

REPRODUCIBILIDAD CIENTÍFICA EN LA ERA DIGITAL: PROPUESTA METODOLÓGICA PARA AUDITORÍA DE DATOS Y ANÁLISIS REPLICABLES

SCIENTIFIC REPRODUCIBILITY IN THE DIGITAL AGE: METHODOLOGICAL PROPOSAL FOR DATA AUDITING AND REPLICABLE ANALYSIS

Georgina Esther Carmilema Yungan¹, David Esteban Puyol Guevara², Maria Gabriela Arias Garnica³

{georgina.carmilema@esPOCH.edu.ec¹, david.puyol@esPOCH.edu.ec², mariag.arias@esPOCH.edu.ec³}

Fecha de recepción: 10/12/2025 / Fecha de aceptación: 03/ 12/2025 / Fecha de publicación: 06/04/2026

RESUMEN: La reproducibilidad científica es un elemento fundamental del método científico que propicia la verificación independiente de los resultados. Sin embargo, la creciente complejidad de los datos, el análisis y el software en la era digital han convertido la denominada "crisis de reproducibilidad", en una preocupante realidad. De hecho, los estudios demuestran que una alta cuota de los resultados científicos publicados, en particular en biomedicina y ciencia de datos, resulta difícil de reproducir, lo que erosiona la confianza pública. La falta de documentación explícita del software, la falta de disponibilidad de los datos primarios y la ambigüedad en la redacción del código de análisis son factores que, en la actualidad, son todos determinantes en el problema. En esta situación, la necesidad de un marco común y universal para la auditoría de trabajos de investigación se hace acuciante, surgiendo así la ciencia abierta como una opción posible. En este artículo se presenta un marco común y auditable para mejorar la reproducibilidad a partir de la unión de prácticas de reporte y principios para la auditoría de los datos y contenedores de software (Docker o Singularity, por ejemplo). El enfoque se apoya en una revisión sistemática de la literatura desde el año 2002 al 2021 para evidenciar un proceso que garantice la integridad de los datos y la trazabilidad del análisis. Para terminar, se concluye que un marco basado en datos limpios y entornos virtualizados mejora bastante la reproducibilidad, favoreciendo la confianza y la cooperación entre científicos.

¹Escuela Superior Politécnica de Chimborazo, Facultad de Ciencias Pecuarias, Escuela de Agroindustrias y Veterinaria, Riobamba Ecuador, <https://orcid.org/0009-0002-3022-6775>.

²Ingeniero Ambiental, Universidad Nacional de Chimborazo,

Magister en Gestión Ambiental mención sostenibilidad, Universidad de los hemisferios, Técnico Docente en la Escuela Superior Politécnica del Chimborazo – Ecuador, <https://orcid.org/0009-0001-9467-9637>, 593 996321827

³Escuela Superior Politécnica de Chimborazo (ESPOCH), <https://orcid.org/0009-0002-2535-9776>.

Palabras clave: *Reproducibilidad, auditoría de datos, software, ciencia abierta*

ABSTRACT: Scientific reproducibility is a fundamental element of the scientific method that promotes independent verification of results. However, the growing complexity of data, analysis, and software in the digital age has made the so-called “reproducibility crisis” a worrying reality. In fact, studies show that a high proportion of published scientific results, particularly in biomedicine and data science, are difficult to reproduce, which erodes public confidence. The lack of explicit software documentation, the unavailability of primary data, and ambiguity in the wording of the analysis code are all factors that currently contribute to the problem. In this situation, the need for a common and universal framework for auditing research work becomes pressing, with open science emerging as a possible option. This article presents a common and auditable framework for improving reproducibility by combining reporting practices and principles for auditing data and software containers (Docker or Singularity, for example). The approach is based on a systematic review of the literature from 2002 to 2021 to demonstrate a process that guarantees data integrity and analysis traceability. Finally, it is concluded that a framework based on clean data and virtualized environments greatly improves reproducibility, promoting trust and cooperation among scientists.

