



Más allá de los patrones y las correlaciones: de los datos al *understanding*

Julián Reynoso*

Andrés A. Ilcic‡

Resumen

El foco de este trabajo está puesto en la manera en la que el concepto de “dato” es caracterizado, dado que es una noción que se utiliza como “primitiva”, en tanto rara vez es trabajada, pero al mismo tiempo pareciera cargar con una gran responsabilidad. Un dato, por sí solo (aislado), es incapaz de proveer información o nuevo conocimiento, sino que debe ser considerado en la constelación de herramientas (tanto teóricas como informáticas) en las que nace y se puede interpretar. De esta manera, un dato adquiere significado solo cuando se sitúa dentro de un conjunto más amplio de instrumentos y conceptos que facilitan su comprensión, permitiendo que emerja un conocimiento útil y significativo. Sin este contexto, el registro queda reducido a un elemento sin mayor relevancia o capacidad explicativa.

Palabras clave: Dato, comprensión, patrones, modelos.

Abstract

The focus of this paper is on the way in which the concept of ‘data’ is characterised, given that it is a notion that is used as ‘primitive’, in that it is rarely worked with, but at the same time seems to carry a great responsibility. A piece of data, on its own (in isolation) is incapable of providing information or new knowledge, but must be considered in the constellation of tools (both theoretical and informational) in which it is born

* Universidad Nacional de Córdoba - Instituto de Humanidades – CONICET

‡ Universidad Nacional de Córdoba - Instituto de Investigaciones Filosóficas – CONICET/
SADAF

and can be interpreted. In this way, a piece of data acquires meaning only when it is placed within a broader set of tools and concepts that facilitate its understanding, allowing useful and meaningful knowledge to emerge. Without this context, the record is reduced to an element without greater relevance or explanatory capacity.

Keywords: Data, understanding, patterns, data travel, models.

I. Introducción

El interés por lo que se constituye como “dato” ha sido una preocupación relativamente reciente para la filosofía de las ciencias dado que el interés estuvo centrado, por lejos, en la compleja relación entre teorías, modelos, y aquello que pretenden explicar, representar o, en definitiva, dar cuenta. Al desenmarañar ese nudo han ido surgiendo nuevas preocupaciones que permitieron sofisticar la caracterización de las prácticas científicas contemporáneas y la noción de dato es una de ellas. A los efectos de este trabajo, ensayaremos una periodización para ordenar la discusión, cuyos criterios son enteramente arbitrarios. Tomaremos como eje la idea del “cuarto paradigma” de la investigación que propusiera Jim Gray según el cual el impacto de las tecnologías de la información ha causado un cambio profundo en la manera de hacer ciencia.

El “diluvio de datos” al que se enfrentan investigadores en un gran porcentaje de las áreas de investigación obliga a desarrollar técnicas, infraestructura y metodologías para garantizar la confiabilidad del conocimiento producido. El trabajo en filosofía de las ciencias, durante la primera mitad del siglo XX, estuvo marcado fuertemente por los postulados del positivismo lógico cuya piedra de toque estaba en el estudio de las teorías científicas y de qué manera era posible establecer relaciones entre los enunciados con contenido empírico a la estructura axiomática a la que las teorías debían aspirar. Los datos, para dicha escuela, podían ser enteramente capturados por los llamados “enunciados observacionales”, que luego podían ser introducidos a la axiomática garantizada por la lógica.

En ese marco, la noción de dato no resultó de particular interés para los filósofos de la ciencia, sino que el concepto de explicación recibió más atención, en tanto permitía dar cuenta del *porqué* de determinados fenómenos, bajo parámetros muy concretos de cómo se produce tal explica-

ción. No realizaremos un recorrido exhaustivo por las transformaciones que experimentó la disciplina, pero baste señalar que, a medida que se fue sofisticando la comprensión sobre las prácticas científicas, afloró un gran conjunto de problemáticas y nociones que no habían sido tratadas anteriormente.

Lo que nos ocupa en el presente trabajo es, por lo tanto, delinear algunas características de la noción de dato que opera en el marco de las prácticas científicas contemporáneas, con especial atención a la astronomía. Su largo y rico historial en la acumulación de registros y observaciones que han servido para apuntalar y acunar una multitud de hipótesis y teorías. En la actualidad, el despliegue de infraestructura computacional ha permitido la captura y procesamiento de cantidades inconcebibles de registros. Tomaremos, a modo de ejemplo, un estudio reciente que llevó a cabo Suárez (2023) sobre avances en el campo de la astrosismología, una especialidad de la astrofísica ocupada de la estructura interna de las estrellas.

