

Artículo de investigación

El camino hacia la ciencia abierta

Gonzalo L. Villarreal^{1,2*}

¹Universidad Nacional de La Plata (Argentina)

²Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (Argentina)

*Correspondencia: gonzalo@prebi.unlp.edu.ar

Recibido: 3 abr. 2024 | 1ra decisión: 20 may. 2024 | Aceptado: 20 may. 2024 | Publicado: 27 dic. 2024



Resumen

Esta ponencia busca contrastar el circuito científico tradicional con las propuestas de la ciencia abierta. Se introduce como hipotético caso de estudio a Pablo, un doctorando que recopiló datos, publicó artículos y presentó su tesis doctoral, pero dado que no difundió en abierto ni los datos recopilados ni las publicaciones, detiene el avance de las investigaciones y obliga a duplicar esfuerzos. Se propone el uso de licencias abiertas y de software abierto, el depósito de datos y la incorporación de vías alternativas de publicación, a fin de acelerar el avance de la ciencia y democratizar el conocimiento generado.

Palabras clave: acceso abierto, colaboración, reproducibilidad, conocimiento compartido.

O caminho para a ciência aberta

Resumo: Esta apresentação busca contrastar o circuito científico tradicional com as propostas da ciência aberta. Introduz-se como caso de estudo hipotético Pablo, um doutorando que coletou dados, publicou artigos e apresentou sua tese de doutorado, mas, como não divulgou abertamente os dados coletados nem as publicações, impede o avanço das pesquisas e obriga a duplicação de esforços. Propõe-se o uso de licenças abertas e de software livre, o depósito de dados e a incorporação de vias alternativas de publicação, a fim de acelerar o avanço da ciência e democratizar o conhecimento gerado.

Palavras-chave: acesso aberto, colaboração, reprodutibilidade, conhecimento compartilhado.

The path to open science

Abstract: This presentation aims to contrast the traditional scientific circuit with the proposals of open science. A hypothetical case study introduces Pablo, a doctoral student who collected data, published articles, and defended his dissertation. However, because he did not openly share the collected data or publications, he hinders research progress and forces duplication of efforts. The presentation advocates for the use of open licenses and open software, data repositories, and the adoption of alternative publication avenues to accelerate scientific progress and democratize the knowledge produced.

Keywords: open access, collaboration, reproducibility, shared knowledge.

En esta presentación¹ voy a hablar de ciencia abierta con un caso más o menos real (por suerte hoy ya no tan real). La idea de esta presentación es hablar sobre la importancia de adoptar prácticas de ciencia abierta, cuáles son sus consecuencias positivas y mostrar las consecuencias negativas de no hacerlo. Hablaremos también de las razones por las cuáles un/a alumno/a de grado, de postgrado, o un/a investigador/a, debería involucrarse en ciencia abierta. ¿Qué ventajas tiene y qué desventajas tiene para ellos/as?

El caso de Pablo

Voy a tomar un personaje de ejemplo (cualquier relación con la realidad es pura conciencia). La ciencia se hacía de una manera antes. Al principio, de una única manera: los investigadores desarrollaban las indagaciones en sus universidades y los resultados se compartían sólo en un circuito muy cerrado. Se escribían los avances y, cada tanto, hacían publicaciones que condujeron al surgimiento de algunas revistas científicas. Esto sucedió de esa manera hasta que un buen día vino internet, nos conectó a todos y eso cambió cómo hacemos ciencia. Vamos a poner el ejemplo: Pablo es un estudiante de doctorado en psicología y lo está haciendo con una beca del CONICET, o de la UNLP, o de la CIC.

Como parte de su investigación, Pablo tuvo que ponerse durante años a relevar datos de un montón de pacientes y formó una base de datos con Excel, o algún programa similar. Tomó datos, visitó a los pacientes, se contactó con otros profesionales que tenían datos de los pacientes y que se los podían compartir. Esto le demandó mucho trabajo. Son años de recolectar toda esta información para hacer este relevamiento. Pablo tuvo que moverse por el país, ir a un centro de investigación en Córdoba, quizás tuvo que ir a Ushuaia a una clínica especializada, o entrevistarse con otros pares en Salta, y eso requiere tiempo y dinero. Con lo cual, la institución que financia su trabajo doctoral le otorgó un subsidio para viajes y, gracias a eso, pudo costear el traslado a los distintos lugares y obtener información directamente desde la fuente. Todo el trabajo de Pablo no es gratis, alguien lo tiene que pagar y, en este caso, lo costean las instituciones científicas y académicas del país. Esto está buenísimo, gracias a ese dinero Pablo puede hacer su tesis doctoral.

