8TH INTERNATIONAL CONGRESS
53ª JORNADAS NACIONALES
SEVILLA 15/17 OCTUBRE 2026
SOCIDROGALCOHOL
Sociedad Científica Española de Estudios sobre el Alcohol, el Alcoholismo y las otras Toxicomanías

Trazabilidad de los datos en adicciones:
vinculación de datos brutos, material suplementario y publicaciones científicas

Liu, Yiming[1,2], Vidal-Infer, Antonio[1,2,3], Rius, Cristina[1,2,4,5], Sanz-de-Ramón, Aitana[1,2], Cirilo, Ramón[6], Lucas-Domínguez, Rut[1,2,5]

1. Universitat de València. Unidad de Información e Investigación Social y Sanitaria (Grupo UISYS). Unidad asociada al Instituto Interuniversitario de Investigación Avanzada sobre Evaluación de la Ciencia y la Universidad (INAECU) UC3M-UAM. España. 2. Departamento de Historia de la Ciencia y Documentación. Facultad de Medicina y Odontología. Universitat de València. España. 3. Red de Investigación de Adicciones en Atención Primaria (RIAPAd). España. 4. CIBERCV. España. 5. CIBERONC. España. 6. Instituto de Robótica y TIC (IRTIC). Universitat de València. España.

Introducción

Compartir los datos de investigación es un pilar de la ciencia abierta que fortalece la integridad y la reproducibilidad científica. En el campo de las adicciones, resulta clave para acelerar los descubrimientos y mejorar la práctica clínica, aunque sigue siendo poco frecuente debido a la sensibilidad de los datos, la falta de incentivos y la competencia. Los principios FAIR fomentan la calidad y reutilización de los datos (Shreeve, Hurley y Adams, 2025).

Objetivos

El presente estudio evalúa los datos compartidos sobre adicciones, analizando el material suplementario a las publicaciones científicas indexadas en la base de datos bibliográfica Web of Science Core Collection (WoSCC) y los datasets depositados en repositorios utilizando Data Citation Index (DCI) y comparando los años 2014 y 2023. Paralelamente se analizó la procedencia y trazabilidad de los datos de investigación mediante la vinculación entre publicaciones científicas y datos brutos depositados.