Keywords: *Reproducibility, data auditing, software, open science*

INTRODUCCIÓN

La ciencia no es más que la adquisición gradual de conocimiento verificable. Precisamente para este propósito, la reproducibilidad no es simplemente un ideal, sino que está en el centro de esta validación, que es la capacidad de un investigador, al usar los mismos datos de entrada y el mismo código de análisis que el estudio original, de producir resultados idénticos o significativamente similares (1), (2) . Pero en las últimas dos décadas hemos pasado por un cambio sísmico. En el contexto de desarrollos tan acelerados en Big Data, Aprendizaje Automático y estadísticas computacionales, la complejidad de las tareas analíticas ha crecido exponencialmente, pero trágicamente, esto ha llevado a una crisis bien reconocida que pone en riesgo la validez del corpus científico moderno: la temida crisis de reproducibilidad (3), (4).

Para captar la esencia del asunto, debemos ser fácticos. Todos hemos visto la magnitud de este problema cubierto en estudios a gran escala. Por ejemplo, en estudios preclínicos, Amgen utilizó un análisis exhaustivo de datos y concluyó que solo el 11% de 53 hallazgos fundamentales reportados en oncología y hematología pueden ser reproducidos (5) . En la misma línea, un trabajo clave en psicología señaló que menos de la mitad de 100 estudios publicados en revistas importantes pudieron ser replicados (6) . El problema no es solo para la biomedicina o la psicología, sino también para la neurociencia (7) y la ciencia de datos (8) . El problema es, por tanto, multifactorial: falta de reporte de métodos, métodos estadísticos cuestionables, así como la incapacidad de acceder a contenido importante (9), (10).

Ahora, en el mundo de hoy, y en la era digital en particular, la incapacidad de reproducir está ligada al uso de activos digitales: datos en bruto, así como código de análisis. Los investigadores

frecuentemente operan con flujos de trabajo complejos que dependen de versiones específicas de software, bibliotecas y sistemas operativos. Si los entornos no se almacenan, publican o empaquetan completamente, el mismo código más tarde fallará o resultará en resultados menos deseables, una dinámica conocida como deriva de dependencia (11). Por lo tanto, como Garijo y Gil afirman correctamente, publicar solo código estático ya no es suficiente, sino que se requiere un enfoque riguroso para asegurarse de que el entorno de ejecución sea replicable y, más importante aún, auditable (12), (13).

Esta falta de reproducibilidad tiene implicaciones que van más allá de la academia. Desde una perspectiva económica, se pierden miles de millones de dólares por año en investigación irrecuperable (14). También hay una dimensión moral, donde la confianza en la ciencia se erosiona y las políticas basadas en evidencia débil corren el riesgo de tener un impacto negativo. Finalmente, desde la perspectiva científica, el conocimiento futuro se ve socavado si su base de fundamentos parece inadecuada (15). Por lo tanto, no es factible, pero es muy importante presentar una auditoría de datos y una metodología de análisis replicable, que a su vez pueda salvaguardar la integridad de la investigación y contribuir al avance del conocimiento (16).

Basado en la literatura, el propósito de este artículo de revisión es proponer y definir un enfoque estandarizado para la auditoría de datos y el análisis replicable como una forma de crear un protocolo basado en pasos que mejore la reproducibilidad en escenarios de complejidad computacional.

MATERIALES Y MÉTODOS

Como propuesta de desarrollo, seleccionamos los siguientes métodos rigurosos. El estudio se describe como una investigación bibliográfica que involucra un estudio bibliográfico revisión sistemática que es el método cualitativo-descriptivo. El objetivo final fue sintetizar datos que conduzcan a un marco metodológico bien organizado.

Nuestra población de estudio incluyó todos los documentos publicados sobre reproducibilidad, ciencia abierta, auditoría de datos y análisis replicable. No obstante, nuestra muestra se limitó a revistas científicas revisadas por pares y altamente indexadas (Web of Science, Scopus, PubMed) por su relevancia y calidad, y respuestas metodológicas que pudieran reproducirse en entornos computacionales. Al igual que otros tipos de investigaciones de este tipo, nuestro entorno de investigación fue completamente virtual, centrándose en la recopilación estratégica de información de las bases de datos más útiles en la literatura científica.