Por supuesto, Pablo también tuvo que hacer cursos que le enseñaron cosas vinculadas a su tema investigación y también, muchas veces, cosas vinculadas a trabajar con datos, estadística, ciencia de datos, el funcionamiento de las bases de datos, cómo recopilar información de manera estructurada, cómo escribir pequeños programas o algoritmos para explotar esos datos sin limitarse a un Excel o un gráfico de torta. Todo esto requiere de una infraestructura tecnológica, redes, servidores, computadoras. Obviamente, el laboratorio donde trabaja Pablo le ofrece un montón de servicios para alojar datos y procesarlos. Es muy difícil encarar una carrera doctoral sin este tipo de apoyo porque aparte de la infraestructura física del edificio, hay que pagar la parte humana: tutores, personal de apoyo, administrativos, etc. Las instituciones ponen todo ese marco a disposición de los estudiantes y eso está buenísimo.

Un buen día Pablo empezó a analizar sus datos y encontró algunos resultados que eran bastante interesantes. Justo hubo un simposio nacional donde se discutían estos temas, él fue y presentó los primeros resultados. Está buenísimo compartir

con sus pares los avances que uno va logrando. Más adelante, los resultados preliminares empezaron a afianzarse, que fueron muy prometedores; Pablo los presentó en un congreso y pudo hablar con sus pares. Con todo este background y todos esos intercambios, escribió un artículo para una revista internacional. Reunió sus publicaciones y escribió su tesis doctoral, que luego defendió. Ahora, Pablo es doctor. Sin embargo, la historia no termina acá. El trabajo que hizo estuvo buenísimo y una editorial le ofreció escribir un libro para un público más amplio, que pueda ser vendido y así poder darle difusión. Pablo escribió el libro y, junto con eso, vinieron reportajes en los medios, lo invitaron a un podcast y a escribir artículos en un diario de tirada nacional. En síntesis, el trabajo de Pablo estaba resonando.

Como no puede ser de otra manera, cuando uno empieza a tener avances tan prometedores, muchas personas comienzan a interesarse por el trabajo de uno y a seguir esa misma línea. Surgen nuevos becarios doctorales que quieren continuar la línea de investigación y extenderla, por ejemplo, sumando nuevas variables al análisis de Pablo. Sin embargo, puede darse el caso que esos nuevos becarios no tengan los mismos recursos que Pablo, ni cuenten con un subsidio para viajar a las mismas clínicas que él visitó, a lo largo y ancho del país. Otros querrán sumar datos de otros estudios y relacionarlos con los de Pablo, y otros más buscarán diseñar sus propios algoritmos para analizar los mismos datos que él usó, pero de otra manera.

Sin embargo, no todo es tan sencillo. Los estudiantes nuevos, que quieren empezar a trabajar sobre las bases que sentó Pablo, no tienen tan fácil el acceso a todos sus avances. Resulta ser que ni el simposio ni el congreso, en los que participó Pablo, publicaron actas, sino que se limitaron a publicar un libro de resúmenes de 200 palabras por trabajo. Además, las conferencias no se grabaron, o las grabaciones no están disponibles para el público. Con lo cual, los investigadores nuevos solamente saben que Pablo estuvo en esos eventos; quizá pudieron escucharlo, pero todo quedó ahí. Además, la revista tan reconocida en la que publicó trabaja con un modelo de suscripción. Eso quiere decir que hay que comprar el artículo, que cuesta entre 20 y 50 dólares, para poder leerlo, o hay que formar parte de una universidad con suscripciones a base de datos para acceder a la investigación de Pablo. Como si eso fuera poco, la tesis doctoral no está en el repositorio de su universidad, o está allí pero tiene un embargo de 6 meses que impide verla completa. El libro en el que Pablo publicó su tesis también hay que pagarlo. De esta forma, quedarían los reportajes, el artículo en los medios y el podcast, como herramientas muy valiosas para dar a conocer su labor, pero que resultan insuficientes para quien desea desarrollar la línea de investigación que trazó nuestro investigador. Eso es difusión.