En este sentido, podemos señalar que la noción de “dato” cobró una gran relevancia a principios de la segunda década del presente siglo, cuando coincidieron en la esfera pública una serie de desarrollos e innovaciones que habían conseguido cierta inercia en las dos o tres décadas previas. Factores tales como el aumento en la capacidad de cómputo de los equipos disponibles, a un precio más accesible; junto con la disminución del costo de almacenamiento y la digitalización facilitada por la ubicuidad de los dispositivos de captura y registro en ámbitos tan diversos como las ciencias, la industria, el comercio y hasta nuestra propia vida cotidiana¹ confabularon para permitir el surgimiento de fenómenos *data-driven*, una multiplicidad de manifestaciones, procesos y técnicas cuya particularidad radica en “ser guiadas por datos”². Existen numerosos ejemplos, como son las “ciencias *data driven*”, procesos de toma de decisiones *data driven* y hasta estrategias *data driven* para el deporte y la política. Todo este abanico de manifestaciones se agrupó bajo el término *big data*, aunque esa expresión perdió algo de tracción en los últimos años. La noción de dato, sin embar-

1 Puede pensarse en la cantidad de dispositivos biométricos que registran pasos, pulsaciones, respiraciones y hasta calidad de sueño que usamos a diario (por no mencionar también dispositivos estrictamente médicos como medidores de glucosa o de tensión sanguínea).

2 Ver Patil y Mason (2015).

go, parece haber quedado sepultada bajo este “diluvio” que abrigó un fuerte optimismo en torno a las posibilidades que estos desarrollos facilitarían.

Pese a la enorme cantidad de bibliografía que existe sobre estas tendencias, sin embargo, no parece haber recibido mucha atención a qué se denomina “dato”. La propia noción, sin embargo, dificulta a tal análisis puesto que parece ser algo elusiva: resulta difícil echar luz a qué es un dato sin inmediatamente considerar el entramado de relaciones en las que se enmarca. En otras palabras, no resulta sencillo hablar de qué es un dato sin hablar acerca de lo que ese dato “refiere”. Tal como mencionamos anteriormente, la computarización y digitalización despertó una serie de cambios fundamentales en la manera de hacer ciencia, con efectos diferentes en los múltiples campos que fueron afectados y la astronomía no estuvo exenta de tales modificaciones. El paso de las computadoras de ser simples máquinas auxiliares para realizar cálculos complejos en menor tiempo, a ser dispositivos capaces de simular fenómenos para su estudio significó la posibilidad de explorar y aumentar el espectro del campo de estudio. Esta transformación fue fundamental en campos como la astronomía y las ciencias de la atmósfera, disciplinas en las que es prácticamente imposible intervenir experimentalmente de manera directa y la observación de determinados fenómenos puede llegar a ser prohibitiva por las escalas temporales involucradas.

II. Datos sobre los datos

Resulta ineludible dar comienzo a esta sección haciendo un rastreo casi etimológico de la noción dado que es un término con una larga historia, aunque sus acepciones han ido variando con el correr de las épocas. Luego rescataremos algunas características significativas de la noción de dato a partir de los aportes de quienes identificamos como pilares en el estudio de esta problemática. Esto nos permitirá, en la sección siguiente, ilustrar con un ejemplo a partir del trabajo que realizó Mauricio Suárez sobre la misión CoRoT.

II.1 Breve perspectiva histórica

Dado el foco de la investigación, de la que el presente trabajo forma parte, dejaremos de lado la discusión ontológica que puede surgir de inmediato al leer la pregunta que encabeza esta sección. En definitiva, no nos detendremos en la discusión sobre el estatus ontológico de los datos, sino que intentaremos caracterizarlos de manera tal que permita entender la forma en la que la noción existe hoy en las prácticas científicas. La literatura en español no es abundante, sin embargo, tanto “dato” en español como su contraparte anglosajona “*data*” comparten un origen en común en el Latín *datum*, “lo dado”.³ En idioma inglés, en cambio, podemos encontrar el trabajo de Rosenberg (2013) en el que hace un rastreo del uso de la noción de *data* hasta las *Lectures on History and General Policy* de Priestley (1788) en el que se refiere a los hechos de la historia como “datos”. Rosenberg señala que en el contexto de las *Lectures* este uso resulta completamente natural, pese a que pareciera significar lo contrario a lo que hoy entendemos como dato, más ligado con los avances burocráticos y estadísticos del siglo XIX y los desarrollos en informática del siglo XX. Rosenberg señala que el término aparece en una gran variedad de contextos a lo largo del siglo XVIII, en un sentido similar al que emplea Priestley. Rosenberg emprende una breve recorrida por la etimología de la familia de vocablos que parece estar tan íntimamente relacionada: datos, evidencia y hechos: *datum* viene de *dare*; evidencia de *videre* y fact de *facere*. De allí, el autor concluye que “los hechos son ontológicos, la evidencia epistemológica y los datos son retóricos” Rosenberg (2013, p. 18). Esto lleva al autor a señalar que un hecho, al ser falseado, deja de ser un hecho, pero los datos falsos siguen siendo datos.