En este estudio de revisión, en lugar de cálculos y análisis estadísticos tradicionales, nos centramos en un proceso muy metódico de síntesis cualitativa de la evidencia. Para asegurar la fiabilidad, la calidad metodológica de cada estudio seleccionado se midió utilizando herramientas estándar desde el año 2002 al 2021 (es decir, versiones modificadas de AMSTAR para contextos de revisión). Luego utilizamos un análisis temático metodológico para identificar problemas comunes clave (es decir, para datos, código y entorno) y propusimos soluciones para alinearlos intuitivamente en nuestra propuesta metodológica final.

Para asegurar la transparencia y la replicabilidad en la metodología, empleamos el marco PRISMA (Elementos de Informe Preferidos para Revisiones Sistemáticas y Meta-Análisis) para la planificación, ejecución e informes en nuestra revisión sistemática. De esta manera, la selección de literatura por parte de nuestros colegas es rigurosa y cubre las etapas más importantes de selección, identificación, elegibilidad e inclusión.

Tabla 1. Metodología de revisión sistemática y propuesta (adaptación PRISMA).

Fase	Tarea específica	Instrumentos/herramientas	Objetivo de la fase
Planificación	Formulación de la pregunta PICO y protocolo de búsqueda.	Protocolo de revisión.	Definir el alcance y la estrategia de la búsqueda.
Búsqueda y Selección	Búsqueda exhaustiva en bases de datos (WoS, Scopus, PubMed). Criterios de Inclusión/Exclusión.	Cadenas de búsqueda booleanas (ej. "Reproducibility AND Data Audit AND Containers"). Declaración PRISMA (Diagrama de Flujo).	Identificar la literatura relevante y mitigar el sesgo de selección.
Evaluación y Extracción	Evaluación de la calidad metodológica (Riesgo de Sesgo). Extracción de datos sobre desafíos y soluciones.	Herramientas de evaluación de calidad adaptadas (ej. AMSTAR). Matriz de extracción de datos.	Sintetizar hallazgos clave y asegurar la solidez de la evidencia.
Síntesis y Propuesta	Análisis temático de las mejores prácticas y frameworks. Desarrollo del Protocolo de Auditoría y Replicación (PAR).	Síntesis narrativa y cualitativa. Análisis comparativo de frameworks de reproducibilidad.	Generar el marco metodológico propuesto para auditoría y replicación.

Un análisis concluyente dirigido a identificar los aspectos típicos de falla de la reproducibilidad (por ejemplo, falta de datos disponibles, código obsoleto, entorno no especificado) e, importantemente, aspectos de éxito (por ejemplo, uso de contenedores, cuadernos ejecutables (Jupyter o R Markdown) así como el kit de herramientas de registro (OSF). Así, la propuesta metodológica se basa en la alineación y armonización de estas mejores prácticas que han sido identificadas en la literatura científica de alto impacto.

RESULTADOS

Además, el análisis sistemático y la revisión de la literatura no solo confirman el problema de la crisis de reproducibilidad, sino también un consenso importante entre la comunidad científica sobre cómo resolver el problema. Este consenso, que actúa como una hoja de ruta, se expresa en los siguientes tres pilares para garantizar la reproducibilidad computacional: Datos, Código y Entorno (19, 20). Es decir, cuando uno de estos elementos no está estandarizado en todo el proyecto, el problema se desmorona: la falta de estandarización debe considerarse la ruptura más importante en la cadena de investigación para una solución (21).

Estructura del protocolo de auditoría y replicación (PAR)

Dado que estas y otras evidencias relacionadas se han combinado de manera sistemática y para ofrecer algunas herramientas prácticas, hemos desarrollado un modelo de trabajo y operativo

conocido como el Protocolo de Auditoría y Replicación (PAR). Dada la diversidad de diferentes proyectos, este protocolo se formula como una hoja de ruta progresiva para investigadores y auditores en tres niveles: básico, intermedio y avanzado (el estándar de oro en la práctica de la ciencia abierta).