Metodología

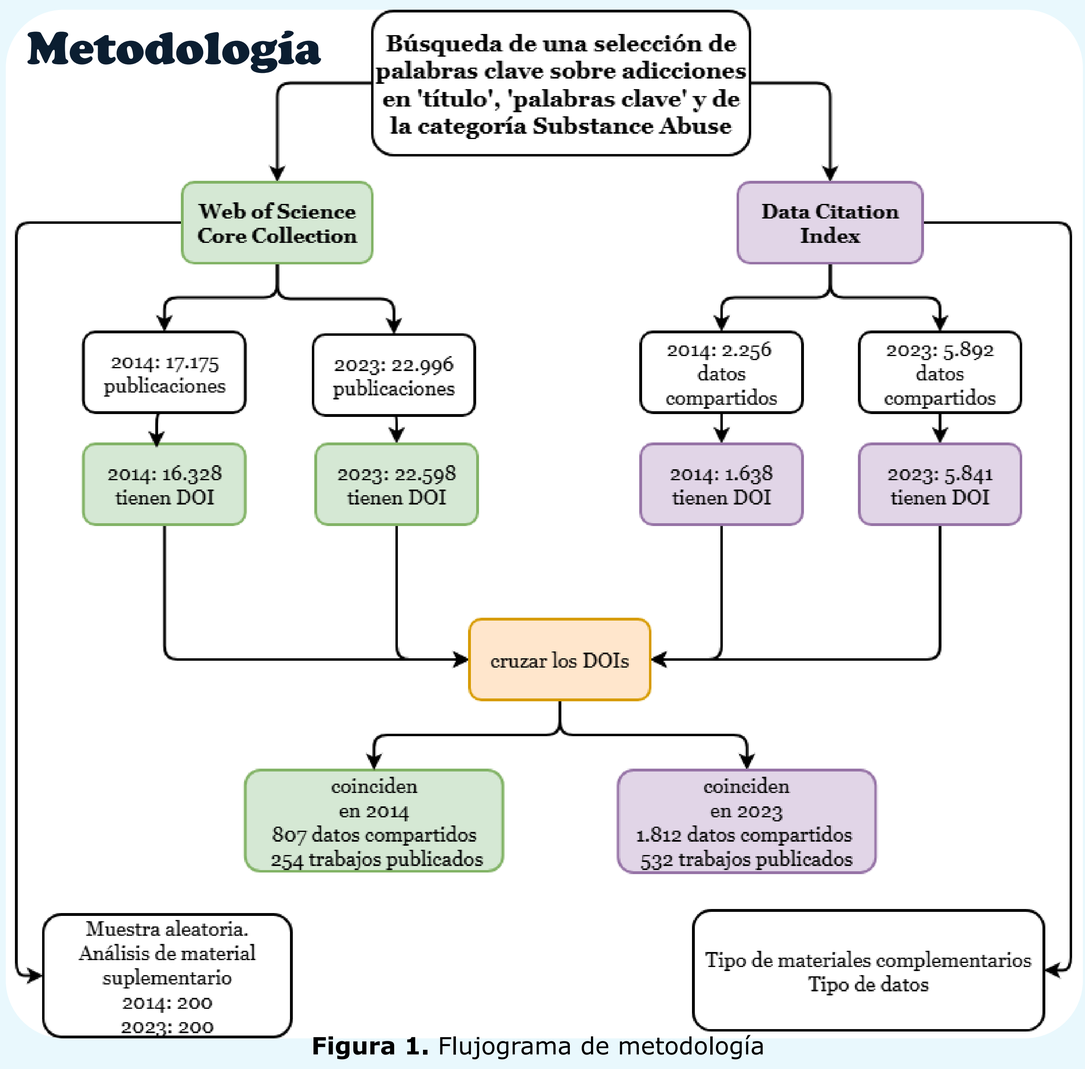


Figura 1. Flujograma de metodología

Resultados

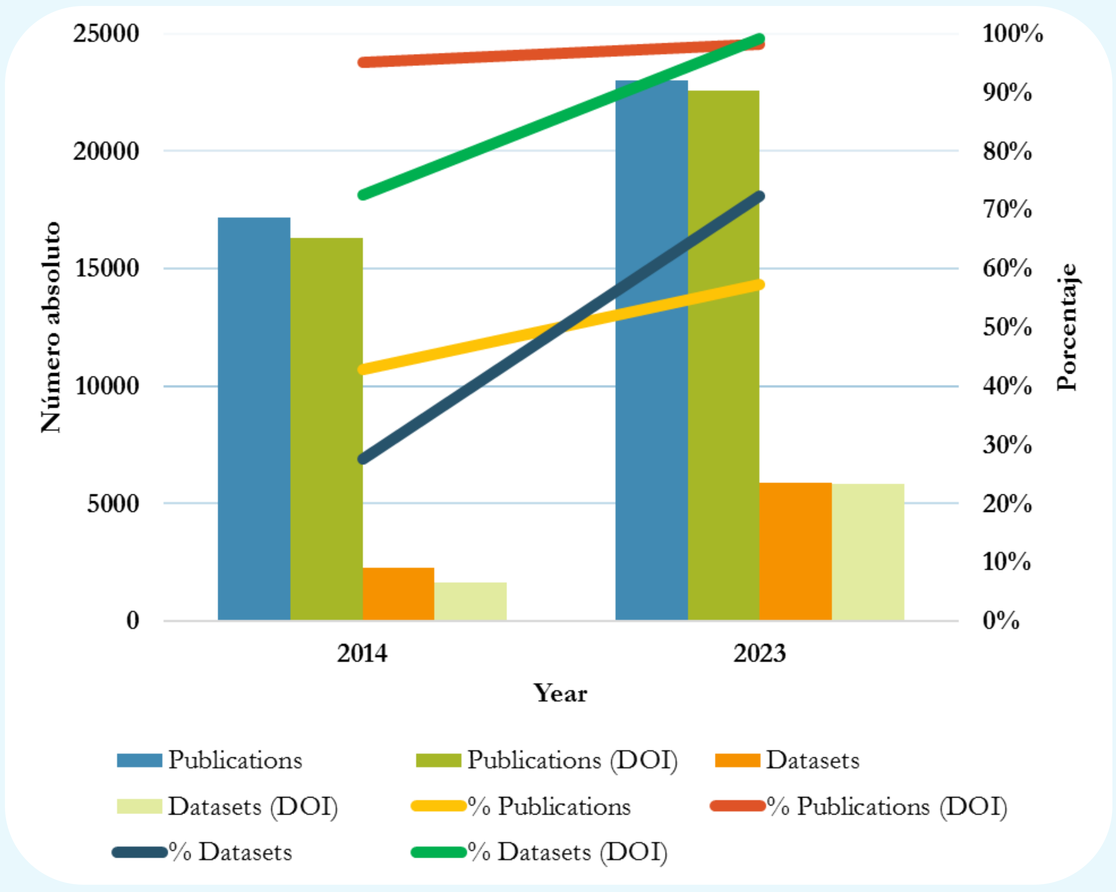


Figura 2. Comparación entre número de publicaciones y datos compartidos entre 2014 y 2023