Siguiendo el recorrido histórico, en el siglo XVII el término *data* se utilizaba en un sentido técnico para referirse a las cantidades dadas en un problema matemático concreto, en oposición a la *quaestio*, aquello que el problema intentaba responder. En filosofía natural se empleaba en un sentido similar, para referir la categoría de hechos y principios que estaban fuera de discusión. Un sentido similar se retuvo en teología, al referir a las verdades provenientes de las Escrituras. Rosemberg señala que, en la filosofía natural del siglo XVII -y tal como sucedía en teología o en ma-

³ *Datum* es el participio pasado del verbo *Dare*, “dar” en español.

temáticas- la noción de *data* identificaba una categoría de hechos y principios que no eran pasibles, por convención, de ser puestos en cuestión.

Para rastrear el uso del término *data* en idioma inglés Rosenberg utilizó un enfoque cualitativo con distintas herramientas con el objetivo de dar cuenta del uso que hizo Priestley. Sus principales fuentes fueron el “Visor de *Ngram*” provisto por Google Books y la *Eighteenth Century Collections Online*, un repositorio digital de volúmenes publicados en Gran Bretaña durante el siglo XVIII. El autor concluye que, desde una perspectiva estadística, *data* no fue un término ni particularmente raro ni muy común en la literatura inglesa del siglo XVIII y fue ampliamente opacado en su uso por otros como *fact*, *evidence* y *truth*.

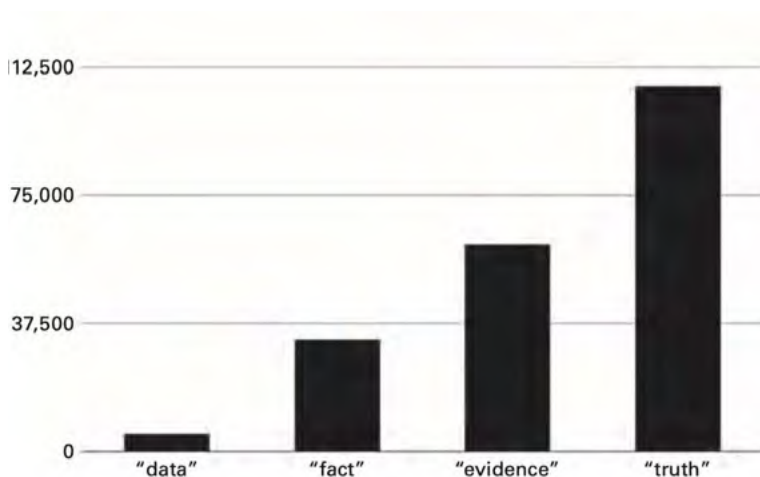


Figura 1. Fuente: Rosenberg (2013).

El término termina de establecerse hacia el siglo XVIII, pero hacia finales de este siglo se hizo patente una transformación y pasó de referir a cuestiones indiscutibles a referir a hechos evidenciados por experimentos, experiencias o recopilación: “Era habitual referirse a los datos como el resultado de una investigación y no como su premisa” Rosenberg (2013, p. 33). No fue hasta el siglo XX, con la aparición de las computadoras digitales y la teoría de la información que el concepto adquirió la importancia y

ubiquidad que tiene hoy, pero cuyas bases ya fueron sentadas en por el uso durante el siglo XVIII. Ese camino fue allanado en el siglo XIX mediante el desarrollo de métodos estadísticos para el procesamiento de los censos y la confección de anales estadísticos (cf. Hacking, 1991). Meyns (2019) emplea un método similar al de Rosenberg, pero circunscrito a las *Philosophical Transactions* de la Royal Society, con el mismo objetivo de mostrar cuando comienza a utilizarse el término *data* en la publicación. La historia, sin embargo, no terminaría aquí y con el advenimiento de avances en la estadística, la informática y la computación contribuyeron al “diluvio de datos” que caracterizará el cuarto paradigma de la investigación del que hablara Jim Gray.