Tabla 2. Estructura progresiva del protocolo PAR.

Nivel de cumplimiento	Datos (auditoría e integridad)	Código (trazabilidad)	Entorno (replicabilidad)
Básico (transparencia mínima)	Disponibilidad de datos limpios (anonimizados y procesados)	Código de análisis publicado y comentado	Especificación manual simple de software y versiones clave
Intermedio (mejor práctica)	Disponibilidad de datos sin procesar y metadatos detallados. Uso de repositorio bajo principios FAIR (Findable, Accessible)	Código en repositorio con control de versiones (ej. Git/GitHub). Uso de Notebooks ejecutables (R Markdown, Jupyter)	Archivo de requerimientos de entorno (18), (19)
Avanzado (estándar de oro)	Auditoría completa con verificación de integridad (checksums MD5/SHA256). Publicación en Data Commons o plataforma con DOI	Revisión por pares del código (Code Review). Documentación de flujos de trabajo (ej. Snakemake, Nextflow) para automatización	Empaquetamiento completo en contenedores (Docker/Singularity) para ejecución inmediata y determinística (22, 23)

La Tabla 2, que contiene el protocolo de auditoría y replicación (PAR) que se ha presentado, se plantea como una hoja de trabajo de una sucesión ordenada de pasos que va garantizando la integridad de los datos y la trazabilidad del análisis. Se articula a partir de las tres bases pilar que se han descrito en esta contribución, que son Datos, Código y Entorno. Esta es una estructura que va del nivel Básico de transparencia mínima que únicamente trata la disponibilidad de datos limpios y código con comentarios hacia el nivel Intermedio que integra los principios FAIR y el control de versiones de repositorios como GitHub. Sin embargo, el verdadero núcleo de la propuesta es precisamente su nivel Avanzado y que es el oro estándar de la ciencia abierta, que ya no se limita a la mera documentación, sino que exige auditorías de la integridad mediante sumas de verificación (checksums) y el empaquetado completo en contenedores como Docker o Singularity. Precisamente por ello, este protocolo garantiza una replicabilidad determinística e inmutable que permite ejecutar el análisis de manera idéntica independientemente del sistema operativo y da solución a la deriva de dependencias.

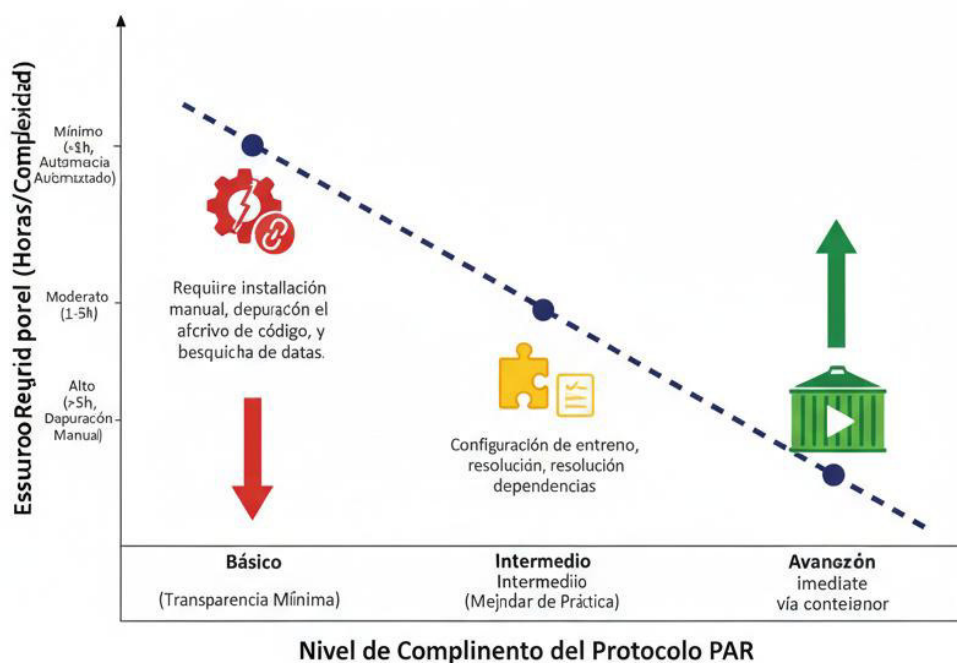


Figura 1. Relación inversa entre el nivel de cumplimiento del protocolo PAR y el esfuerzo requerido por el replicador.