Entonces, la conclusión aparente sería que debemos repetir el esfuerzo que hizo Pablo. Él tiene todos los datos que relevó, seguramente en su computadora y en algún backup, pero no los publicó. La información que recopiló la tiene él, solamente. Como los investigadores jóvenes no poseen esos datos, deben volver a recopilarlos, ir a las mismas instituciones, hacer todo el trabajo que Pablo ya hizo. Lo mismo ocurre en cuanto al código: Pablo desarrolló programas y algoritmos que tenía guardados en su computadora, pero que jamás publicó. Por lo tanto, nadie puede ver ni evaluar el algoritmo que creó, ni intentar mejorarlo. Debemos confiar ciegamente en Pablo y creer que todos sus algoritmos están bien.

No tiene sentido que los estudiantes deban perseguir a Pablo por todos lados, mandarle un whatsapp o escribirle por Twitter para que le pase una copia de los trabajos que escribió (incluso, muchas veces, no es del todo legal porque las editoriales no siempre permiten a los autores difundir los trabajos que ellos mismos escribieron). ¿Tiene sentido sostener con nuestros impuestos una misma actividad dos veces? La verdad es que resulta inviable para el país, no sólo por el costo económico, sino también por el tiempo que conlleva hacer una investigación doctoral. Además, nosotros pagamos por un conocimiento que luego no nos es accesible porque queda en manos de empresas privadas. ¿Tiene sentido que Pablo, como investigador, no intente hacer lo imposible por conectarse con otros investigadores, compartir su trabajo y nutrirse del feedback de otros investigadores?

Pablo podría haber hecho muchas cosas para mejorar esta situación y evitar estos problemas. En vez de publicar en una revista cerrada o en un libro, podría haber publicado en abierto. De esa manera, su trabajo puede ser reutilizado muchas veces. Todo el mundo puede descargar, leer, citar y comentar su trabajo. Es importante tener en claro que todo esto se paga; lo pagan las instituciones académicas con el dinero que se recauda de los impuestos.

No tiene sentido, como decía antes, volver a pagar por el trabajo que ya se hizo. Una vez que uno hace un trabajo, ya lo tiene, no lo vuelve hacer, sino que lo trata de recuperar. El trabajo ya hecho es importantísimo. Por eso, cuando uno publica hay que fijarse la licencia que uno firma con las editoriales. Uno no puede ceder exclusivamente los derechos a editoriales privadas. Hay que asegurarse de que la licencia permita la lectura y la reutilización de esos materiales para investigaciones, para dar clases, para dar talleres, o lo que sea.

Existen repositorios institucionales en casi todas las instituciones públicas académicas: el CONICET, la CIC, la UNLP, la UNMdP, la UNR, la UNC, la UNCuyo, la editorial de la UBA. En síntesis, hay repositorios institucionales en todo el país y son plataformas ideales para alojar, preservar y difundir el contenido. Uno no tiene que preocuparse de que Google o una base de datos internacional vea lo que uno publicó, sino que, simplemente, lo pone en un repositorio y este se encarga de la difusión. Y cuidado porque no estamos hablando solamente de los artículos en revistas o de la tesis doctoral, sino los datos, el código y los algoritmos.

Existen un montón de plataformas para publicar datos y a la hora de hacerlo es importantísimo que sea en un formato que asegure su uso. Supongamos que trabajo con una plataforma de estadística, al que le compré la licencia, si publico los datos sólo bajo su formato, únicamente lo van a poder utilizar quienes tengan el software con esa licencia. Si publico una planilla de Excel, solamente van a poder ver esos datos quienes tengan Excel o quienes tengan una versión pirata, o quienes tengan alguna herramienta que asegure compatibilidad con los archivos que este genera. Sin embargo, existen muchos formatos abiertos que hacen lo mismo que dicha plataforma, pero que no requieren una licencia paga, por ejemplo, Libreoffice. Incluso, compartir un archivo csv separado por comas es ideal pues no requiere de ningún tipo de licencia para emplearlo.