II.2 Digitalización y datificación

Se ha llamado “datificación” al proceso mediante el cual “se cuantifica un fenómeno de manera tal de tabularlo y analizarlo”, y se le atribuye a un paneo que realizaron Mayer-Schönberger y Cukier (2013) sobre *big data* en distintas áreas. Según los autores, es un proceso que involucra mucho más que convertir materiales simbólicos en una forma digital. Hacking, sin embargo, ya había hecho referencia a esta cuestión allá por 1991 y señala como desde principios del siglo XIX, los incipientes estados-nación⁴ en Europa y América del norte no sólo comenzaron a llevar registros más puntillosos de su población, sino que también empezaron a hacerlos disponibles a la población en general, en forma de anuarios estadísticos y una avalancha de “números impresos” que fueron posibilitados por un desarrollo de nuevas técnicas y tecnologías para clasificar y enumerar a la población, como así también el desarrollo de dependencias burocráticas con la autoridad suficiente para el despliegue de estas tecnologías. Toda una infraestructura teórica y administrativa que permitió que la población sea contada y clasificada dentro de las nuevas categorías estadísticas. La digitalización de estos registros significó otro paso en el camino al procesamiento de los registros y la posibilidad de “accionar” sobre ellos al facilitar las operaciones de cálculo que se podían realizar sobre dichos registros.

4 Aquí podría pensarse también en una co-aparición entre estadísticas y estados.

II.3 Datos y fenómenos

Volcaré mi atención ahora a las discusiones más recientes en torno a la noción de dato en filosofía de la ciencia, foco de interés principal del presente trabajo. En tal sentido, el trabajo de Bogen y Woodward (1988) es central para comprender cómo la noción de dato aparece en el escenario en las discusiones en filosofía de las ciencias. Si bien en otras ramas de la filosofía, como en teoría del conocimiento, se trabajó extensamente la idea de dato en relación con los sentidos, en el sentido más técnico que se emplea actualmente aún no había aparecido. Para la presente sección tomaremos como base el artículo de 1988, pero es importante tener en cuenta que ese fue sólo el puntapié inicial de una discusión que sigue hasta el día de hoy⁵.

La tesis principal de Bogen y Woodward en el artículo de 1988 es bastante directa: las teorías científicas no proveen una explicación sistemática de los hechos acerca de los cuales hacen referencias los datos y las observaciones. Apuntan contra una versión ampliamente difundida a finales de los '80 según la cual las teorías científicas predicen y explican hechos (*facts*) acerca de “observables”, objetos y propiedades que pueden ser percibidos por los sentidos o aumentadas por el uso de distintos instrumentos. Los autores señalan que, si bien esta versión “positivista” de la labor científica ha sido abandonada en su mayoría, aún hay -incluso entre los críticos acérrimos al positivismo- quienes sostienen que la función principal de las teorías científicas es predecir y explicar afirmaciones acerca de lo que se observa. Han logrado esto al ampliar la noción de “observación” y “observabilidad”. Bogen y Woodward apuntan a mostrar que, si “observar” significa “percibir o detectar por medio de procesos que pueden ser entendidos como extensión de la percepción”, entonces las teorías científicas no suelen predecir ni explicar hechos sobre lo que observamos. Los autores afirman que la categoría de “observables” (y lo que no es observable por extensión) y las discusiones que se generaron en torno a la carga teórica de la observación son mucho menos centrales para comprender la actividad científica de lo que se suponía en aquel entonces.

5 Sólo porque las discusiones en filosofía no se terminan, sino que se dejan a un lado.

El argumento de los autores descansa en una distinción que hasta el momento había sido ignorada y que será crucial en el desarrollo posterior de este trabajo: la distinción entre fenómenos y datos. Aquí es que Bogen y Woodward estipulan una definición de dato que tomaré como base. Para ellos los datos se caracterizan por ser observados de manera más o menos directa y juegan el rol de evidencia para la existencia de los fenómenos. La contrapartida son los fenómenos, detectables a través de los datos, pero no observables en ningún sentido interesante del término. Bogen y Woodward comienzan con algunos ejemplos con los que pretendan ilustrar tal distinción. El primero trata sobre el descubrimiento de la Corriente Neutral Débil en 1973 a partir de dos experimentos llevados de manera independiente en el CERN y el NAL, un laboratorio de Fermilab. Los datos obtenidos en el CERN consistían en unas trescientas mil fotos de una cámara de burbujas, de las cuales unas 100 fueron consideradas como evidencia a favor de estas corrientes neutrales. En el NAL, por su parte, se registraron patrones de descargas en detectores de partículas electrónicos. De trescientos registros, se consideró que 8 eran evidencia de dichas corrientes. En otras palabras, ambos laboratorios encontraron datos distintos que apuntaban al mismo fenómeno.

