque, alineado con las exigencias actuales y las oportunidades tecnológicas, permite proyectar un sistema más robusto, dinámico y orientado al bien común, posicionando a la República Argentina como un referente en la gestión de datos en el ámbito científico.

Bibliografía

- Angelozzi, S. M. (2020). La gestión de datos de investigación en abierto: introducción al rol emergente para las bibliotecas universitarias y científicas argentinas. *Palabra Clave (La Plata)*, 9(2). <https://doi.org/10.24215/18539912e091>.
- Banco Interamericano de Desarrollo. (2017). *El uso de datos masivos y sus técnicas analíticas para el diseño e implementación de políticas públicas en Latinoamérica y el Caribe*. <https://publications.iadb.org/es/publications/spanish/viewer/El-uso-de-datos-masivos-y-sus-tecnicas-analiticas-para-el-diseno-e-implementacion-de-politicas-publi.pdf>.
- Barulli, M. (9 de febrero de 2021). La PI es un viaje: la cadena de bloques y el almacenamiento cifrado son sus mejores amigos. *Revista de la OMPI*. https://www.wipo.int/wipo_magazine_digital/es/2021/article_0002.html.
- Big data, la vida en datos*. (2018). DEF Revista (N° 122).
- Busaniche, B. (2013). *Es Ley el Acceso Abierto a investigación científica sostenida con fondos públicos en Argentina*. DIGITAL RIGHTS LAC. <https://digitalrightslac.derechosdigitales.org/es/es-ley-el-acceso-abierto-a-investigacion-cientifica-sostenida-con-fondos-publicos-en-argentina/>.
- Cabanellas de las Cuevas, G. (2023). *Compendio del Derecho de Autor*. Heliasta.

- Campo Varela, M., Rodríguez Rodríguez, I. y Rodríguez Rodríguez, J. V. (2023). *Conceptos Básicos de la Ciencia del Sonido en el Mar*. UMA Editorial. <https://monografias.uma.es/index.php/mumaed/catalog/book/47>.
- CEPAL Bibliografía. (2020). *Gestión de datos de investigación*. Bibliogúias - Biblioteca de la CEPAL. <https://biblioguias.cepal.org/gestion-de-datos-de-investigacion>.
- Clark, B. (2018). La tecnología de la cadena de bloques y el Derecho de propiedad intelectual: ¿una pareja perfecta en el criptoespacio? *Revista de la OMPI*. https://www.wipo.int/wipo_magazine/es/2018/01/article_0005.html.
- Cognodata Consulting. (2019). *El futuro del procesamiento de datos: blockchain y big data*. <https://www.cognodata.com/blog/futuro-procesamiento-datos-blockchain-big-data>.
- Corvalán, J. G. (2021). *Tratado de Inteligencia Artificial* (Tomo I). La Ley.
- De Filippi, P. y Wright, A. (2018). *Blockchain and the Law: The Rule of Code*. Harvard University Press.
- Emery, M. A. (2017). *Propiedad Intelectual Ley 11.723. Comentada, Anotada y Concordada con los Tratados Internacionales*. Astrea.
- Gerhardt, B., Griffin, K. y Klemann, R. (2012). *Descubrir el valor en el fragmentado mundo del análisis de los Big Data. Cómo los infomediarios de la información crearán un nuevo ecosistema de datos*. Grupo de Soluciones Empresariales para Internet (IBSG) de Cisco. https://www.cisco.com/c/dam/global/es_es/assets/executives/pdf/Unlocking_Value_in_Big_Data_Analytics.pdf.
- Habla el mercado. (30 de septiembre de 2021). Computación inspirada en física: un paso pragmático antes del ordenador cuántico. *MIT Technology Review*. <https://www.technologyreview.es/s/13703/computacion-inspirada-en-fisica-un-paso-pragmatico-antes-del-ordenador-cuantico>.
- Hellerstein, J. (2008). *The Commoditization of Massive Data Analysis*. <https://timshowers.com/2008/11/future-of-data-analysis/>.
- INTEL. (2017). *Intel Technology Innovations*. <https://www.intel.la/content/www/xl/es/innovation/leadership/overview.html>.
- Lacquaniti, L. G. (2017). La Ley de propiedad intelectual de 1933. Proyectos y debates parlamentarios sobre los derechos autorales en Argentina. *Revista Estudios Sociales Contemporáneos*, (17), 69-87. <https://ri.conicet.gov.ar/handle/11336/49560>.
- Leaders Activating Research Networks. (2017). *Toolkit of Best Practice for Research Data Management*. <http://learn-rdm.eu/wp-content/uploads/RDMToolkit.pdf>.
- Microsoft. (s.f.). ¿Qué es la computación cuántica? <https://azure.microsoft.com/es-es/overview/what-is-quantum-computing/>.
- Mashey, J. (1998). *Big Data and the Next Wave of Infostress*. https://www.usenix.org/legacy/publications/library/proceedings/usenix99/invited_talks/mashey.pdf.
- Mims, C. (19 de septiembre de 2020). Huang's Law Is the New Moore's Law, and Explains Why Nvidia Wants Arm. *The Wall Street Journal*. <https://www.wsj.com/articles/huangs-law-is-the-new-moores-law-and-explains-why-nvidia-wants-arm-11600488001>.