Además, es importante entender que cuando recopilo datos, estos pueden evolucionar, agregar más variables, o más registros. Por lo tanto, necesitamos

alguna forma de tener una trazabilidad de cómo fue cambiando ese conjunto de datos. No puedo tener una foto de un conjunto de datos, porque si cambian, la foto también. Es importantísimo entender qué versión del dataset, o del conjunto de datos, se utilizó para hacer una investigación. Para eso, existen un montón de repositorios que ya hacen este tipo de cosas, que publican datos, y también existen revistas científicas que se dedican a poner en línea artículos que hablan sobre datos de investigación donde uno describe cómo los recopiló, qué técnicas utilizó para depurarlos, cuáles para filtrarlos, qué software, qué algoritmos. Hay revistas dedicadas a esto. Todas estas técnicas y algoritmos que se utilizaron también deben publicarse. ¿Por qué? Para asegurarnos de que la comunidad académica pueda verificar si el tratamiento que le hicimos a los datos fue el correcto. De vuelta, existen repositorios dedicados a publicar código de programación, por ejemplo. Y existen revistas científicas que se dedican a publicar artículos donde uno narra el uso de estas técnicas de tratamientos, depuración y filtrado de datos, de análisis estadístico. Las hay para todo. Es importante entender que el objetivo de cualquier publicación científica, o de la ciencia en general, es que sea reproducible, que sea repetible y que sea falsable, si no, no es ciencia.

Para tener en cuenta

Es importante entender que el conocimiento se construye sobre nociones previas. No existe un mega genio de la ciencia, ni Einstein, ni Newton. Nadie creó el conocimiento desde cero. Todos se apoyan sobre uno previo, a hombros de gigantes. También, hay que entender que la ciencia es una actividad colaborativa. Uno no puede aislarse y decir “bueno, yo trabajo solo”, o “yo no comparto lo mío con el mundo”, o “a mí no me interesa lo que opinan mis pares”. La ciencia no funciona así; es colaborativa y cada vez esto se incrementa. Todo lo que aprendamos tenemos que compartirlo con la humanidad. No nos podemos quedar con ese conocimiento.

En este sentido, hay que tener en cuenta que estamos en un mundo muy conectado y muy diverso y amplio. Cuando uno publica, debe asegurarse que toda la humanidad tenga acceso a eso, incluso a los pares, personas de otras culturas que pueden interpretar las cosas de otra manera, o que emplean otro idioma, gente con distintos tipos de discapacidades que también tiene que tener acceso al conocimiento. Uno de nuestros objetivos es llegar a todos, al menos a mayor cantidad posible. Obviamente, uno no puede adaptar todo lo que publica a todos los formatos y a todas las herramientas y sistemas del planeta. Para eso, hay que saber apoyarse en un montón de plataformas de servicios especializados, como los repositorios institucionales, repositorios de datos o revistas especializadas. Hay mucho allá afuera para aprovechar para apoyarnos y para no volvernos locos a la hora de publicar algo, tratando de llegar a todo el mundo.

Espero que hayan disfrutado esta presentación. Muchas gracias.

Notas

¹Este artículo es una elaboración sobre una ponencia pronunciada durante la sesión “Una invitación a la ciencia abierta”, el 16 de septiembre de 2022, durante el evento *Psicología con(Ciencia) Abierta*, llevado a cabo en la Facultad de Psicología (UNLP, Argentina) y transmitido en línea [[HTTPS://CONGRESOS.UNLP.EDU.AR/PSICOABIERTA/](https://congresos.unlp.edu.ar/psicoabierta/)].

Agradecimientos

El evento *Psicología (con)Ciencia Abierta* se realizó con el respaldo de una *Community Engagement Grant* otorgada por DORA (The Declaration of Research Assessment).

[Más información disponible en el siguiente link: [HTTPS://SFDORA.ORG/2023/02/16/COMMUNITY-ENGAGEMENT-GRANT-REPORT-PSICOLOGIA-CONCIENCIA-ABIERTA-ARGENTINA-AN-EVENT-TO-ADVANCE-THE-IMPLEMENTATION-OF-OPEN-SCIENCE-PRACTICE-AND-RESEARCH-ASSESSMENT-REFORM-IN-PSYCHOLOGY-AND-THE/](https://sfdora.org/2023/02/16/community-engagement-grant-report-psicologia-conciencia-abierta-argentina-an-event-to-advance-the-implementation-of-open-science-practice-and-research-assessment-reform-in-psychology-and-the/)]