

## Repositorios del futuro

Dr. Francisco José García-Peñalvo

Profesor del Departamento de Informática y Automática de la Universidad de Salamanca (USAL)

Director del Grupo GRIAL de la USAL y de la Unidad de Investigación Consolidada de la Junta de Castilla y León

[fgarcia@usal.es](mailto:fgarcia@usal.es)

### Resumen

Intervención en la mesa “Repositorios del futuro”, celebrada en el congreso **Ecosistemas del Conocimiento Abierto (ECA 2017)** en Salamanca (España) el 27 de octubre de 2017.

### Palabras clave

Repositorios institucionales; Experiencia de usuario; Ecosistema tecnológico; Inteligencia artificial

### Referencia recomendada

García-Peñalvo, F. J. (2017). *Repositorios del futuro*. Paper presented at the Ecosistemas del Conocimiento Abierto (ECA 2017), Salamanca, España. doi:10.5281/zenodo.1042121

### Posicionamiento

La mesa de debate surge como respuesta al informe de COAR “Next Generation Repositories” (Rodrigues et al., 2017).

Para pensar en los repositorios institucionales del futuro, primero deberíamos plantearnos los problemas de los repositorios del presente (Ferrerías-Fernández & Merlo-Vega, 2015; Ferrerías-Fernández, Merlo-Vega, & García-Peñalvo, 2013). Los repositorios a nivel técnico han experimentado un avance muy importante en cuanto a interoperabilidad, recolección, intercambio y ayuda a la promoción del acceso abierto mediante la ruta verde (García-Peñalvo, García de Figuerola, & Merlo-Vega, 2010; García-Peñalvo, Merlo-Vega, et al., 2010; García-Peñalvo & Tena-Espinoza-de-los-Monteros, 2017). Son una herramienta muy conocida en el ámbito bibliotecario, pero son unos auténticos desconocidos para los investigadores (García-Peñalvo, 2017a, 2017b) y aquellos que se “atreven” a acercarse, vuelven a alejarse porque los procesos de autoarchivo son pesados y poco amigables y las interfaces de búsquedas totalmente alejadas de la realidad que les ofrecen otros servicios que usan (en el mejor de los casos) o de la simplicidad de utilizar el buscador Google (González-Pérez, Ramírez-Montoya, & García-Peñalvo, 2016b, 2017a).

Las interfaces y los sistemas de clasificación de los contenidos están diseñados por y para profesionales de la catalogación, es decir, centrados en una estructura de colecciones muy estática y normalmente desconocida fuera del ámbito de decisión, y orientada a envolver al recurso final bajo una capa de metadatos extremadamente rica, pero orientada a ser alimentada de forma manual, lo que crea enormes cuellos de

botella cuando se quieren digitalizar grandes cantidades de documentos y conlleva numerosos errores (que repercuten en la calidad del dato) cuando se abre el auto-archivo a los investigadores.

Todo ello hace que el repositorio como tal cumpla una misión imprescindible en difusión de la Ciencia del siglo XXI, como garantes de la preservación de la producción científica (especialmente de la literatura gris (Ferrerías-Fernández, 2016; Ferrerías-Fernández, García-Peñalvo, Merlo-Vega, & Martín-Rodero, 2016; Ferrerías-Fernández, Martín-Rodero, García-Peñalvo, & Merlo-Vega, 2016)) y de la difusión por el medio digital, de la que se benefician terceros servicios (Google, Google Scholar, ResearchGate, etc.) y desde estos llegue el efecto a los perfiles digitales de los investigadores (métricas, visibilidad, citas recibidas, etc.). Pero a su vez, los repositorios en su estado actual están completamente alejados de los usuarios.

Para los repositorios del futuro, partiendo del hecho de que se haya diseñado una estrategia de formación, información, puesta en valor, etc. de los repositorios, incluso más allá de estos, hacia una necesidad de definir el perfil digital de los investigadores (Tena-Espinoza-de-los-Monteros, García-Holgado, Merlo-Vega, & García-Peñalvo, 2017), habría que poner el centro de atención en los usuarios finales y en maximizar su experiencia de usuario a través de flujos de trabajo bien definidos, claros, simples y que se basen en servicios de valor añadido, esto es, que acabe encontrando en el repositorio una visión integral del proceso de Ciencia 2.0 (Shneiderman, 2008), con un doble flujo, el que le ayuda a construir su identidad digital como investigador-autor, y el que le permite sacar provecho en un contexto de *discovery* (González-Pérez, Ramírez-Montoya, & García-Peñalvo, 2016a): descubrir trabajos/investigador relacionados, establecer relaciones, recibir recomendaciones, realizar búsquedas avanzadas para estudios sistemáticos (González-Pérez, Ramírez-Montoya, & García-Peñalvo, 2017b, 2018).