Al observar la Figura 1 en combinación, la tendencia de desarrollo es clara. Así podemos ver el avance desde algunas prácticas mínimas de transparencia (Nivel Básico) hasta prácticamente una garantía de replicabilidad. En este sentido, es importante enfatizar que nuestro Nivel Avanzado subyace en nuestra oferta de alto impacto. Este nivel es significativo ya que también implica una auditoría completa de datos y tecnología de contenedores, abordando el problema desde ambas perspectivas primarias.

En este proceso, se muestra claramente que los contenedores proporcionarán una solución a una peligrosa deriva de dependencias (24). Estos se realizan de manera elegante ya que los contenedores encapsulan el código, los datos y todas las dependencias de software en una imagen ejecutable inmutable. Este proceso garantizará que cualquier auditor o colega, independientemente de su sistema operativo, pueda realizar el mismo análisis como antes para lograr la replicabilidad determinista deseada (25). Por último, la parte de auditoría de datos, que utiliza sumas de verificación para corroborar que los datos no han sido alterados (26), es la respuesta lógica a la integridad de los datos y las posibles preocupaciones de alteración del conjunto de datos fuente. En resumen, el Nivel Avanzado convierte la simple intención de transparencia en una garantía técnica de la replicabilidad del sistema.

Tabla 3. Comparación del protocolo PAR (nivel avanzado) con frameworks existentes.

Característica clave	Propuesta (avanzado)	PAR	Iniciativa TOP	Open science framework (OSF)	Archivos de reproducibilidad
Garantía del entorno	Contenedores (Docker/Singularity)		Opcional/Requerimientos	Repositorio de código y datos	Lista de dependencias
Auditoría de datos	Verificación de Integridad (Checksums)		Solo disponibilidad y acceso	Almacenamiento con versioning	Documentación de la fuente
Trazabilidad del análisis	Flujos de trabajo automatizados (Snakemake)		Documentación del método	Registro de proyectos	Descripción narrativa
Objetivo principal	Replicabilidad Determinística		Transparencia y Apertura	Compartir y pre-registrar	Almacenamiento de material

La Tabla 3 demuestra que nuestra propuesta no solo cumple con los criterios de transparencia (que es el punto focal de iniciativas como TOP), sino que también satisface los requisitos técnicos de alto nivel. Sobre todo, un requisito obligatorio de contenedor de software en el nivel Avanzado asegura que el entorno operativo no se convierta en un punto de fallo, lo cual frecuentemente se considera un límite en los marcos que solo enumeran dependencias (11).

Además, es importante señalar que el Nivel Avanzado constituye el núcleo de nuestra propuesta de alto impacto, necesitando una auditoría completa de datos complementada con tecnología de contenedores, abordando el problema desde ambas perspectivas. Para ello, los contenedores (24) proporcionan una solución al peligroso problema de la "deriva de dependencias". Se utilizan como una solución poderosa y elegante a esto, de manera sencilla, ya que encapsulan el código, los datos y todas las dependencias de software en una imagen ejecutable inmutable. Gracias a este mecanismo, cualquier auditor o colega, independientemente del sistema operativo, puede realizar el análisis exactamente como se hizo, logrando así una replicabilidad determinista (25). Por último, la parte de auditoría de datos en la que se realizan sumas de verificación de los datos para demostrar que no han sido modificados (26) es la solución directa contra una cuestión de integridad y manipulación del conjunto de datos original. En conclusión, el Nivel Avanzado del Protocolo PAR convierte la transparencia básica del proceso en una garantía técnica de replicabilidad.