Estos ejemplos los llevan a Bogen y Woodward a concluir que “Los datos son (...) idiosincráticos de contextos experimentales particulares, y normalmente no pueden darse fuera de esos contextos.” (1988, p. 317). ¿Por qué? Recordemos que Bogen y Woodward están preocupados en resaltar el carácter evidencial de los datos en tanto sirven para dar cuenta de fenómenos. En pos de rescatar tal carácter, para que un dato pueda servir como evidencia confiable (*reliable*) de la ocurrencia del fenómeno en cuestión, los autores señalan que es necesario tener en cuenta que la captura de ese dato es el resultado de un proceso altamente complejo y una “coincidencia inusual de circunstancias”⁶.

Bogen y Woodward señalan también que los datos deben ser “relativamente fáciles de identificar, clasificar, medir, agregar y analizar” (cf. 1988, p. 320) de forma tal que sean confiables y reproducibles por demás integrantes de la comunidad científica. Profundizaré sobre esta cuestión más adelante, pero por ahora basta con señalar que el tiempo no sólo les dio

6 El complejo proceso de captura, almacenamiento y procesamiento de datos será objeto de un trabajo posterior, pero ciertamente amerita una mirada más en detalle.

la razón, sino que ya había en funcionamiento una gran maquinaria destinada a fomentar la cooperación, el intercambio y la gestión de datos. La “tratabilidad” de los datos es otra de las características que los autores señalan como parte de su caracterización, como así también la necesidad el proceso de captura haya tenido un control adecuado para los distintos tipos de errores experimentales que puedan darse en el proceso.

En un artículo más reciente James Woodward redobla en la caracterización de dato que propusiera con Bogen en el ya mencionado artículo de 1988:

“Datos son registros públicos producidos por mediciones y experimentos que sirven de evidencia para la existencia de, o características de los fenómenos.” (Woodward, 2011, p. 166)

Es importante señalar nuevamente el carácter “subsidiario” que tiene la noción de dato aquí, relevante solo en tanto funciona como evidencia para el fenómeno en cuestión. Woodward vuelve a recordarnos también aquel carácter idiosincrático, que los datos capturados también reflejan la influencia de otros factores causales que nada tienen que ver con el fenómeno bajo estudio: “los datos son científicamente útiles e interesantes en tanto proveen información sobre características de los fenómenos” dice Woodward (2011, p. 168).

II.4 Vida y viaje de los datos

La discusión iniciada por Bogen y Woodward con su artículo de 1988 fue adquiriendo nuevas dimensiones a medida que los métodos y herramientas continuaron avanzando a un ritmo acelerado. A medida que los costos de almacenamiento y procesamiento computacional fue bajando, cada vez más investigadores tuvieron acceso a herramientas computacionales y explorar áreas del conocimiento que, de otra manera, hubieran permanecido lejanas e inaccesibles. Los cuadernos y bitácoras de investigación donde se registraban los avances fueron mutando a hojas de cálculo y bases de datos que dejaron de ser exclusividad de los grandes proyectos de *big science* y empezaron a estar a disposición de mayor cantidad de científicos. Como ya hemos mencionado, parte del interés en la noción de “dato” viene dada por esta transformación.



Sabina Leonelli ha sido una pionera en el análisis que significaron estos avances en las llamadas “ciencias de la vida”, cuyo trabajo se cristalizó en un libro publicado en 2016 en el que analiza en detalle la forma en la que se transformó la investigación en estas áreas del conocimiento a partir del influjo de herramientas. Su tesis principal, sin embargo, va en un sentido original contrario a la corriente más establecida. Para Leonelli, la fuente de innovación en la biología contemporánea es “la atención que se le presta al manejo y disseminación de los datos (...) en lugar del surgimiento del *big data* y los métodos asociados” (Leonelli, 2016, p. 8).

