¿Cómo preparamos nuestros repositorios para un ecosistema transformado por la inteligencia artificial?

Francisco José García-Peñalvo
GRupo de investigación en InterAcción y eLearning (GRIAL)
Departamento de Informática y Automática
Instituto Universitario de Ciencias de la Educación (IUCE) (https://ror.org/00xnj6419)
Universidad de Salamanca (https://ror.org/02f40zc51), Salamanca, España
fgarcia@usal.es
http://twitter.com/frangp
https://orcid.org/0000-0001-9987-5584

Resumen

En la conferencia «Repositorios digitales: de la infraestructura al liderazgo de la Ciencia Abierta en América Latina», organizada por ICI Academy el 11 de noviembre de 2026, Francisco José García-Peñalvo, de la Universidad de Salamanca y del Grupo GRIAL, presentó la intervención de once minutos «¿Cómo preparamos nuestros repositorios para un ecosistema transformado por la inteligencia artificial?». Su tesis central fue que los repositorios deben evolucionar desde infraestructuras de depósito hacia servicios confiables, interoperables y gobernables, capaces de responder a las dinámicas de la inteligencia artificial.

Cambian simultáneamente los mecanismos de descubrimiento, los consumidores de información y las cargas técnicas. Los buscadores generativos sintetizan respuestas y pueden invisibilizar las fuentes; los agentes automatizados procesan colecciones a gran escala; y los rastreadores comprometen la disponibilidad de los servidores. Según una encuesta de COAR de 2025, más del 90 % de los 66 repositorios participantes reportaron problemas con rastreadores agresivos. Frente a ello, se propone una defensa escalonada, con cortafuegos de aplicaciones, limitación de peticiones y canales específicos para accesos masivos legítimos.

Un eje sensible es la reutilización de contenidos abiertos para entrenar modelos comerciales. El acceso abierto no resuelve por sí solo las condiciones jurídicas de explotación. La ponencia examinó, como referencia, los artículos 3 y 4 de la Directiva europea 2019/790 sobre minería de textos y datos, así como el artículo 53 del Reglamento europeo de inteligencia artificial. Planteó reservas de derechos legibles por máquina y mecanismos de protección de las bases de datos, sin confundir la apertura del conocimiento con la extracción indiscriminada de valor.

En el plano técnico, defendió una secuencia de preparación: primero, identificadores persistentes, metadatos de calidad, licencias y procedencia; después, interoperabilidad mediante OAI-PMH, JSON-LD, Signposting y COAR Notify; posteriormente, preservación y confianza; y, finalmente, aplicaciones como la búsqueda semántica, la generación aumentada por recuperación y la curación asistida. El principio rector es que la IA sugiera, el repositorio valide y una persona decida, manteniendo la trazabilidad y la independencia tecnológica.

Las recomendaciones combinaron los principios FAIR y CARE para equilibrar la reutilización automatizada y los derechos de las comunidades, con especial atención a las

lenguas, las disciplinas y los saberes locales. Como primeros pasos, propuso auditar los metadatos, validar la recolección, aprobar una política institucional de IA y medir el tráfico automatizado. Antes de desplegar sistemas inteligentes, pidió evaluar la calidad, los costes, los riesgos y la reversibilidad. El mensaje final sintetizó la orientación estratégica: repositorios preparados para la IA, pero no dependientes de ella; abiertos al conocimiento, no al extractivismo.

Palabras clave

Repositorios institucionales; Ciencia abierta; Inteligencia artificial; Gobernanza de datos; Extractivismo digital

Enlace a la presentación

https://zenodo.org/record/24136235

DOI

10.5281/zenodo.24136235

Cita recomendada

F. J. García-Peñalvo, "¿Cómo preparamos nuestros repositorios para un ecosistema transformado por la inteligencia artificial?," presentado en Repositorios digitales: de la infraestructura al liderazgo de la Ciencia Abierta en América Latina (ICI Academy, 11 de noviembre de 2026). Salamanca, España: Grupo GRIAL, 2026. Disponible: https://d66z.short.gy/IjjFA0. doi: 10.5281/zenodo.24136235.

Referencias

[1] COAR, "AI and Repositories," Confederation of Open Access Repositories. Disponible en: https://coar-repositories.org/what-we-do/ai-and-repositories/.
[2] European Parliament y The Council of the European Union, "Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market," Official Journal of the European Union, L 130, 2019. Disponible en: https://eur-lex.europa.eu/eli/dir/2019/790/oj.
[3] European Parliament y The Council of the European Union, "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act)," Official Journal of the European Union, 2024. Disponible en: https://eur-lex.europa.eu/eli/reg/2024/1689/oj.
[4] "Protecting University Repositories from Aggressive Web Scraping: Using Database Rights to Retain Control over Academic Content," Kluwer Copyright Blog. Disponible en: https://legalblogs.wolterskluwer.com/copyright-blog/protecting-university-repositories-from-aggressive-web-scraping-using-database-rights-to-retain-control-over-academic-content/.
[5] "Tensions Between Digital Libraries and Generative AI," Project MUSE. Disponible en: https://muse.jhu.edu/article/992004.
[6] "Generative AI and intellectual property: a systematic literature review of possible issues and mitigations." Disponible en: https://www.researchgate.net/publication/408197294.

[7] COAR, "Report of the COAR Survey on AI Bots," June 2025. Disponible en: https://coar-repositories.org/wp-content/uploads/2025/06/Report-of-the-COAR-Survey-on-AI-Bots-June-2025-1.pdf.
[8] "Mitigating Aggressive Crawler Traffic in the Age of Generative AI," The Code4Lib Journal. Disponible en: https://journal.code4lib.org/articles/18489.
[9] LibLynx, "Bot wars: Will AI scraping change the Scholarly landscape forever?" Disponible en: https://www.liblynx.com/news/bot-wars-will-ai-scraping-change-the-scholarly-landscape-forever.
[10] COAR, "Mitigating the impact of AI bots." Disponible en: https://coar-repositories.org/news-updates/mitigating-the-impact-of-ai-bots/.
[11] COAR, "USRN Discovery Pilot: Outcomes and Next Steps." Disponible en: https://coar-repositories.org/news-updates/usrn-discovery-pilot-outcomes-and-next-steps/.
[12] M. D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship," Scientific Data, vol. 3, art. 160018, 2016. doi: 10.1038/sdata.2016.18.
[13] D. Lin et al., "The TRUST Principles for digital repositories," Scientific Data, vol. 7, art. 144, 2020. doi: 10.1038/s41597-020-0486-7.
[14] "Generative Engine Optimization in digital repositories," Infonomy. Disponible en: https://infonomy.scimagoepi.com/index.php/infonomy/en/article/view/115.
[15] Digital Preservation Coalition, "AI-powered metadata workflows at the ICAEW Digital Archive." Disponible en: https://www.dpconline.org/blog/ai-powered-metadata-workflows-at-icaew-craig-mccarthy.
[16] S. R. Carroll et al., "The CARE Principles for Indigenous Data Governance," Data Science Journal, vol. 19, art. 43, 2020. doi: 10.5334/dsj-2020-043.
[17] ESIP, "Are Your Data Ready? Take Stock With ESIP's New AI-Ready Checklist." Disponible en: https://www.esipfed.org/checklist-ai-ready-data/.
[18] "Bridging FAIR and CARE in ETD Metadata: An LLM assisted Cross [...]." Disponible en: https://dcpapers.dublincore.org/article/952576294.