- Ministerio de Ciencia y Tecnología. (2024). *Glosario del acceso abierto*. Sistema Nacional de Repositorios Digitales. <https://repositoriosdigitales.mincyt.gov.ar/vufind/Content/glosario>.
- Ministerio de Ciencia, Tecnología e Innovación Productiva. (2015). *Directrices para proveedores de contenido del sistema nacional de repositorios digitales*. https://repositoriosdigitales.mincyt.gov.ar/files/Directrices_SNRD_2015.pdf
- National Science Board. (2005). *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*. National Science Foundation. <https://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>.
- Organización de las Naciones Unidas. (2018). *Promoción y protección del derecho a la libertad de opinión y expresión*. Septuagésimo tercer período de sesiones. <https://documents.un.org/doc/undoc/gen/n18/270/45/pdf/n1827045.pdf>.
- Oficina de Conocimiento Abierto. (2018). *Acceso Abierto en Europa*. Universidad Nacional de Córdoba. <https://oca.unc.edu.ar/2018/10/23/acceso-abierto-en-europa/>.
- Organización para la Cooperación y el Desarrollo Económicos. (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. https://www.oecd-ilibrary.org/science-and-technology/oecd-principles-and-guidelines-for-access-to-research-data-from-public-funding_9789264034020-en-fr.
- Real Academia Española. (s.f.-a). Dato. En *Diccionario de la lengua española*. <https://dle.rae.es/dato>.
- Real Academia Española. (s.f.-b). Eficiencia. En *Diccionario de la lengua española*. <https://dle.rae.es/eficiencia?m=form>.
- Real Academia Española. (s.f.-c). Eficacia. En *Diccionario de la lengua española*. <https://dle.rae.es/eficacia?m=form>.
- Real Academia Española. (s.f.-d). Metadato. En *Diccionario panhispánico del español jurídico*. <https://dpej.rae.es/lema/metadato>.
- Skytland, N. (2012). *What is NASA doing with Big Data today?* <https://skytland.medium.com/what-is-nasa-doing-with-big-data-a6c748588588>.
- Springer Nature. (2024). *Green or Golden routes to open access*. <https://www.springernature.com/la/open-science/about/green-or-gold-routes-to-oa>.
- Tapscott, D. y Tapscott, A. (2016). *Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World*. Portfolio.
- Terragno, R. H. (1998). *Maitland & San Martín*. Universidad Nacional de Quilmes.
- Universidad de Chile. (s.f.). *Gestión de datos de investigación. Ciclo de vida*. <https://uchile.cl/informacion-y-bibliotecas/gestion-de-datos-de-investigacion/datos/ciclo-de-vida#>.