Leonelli entiende a los datos como “herramientas para la comunicación, cuya principal función es permitir intercambios materiales e intelectuales entre individuos, colectivos, culturas y naciones”. Aclara también que la movilidad de los datos entre estos grupos es un gran logro científico. La novedad de esta caracterización radica en el peso que Leonelli le pone a la posibilidad de compartir y poner en viaje a los datos, a diferencia de las caracterizaciones más tradicionales que se han enfocado solamente en la producción e interpretación de los datos. En este sentido, Leonelli añade que los datos son “artefactos materiales” cuyas características concretas (como el formato o el medio en el que se encuentran) son tan significativos como sus funciones conceptuales.

Esta caracterización, a la que la propia Leonelli llama “versión relacional” de los datos, tiene por particularidad definir a los datos en términos de la función dentro de procesos específicos en los recuentos de las investigaciones, y no en términos de propiedades intrínsecas. “Carece de sentido preguntarse cuales objetos cuentan como datos en abstracto” dice Leonelli, y añade “Esta pregunta sólo puede ser respondida haciendo referencia a situaciones concretas de investigación (...)” (Leonelli, 2016, p. 78)

Bis: datos sobre los datos

Señalaremos, entonces, algunas características que nos resultan significativas para una caracterización de la noción de dato, a la luz de los desarrollos más recientes. En primer lugar, consideramos que los datos no son meros registros ni surgen de un proceso simple de captura del mundo, sino que son un dispositivo mediante el cual *modelamos* el mundo. Son el producto de proceso mediados técnicamente, tanto por instrumentos y teorías, el resultado de procesos organizacionales, categorías y estándares

que suelen ser negociados y debatidos. Kitchin (2022, p. 33) agrega que hay un grado de “desprolijidad” en la generación de los datos: hay un proceso de ajuste, calibración y retroalimentación entre las distintas etapas de la producción de los datos que permiten ajustar de acuerdo con objetivos epistémicos. En segundo lugar nos parece relevante la consideración de Leonelli sobre los “viajes de los datos”, pero no sólo en el sentido de desplazamiento geográfico que facilita la colaboración entre distintos grupos e instituciones, sino también considerando el proceso mencionado en el párrafo anterior: dado que los datos pueden ser obtenidos de múltiples fuentes, lo que en una parte del proceso puede ser considerado un “resultado” (como puede ser el registro de una medición, el *output* de una simulación o un experimento) puede luego convertirse en *input* en otra instancia y ser considerado así también un “dato”. Por último, nos es ineludible plantear un parcial desacuerdo con la versión que plantean Bogen y Woodward según la cual parece haber una distinción tajante entre dato y fenómeno, dado que podría pensarse que hay una co-constitución entre ambos. En otras palabras, podríamos señalar que *dinos como has conseguido tus datos y te diremos que fenómeno tienes*.

III. Datos y comprensión

Señalábamos más arriba que la noción de explicación había recibido históricamente mayor atención por parte de la comunidad epistemológica. En los últimos tiempos, sin embargo, la noción de comprensión (*understanding*) comenzó a ganar terreno en las discusiones. Consideramos que tal noción ofrece un marco apropiado para integrar algunas de las discusiones que han surgido, como por ejemplo la imbricada relación que existe entre datos, modelos y teorías y el bucle de retroalimentación del cual parecemos obtener nuevo conocimiento, tanto de ellos mismos como de los fenómenos en cuestión.

La comprensión también se ha descuidado, aunque en menor medida, y muchas de las razones de este descuido se basan en las mismas razones históricas, a saber, no ser lo suficientemente «objetiva». Consideramos que la comprensión es un objetivo epistémico válido y deseable, que no entra en conflicto con otros, como proporcionar explicaciones, hacer predicciones o luchar por la adecuación empírica, por ejemplo. De hecho,

estos objetivos no sólo no entran en conflicto, sino que creemos firmemente que tienen el potencial de beneficiarse mutuamente.

Como todo concepto que los filósofos toman en consideración, la comprensión ha sido analizada e interpretada desde muchos ángulos diferentes. Para los propósitos del presente trabajo, nos basaremos en los trabajos de De Regt (2015) y Elgin (1993). Estos autores comparten, con nosotros y entre ellos, un objetivo común: defender la idea de que la comprensión tiene un valor epistémico distintivo.