Se obtuvieron 40.171 registros en WoSCC y 8.148 datasets en DCI. El porcentaje de publicaciones aumentó del 42,75% al 57,25% y, el de registros de datos brutos depositados en repositorios se incrementó casi tres veces pasando de 27,69% a 70%. El cruce de los identificadores DOI permitió reconocer 2.619 registros de datos (807 en 2014 y 1.812 en 2023) vinculados a 786 publicaciones científicas (254 y 532, respectivamente), lo que representa el 32,14% del total de registros de datos compartidos y el 1,96% del total de publicaciones (Figura 1, 2).

Tipos de materiales complementarios / Año	2014	% 2014	2023	% 2023
DOC/DOCX	**26**	**48,15%**	**54**	**44,26%**
PPT/PPTX	**10**	**18,52%**	1	0,82%
JPEG/JPG/TIF/TIFF/PNG/GIF	**9**	**16,67%**	6	4,92%
PDF	7	12,96%	32	**26,23%**
XLS/XLSX	2	3,70%	8	6,56%
html	0	0,00%	2	1,64%
csv	0	0,00%	10	**8,20%**
txt/pl	0	0,00%	1	0,82%
r	0	0,00%	8	6,56%
TOTAL	**54**		**122**	

Tabla 1. Número y porcentaje de los tipos de materiales suplementario

Tipo de datos / Año	2014	% 2014	2023	% 2023	Total general
dataset general	113	14,91%	**3111**	**56,08%**	**3224**
gráfico	9	1,19%	2059	37,12%	2068
RNA	**272**	**35,88%**	0	0,00%	**272**
datos genómicos	130	17,15%	3	0,05%	133
datos numéricos	9	1,19%	99	1,78%	108
datos de encuesta	58	7,65%	16	0,29%	74
software	13	1,72%	52	0,94%	65
multimedia	0	0,00%	63	1,14%	63
datos agregados; datos de encuesta	7	0,92%	36	0,65%	43
datos transcriptómicos	16	2,11%	10	0,18%	26
datos administrativos	21	2,77%	4	0,07%	25
imágenes	2	0,26%	21	0,38%	23
datos brutos; datos procesados	20	2,64%	0	0,00%	20
colección de archivos	19	2,51%	0	0,00%	19
colección de datasets	0	0,00%	14	0,25%	14
datos procedados	10	1,32%	0	0,00%	10

El análisis de una muestra aleatoria de 200 publicaciones por año mostró un aumento en la disponibilidad de material suplementario. Los tipos de archivos más frecuentes fueron de texto DOC/DOCX (48,15% en 2014 y 44,26% en 2023) (Tabla 1). Se identificaron los principales tipos de datos brutos derivados del análisis de RNA (35,88%) en 2014. En cambio, los de 2023 fueron conjuntos de datos generales (56,08%) (Tabla 2).

Tabla 2. Número y porcentaje de tipos de datos compartidos de DCI

Conclusión

- Entre 2014 y 2023 aumentaron tanto las publicaciones científicas como los datos compartidos en investigación sobre adicciones.
- Las publicaciones aumentaron un 14,95% y los datos compartidos casi se triplicaron.
- La vinculación mediante DOI permitió relacionar los datasets disponibles en las publicaciones científicas sobre adicciones, poniendo de manifiesto un escaso uso compartido de datos de investigación.
- Casi la mitad de material suplementario estaba en formato propietario DOC/DOCX, lo que sugiere bajo nivel de cumplimiento de los principios FAIR

Referencias

Shreeve, K.N., Hurley, R.W., y Adams, M.C. (2025). Navigating the new NIH Data Management Sharing plan: the NIH HEAL Initiative Data2Action DMPTool template. *JAMIA open, 8*(3), ooaf040. https://doi.org/10.1093/jamiaopen/ooaf040

Financiación

Yiming Liu. Programa de Formación Predoctoral de la Generalitat Valenciana (CIACIF/2023/316).