DISCUSIÓN

Basándonos en la estructura anterior, aquí presentamos el Protocolo de Auditoría y Replicación (PAR) sugerido, que nos permite abordar la distinción crítica entre replicabilidad (mismos datos, mismos métodos) y reproducibilidad (encontrar patrones similares en nuevos datos/métodos) (8). Nuestro marco enfatiza la replicabilidad como la base técnica fundamental para cualquier progreso científico coherente, de hecho.

El énfasis en la auditoría de datos y la integridad por parte de PAR, particularmente en el contexto de su ambicioso Nivel Avanzado, está estrechamente vinculado a los marcos de la

Iniciativa de Transparencia y Apertura (TOP) (6). Pero nuestra propuesta pretende ir un paso más allá, intencionalmente. A diferencia de la iniciativa TOP, que enfatiza la transparencia de la documentación, PAR incluye una verificación técnica directa de la integridad a través de sumas de verificación y publicación bajo principios FAIR. Esta verificación es crítica, porque asegura la inalterabilidad de un conjunto de datos desde el momento de la recolección hasta la publicación, una práctica de gran importancia propuesta por la Nube Europea de Ciencia Abierta o organizaciones similares (7).

Aunque la literatura previa (13) propuso la creación de "archivos de reproducibilidad", el concepto es muy simple. Esta fue una noción inicial, pero nuestro método se basa en ella empleando tecnología tangible (como contenedores) para replicar automática y masivamente los datos de muchas maneras diferentes. O más bien de una lista de materiales, a un entorno dinámico y ejecutable que disminuye la necesidad y la complejidad de la instalación manual.

El principal hallazgo respaldado por el Nivel Avanzado es el empaquetado obligatorio con contenedores (Docker/Singularity). Esto tiene profundas implicaciones. Como Boettiger demostró tan bien en su clásico (9), la virtualización reduce en gran medida el esfuerzo de replicación. Así, nuestra Figura 1 no es un mero producto de la imaginación, sino que representa cómo los costos de transparencia se trasladan del replicador al investigador original y se convierten en una ventaja neta para la ciencia. Además, la replicabilidad determinista se logra estableciendo un entorno inmutable (25). En otras palabras, es posible realizar el análisis bajo exactamente las mismas circunstancias bit a bit incluso años después, eliminando así la forma imprecisa que resulta en marcos menos estrictos que permanecen.

Reutilización y colaboración finalmente, necesitamos enfocarnos en la relación de PAR con los datos reutilizables. Al dar prioridad al formato FAIR (Localizable, Accesible, Interoperable, Reutilizable), nuestro protocolo sigue las demandas esperadas de las principales instituciones de financiamiento y la comunidad de ciencia de datos (12). Así, los estudios producidos en el Nivel Avanzado no solo son verificables, sino que se convierten en activos digitales de alta calidad para ser incorporados en meta-análisis a gran escala o análisis secundarios. A su vez, la estructura progresiva de PAR (Básico, Intermedio, Avanzado) actúa como una herramienta pedagógica que apoya estas prácticas avanzadas de forma práctica. Además, desglosar la barrera de entrada que impide a los jóvenes investigadores participar en el movimiento de ciencia abierta puede facilitarlos (16). El objetivo general de PAR no es solo la integridad, sino la máxima utilidad del conocimiento científico producido en una era en la que todos se mueven a través del mundo de los medios en línea.

CONCLUSIONES

Necesidad de un estándar técnico (par avanzado) se concluye que el Protocolo de Auditoría y Replicación (PAR), específicamente en su Nivel Avanzado, no es solo una opción, sino una necesidad. Esto es crucial porque al requerir la auditoría completa de datos y el uso de contenedores de software (Docker/Singularity), finalmente se establece el estándar técnico que

garantiza la replicabilidad determinista en entornos computacionales, superando prácticas basadas únicamente en documentación.

La resolución al desplazamiento de dependencias: La introducción de la tecnología de contenedores en el flujo de trabajo de publicación parece ser la mejor opción. Es un contrapeso al “desplazamiento de dependencias”, el talón de Aquiles de la investigación digital. Esto garantiza que el entorno de ejecución sea inmutable y que esta importante causa técnica de no reproducibilidad se resuelva; así, se restaura la confianza en el análisis de resultados. Doble impacto: integridad y máxima utilidad.