Hemos identificado tres ideas centrales en sus planteamientos que nos permitirán avanzar en la idea de pericia científica:

1. En primer lugar, una caracterización no factual del conocimiento científico. De Regt sostiene que no es necesario plantear una relación intrínseca entre comprensión y verdad. En su opinión, es posible alcanzar una comprensión genuina de los fenómenos a través de modelos y teorías que desafían una interpretación realista. Elgin también argumenta en esta línea, su visión de la comprensión también es no factual y antirrealista, dado el papel que la idealización juega en la investigación, las teorías falsas pueden contribuir a avanzar en la comprensión científica.
2. En segundo lugar, pero relacionado con el primer punto, el éxito epistémico se evalúa en términos de objetivos epistémicos (y no en términos de factualidad del conocimiento). Esto significa que la comunidad científica, encargada de establecer los objetivos epistémicos que guían la investigación, desempeña un papel clave en su proceso, ideando métodos y estrategias para alcanzar dichos objetivos.
3. Y por último, pero no por ello menos importante, destacamos la idea de las capacidades. De Regt y Dieks caracterizan la comprensión en términos de lo que podemos hacer antes que de lo que podemos saber. Esto sitúa a los profesionales en el centro de la escena, de modo que la comprensión se refiere a sus estados epistémicos y no a las relaciones «objetivas» entre las leyes, las teorías y el mundo.

Así, podemos decir que, dada esta idea no factual del conocimiento científico, que plantea el éxito epistémico como una función de objetivos

epistémicos particulares, los expertos son aquellos que poseen ciertas habilidades para lograr sus objetivos, que pueden variar según la tarea y el dominio en cuestión. Consideramos que incorporar este enfoque como trasfondo de la investigación en curso nos permite reforzar una intuición que presentamos en Ilcic y Reynoso (2018):

“La ventaja epistémica de la postura que presentamos es que permite a los modelos teóricos funcionar como guías heurísticas al sondear los datos o el modelo de análisis de datos.”(Ilcic y Reynoso, 2018, p 191)

Incorporar a la comprensión científica nos permite también incorporar otros elementos al análisis, como el rol de la experticia (cf. Mondaca y Reynoso, 2020) y nos permite pintar un cuadro más completo de cómo dar cuenta de cómo se construye el conocimiento científico.

En tal sentido, el trabajo reciente de Suárez (2023) puede servir para ilustrar algunas de las consideraciones que hemos plasmado en este trabajo. En dicho artículo, el autor analiza con cierto nivel de detalle los resultados de la misión CoRoT (Auvergne et al., 2009), una iniciativa de la Agencia Espacial Francesa (CNES) en conjunto con la Agencia Espacial Europea (ESA) y otros socios internacionales (NASA entre ellos. Los dos objetivos de la misión eran buscar planetas extrasolares con periodos orbitales cortos, en particular los de gran tamaño terrestre, y realizar astrosismología midiendo oscilaciones similares a las solares en estrellas.

La preocupación del artículo de Suárez gira en torno a las condiciones que evitan la circularidad que podría producirse al emplear distintas capas de modelos “anidados” unos en otros al investigar la composición interna de las estrellas. El “anidamiento” se produce porque:

“(...) nuestros modelos de estrellas y su evolución son sensibles a otros modelos de interiores estelares estratificados, a modelos de la física de los materiales radiativos, a modelos de atmósferas estelares en astrofísica coronal, y a modelos de los modos de vibración que producen los datos astrofísicos que a su vez apoyan esos modelos.” (Suárez 2023, p. 112)

Según el autor nada en este «anidamiento» de modelos dentro de modelos puede servir para poner en duda la solidez del conocimiento obtenido sobre la física del interior de las estrellas. Sin embargo, en la cita del

párrafo anterior podemos observar la compleja red de modelos e instrumentos sobre la que se monta la producción del conocimiento científico. Sucede, dice Suárez (p. 121) “que las inferencias de los datos observados a los modelos de datos y las de estos modelos de datos al espacio de parámetros en los modelos teóricos son previsiblemente ricas en supuestos de modelización propios”, que a su vez determinan un rol crítico para determinar la calidad de los datos.

La misión CoRoT nos sirve también para rescatar otro aspecto señalado por Leonelli respecto a la importancia del empaquetamiento y disponibilización de los datos de investigación. Si bien la misión finalizó en 2013, hoy es posible acceder a los datos producidos por los satélites gracias al *Exoplanet Archive* que mantiene hoy el programa de Exploración de Exoplanetas de la NASA⁷ y que hoy gestiona un enorme repositorio de datos de investigación provenientes de múltiples misiones, como CoRoT, las curvas de Kepler y otros.

IV. Comentarios finales

Pocas ciencias gozan a la vez de una historia tan larga y rica y de un presente tan desafiante y vivo como la astronomía, por lo que ofrece campo por demás fértil para las reflexiones que hemos vertido en el presente artículo. Sabemos ya que dar cuenta del mundo es una tarea inmensamente compleja, ensayar explicaciones y formular teorías y modelos acerca de los fenómenos son algunas de las herramientas con las que contamos para la producción del conocimiento científico y esta actividad se ha enriquecido significativamente con el progresivo desarrollo de distintos sistemas tecnológicos en los que hemos podido “descargar” parte del trabajo cognitivo requerido. A su vez, el incorporar estos sistemas nos ha obligado a considerar también de qué manera producimos conocimiento y cuál es la relación entre las distintas mediaciones que aparecen en las diferentes etapas del proceso. La noción de comprensión que presentamos en la sección anterior nos habilita a ubicar estos pequeños bucles de reflexión sobre la propia práctica, no sólo científica sino también filosófica, sobre cómo construimos y por qué “confiamos” en los procesos que desembocan en nuevos *insights*. Ahora bien, un aspecto a tener muy presente es que la irrupción de métodos cada vez más automatizados para la detección de pa-

7 El repositorio se puede acceder en <https://exoplanetarchive.ipac.caltech.edu>.

trones en las bases de datos y el procesamiento en las primeras instancias de los flujos de trabajo que ocurren en la investigación no debe opacar la importancia de los agentes humanos en dicho proceso. Después de todo, aunque sea por ahora, la producción de conocimiento científico continúa siendo una empresa fundamentalmente humana, aunque concedemos que este tipo de ensamblajes cognitivos con fuerte participación de agentes no humanos obliga a tomar ciertos cuidados frente a determinadas afirmaciones.

Referencias

- Bogen, J., & Woodward, J. (1988). Saving the Phenomena. *The Philosophical Review*, 97(3), 303. <https://doi.org/10.2307/2185445>
- De Regt, H. W. (2015). Scientific understanding: truth or dare? *Synthese*, 192(12), 3781–3797. <https://doi.org/10.1007/s11229-014-0538-7>
- Elgin, C. Z. (1993). Understanding: Art and Science. *Synthese*, 95(1), 13–28. <https://www.jstor.org/stable/20117763>
- Hacking, I. (1991). *La domesticación del azar: La erosión del determinismo y el nacimiento de las ciencias del caos*. Gedisa.
- Ilcic, A. A., & Reynoso, J. (2018). Hacia una articulación de modelos: el caso de big data. In S. Chibeni, L. Zaterka, J. Ahumada, & D. Letzen (Eds.), *Filosofía e Historia de la Ciencia en el Cono Sur* (pp. 181–192). Editorial de la Universidad Nacional de Córdoba.
- Kitchin, R. (2022). *The data revolution: a critical analysis of big data, open data & data infrastructures* (Second edition). Sage.
- Leonelli, S. (2016). *Data-centric biology: a philosophical study*. The University of Chicago Press.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: a revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

- Meyns, C. (2019). "Data" in the Royal Society's Philosophical Transactions, 1665–1886. *Notes and Records: The Royal Society Journal of the History of Science*, 74(3), 507–528. <https://doi.org/10.1098/rsnr.2019.0024>
- Mondaca, S., & Reynoso, J. (2020). ¿La práctica hace a le expertx o le expertx hace a la práctica? In A. A. Ilcic, S. Mondaca, P. Torres, & A. N. Venturelli (Eds.), *30^a Jornadas de Epistemología e Historia de la Ciencia* (pp. 65–75). Editorial de la Facultad de Filosofía y Humanidades - UNC. <https://revistas.unc.edu.ar/index.php/jornadaehc/article/view/28681>
- Patil, D. J., & Mason, H. (2015). *Data Driven*. O'Reilly Media, Inc.
- Priestley, J. (1788). *Lectures on History and General Policy: To which is Prefixed, An Essay on a Course of Liberal Education for Civil and Active Life*. P. Byrne.
- Rosenberg, D. (2013). Data before the fact. In L. Gitelman (Ed.), *"Raw Data" is an oxymoron* (pp. 15–40).
- Suárez, M. (2023). Stellar structure models revisited: Evidence and data in asteroseismology. In N.
- Mills Boyd, S. De Baerdemaeker, K. Heng, & V. Matarrese (Eds.), *Philosophy of Astrophysics* (Vol. 472, pp. 111–129). Springer International Publishing. https://doi.org/10.1007/978-3-031-26618-8_7
- Woodward, J. F. (2011). Data and phenomena: a restatement and defense. *Synthese*, 182(1), 165–179. <https://doi.org/10.1007/s11229-009-9618-5>