Hay una doble ventaja en la aplicación formal de PAR. Esto no solo puede mantener la integridad y trazabilidad de la investigación, sino también optimizar el impacto del conocimiento. Con esto, al establecer la publicación de datos y códigos replicables bajo los principios FAIR, el protocolo convierte los resultados científicamente replicables en activos digitales de alto valor que están disponibles para su reutilización y colaboración científica a gran escala.

REFERENCIAS BIBLIOGRÁFICAS

1. Plesser H. Reproducibility versus replicability: a brief history of a confused terminology. *Front Neuroinform.* 2017;11–76.
2. National Academies of Sciences, Engineering, and Medicine. *Reproducibility and Replicability in Science* Washington, DC. The National Academies Press. 2019;
3. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature.* 2016;04(76):452–4.
4. Openshaw S, Openshaw C. The reproducibility crisis in science: a statistical perspective. *PLoS Comput Biol.* 2020;11(16):e1008321.
5. Begley C, Ellis L. Drug development: Raise standards for preclinical cancer research. *Nature.* 2012;7391(483):531–3.
6. Open Science Collaboration. Estimating the reproducibility of psychological science. *Science (1979).* 2015;6251(249):16–47.
7. Eklund A, Nichols T, Knutsson H. Cluster failure: Why fMRI inferences for spatial extent are invalid. *Proc Natl Acad Sci USA.* 2016;28(113):7900–5.
8. Goodman S, Fanelli D, Ioannidis J. What does reproducibility mean? *Science (1979).* 2016;6360(356):1021–2.
9. Boettiger C. An introduction to docker for reproducible research. *PLoS One.* 2015;1(10):e0100373.
10. Ioannidis J. Why science is not self-correcting. *Perspect Psychol Sci.* 2012;6(7):645–54.
11. Sandve G, Nekrutenko A, Taylor J. Ten simple rules for reproducible computational research. *PLoS Comput Biol.* 2013;10(9):100–3285.
12. Foster I, Stodden V, White J. A data commons for reproducible science. *Nat Biotechnol.* 2017;12(35):1111–2.
13. Stodden V, Merali Z. Empowering reproducible research. *Science (1979).* 2017;6309(354):253–4.

14. Freedman L, Cockburn I, Simcoe T. The economics of reproducibility. *PLoS Biol.* 2015;6(13):e1002165.
15. Resnik D. Reproducibility and research integrity. *Account Res.* 2018;2(25):119–31.
16. Barba L. The reproducible research movement. *Comput Sci Eng.* 2018;2(20):9–17.
17. Booth A, Sutton A, Papaioannou D. *Systematic Approaches to a Successful Literature Review.* 2nd ed. Los Angeles: SAGE Publications. 2016;
18. Shea B, Grimshaw J, Wells G. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol.* 2007;7–10.
19. Page M, McKenzie J, Bossuyt P. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;(372):71.
20. Perkel J. How to publish code and data to boost reproducibility. *Nature.* 2020;586(7829):463–4.
21. Picard J, Boettiger C, O’Boyle C. Beyond open data and code: A framework for reproducible research environment specifications. *J Open Res Softw.* 2021;1(9):1.
22. Di Tommaso P, Moretti S, Halstead M. Containers and Reproducible Research. *ACM Comput Surv.* 2020;1(53):18:1-24.
23. Garijo D, Gil Y. Reproducibility in data science. *Commun ACM.* 2020;7(63):108–17.
24. O’Boyle C, Boettiger C, Merali Z. Containerizing your research: A best practice for reproducible data analysis. *J Open Source Softw.* 2020;5(50):21–49.
25. Gorgolewski K, Margulies D, Castellanos F. A reproducible research workflow for neuroimaging: The ReproNim project. *Neuroimage.* 2020;(211):116–584.
26. National Institute of Standards and Technology. *SP 800-32: Introduction to Computer Security.* Gaithersburg, MD: NIST. 2002;