Para conseguir este acercamiento al usuario, el repositorio debe perder su protagonismo para ser una parte más de un ecosistema tecnológico (García-Holgado & García-Peñalvo, 2013, 2016) de ciencia abierta (Crouzier, 2015; García-Holgado & García-Peñalvo, 2017b), es decir, el repositorio como tal es un componente más del ecosistema, especializado en el contenido científico (García-Holgado, García-Peñalvo, & Rodríguez-Conde, 2015), y que interopere con el resto de los servicios que consiguen darle valor al usuario, aislándole de la estructura interna del contenido (García-Holgado & García-Peñalvo, 2017a, 2017c, 2017d).

Por último, para potenciar esos servicios de valor añadido para el usuario, debe haber una capa interna basada en inteligencia artificial que ayude a automatizar procesos para mejorar los servicios y conseguir más valor añadido en ellos, eliminando cuellos de botella en procesos manuales masivos y en simplificar los procesos de auto-archivo. Esta capa sería la responsable de:

- Introducir usuarios no humanos en el ecosistema (orientados a la interoperabilidad).
- Realizar minería de datos (Han, Kamber, & Pei, 2012) para descubrir producción no almacenada en el ecosistema, asociarla a sus autores de forma inequívoca,

pre-llenar los esquemas de metadatos con datos fiables, actualizar los indicadores de los usuarios, para que estos pudieran consultar *dashboards* (Vázquez-Ingelmo, Cruz-Benito, & García-Peñalvo, 2017; Vázquez-Ingelmo, Cruz-Benito, García-Peñalvo, & Martín-González, 2018) fiables en los que se agreguen todos los indicadores (métricas, alt-métricas (Galligan & Dyas-Correia, 2013)) que les solicitan a los investigadores desde las agencias de acreditación, financiación de proyectos, etc.

- Realizar *machine learning* (Bishop, 2006) para hacer recomendaciones precisas en los procesos de descubrimiento, notificaciones, etc.

## Referencias

- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer Science+Business Media.
- Crouzier, T. (2015). *Science Ecosystem 2.0: how will change occur?* Luxembourg: Publications Office of the European Union.
- Ferreras-Fernández, T. (2016). *Visibilidad e impacto de la literatura gris científica en repositorios institucionales de acceso abierto. Estudio de caso bibliométrico del repositorio Gredos de la Universidad de Salamanca*. (PhD), Universidad de Salamanca, Salamanca, España. Retrieved from <http://gredos.usal.es/jspui/handle/10366/132444>
- Ferreras-Fernández, T., García-Peñalvo, F. J., Merlo-Vega, J. A., & Martín-Rodero, H. (2016). Providing open access to PhD theses: visibility and citation benefits. *Program: Electronic library and information systems*, 50(4), 399-416. doi:10.1108/PROG-04-2016-0039
- Ferreras-Fernández, T., Martín-Rodero, H., García-Peñalvo, F. J., & Merlo-Vega, J. A. (2016). The Systematic Review of Literature in LIS: An approach. In F. J. García-Peñalvo (Ed.), *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'16) (Salamanca, Spain, November 2-4, 2016)* (pp. 291-298). New York, NY, USA: ACM.
- Ferreras-Fernández, T., & Merlo-Vega, J. A. (2015). Repositorios de acceso abierto: un nuevo modelo de comunicación científica. *La Revista de la Sociedad ORL CLCR en el repositorio Gredos. Rev. Soc. Otorrinolaringol. Castilla Leon Cantab. La Rioja*, 6(12), 94 -113.
- Ferreras-Fernández, T., Merlo-Vega, J. A., & García-Peñalvo, F. J. (2013). Impact of Scientific Content in Open Access Institutional Repositories. A case study of the Repository Gredos. In F. J. García-Peñalvo (Ed.), *Proceedings of the First International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'13) (Salamanca, Spain, November 14-15, 2013)* (pp. 357-363). New York, NY, USA: ACM.
- Galligan, F., & Dyas-Correia, S. (2013). Altmetrics: Rethinking the Way We Measure. *Serials Review*, 39(1), 56–61. doi:10.1016/j.serrev.2013.01.003
- García-Holgado, A., & García-Peñalvo, F. J. (2013). The evolution of the technological ecosystems: An architectural proposal to enhancing learning processes. In F. J. García-Peñalvo (Ed.), *Proceedings of the First International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'13)*

- (Salamanca, Spain, November 14-15, 2013) (pp. 565-571). New York, NY, USA: ACM.
- García-Holgado, A., & García-Peñalvo, F. J. (2016). Architectural pattern to improve the definition and implementation of eLearning ecosystems. *Science of Computer Programming*, 129, 20-34. doi:10.1016/j.scico.2016.03.010
- García-Holgado, A., & García-Peñalvo, F. J. (2017a). Definición de ecosistemas de aprendizaje independientes de plataforma. In M. L. Sein-Echaluce Lacleta, Á. Fidalgo-Blanco, & F. J. García-Peñalvo (Eds.), *La innovación docente como misión del profesorado. Actas del IV Congreso Internacional sobre Aprendizaje, Innovación y Competitividad. CINAIC 2017 (4-6 de Octubre de 2017, Zaragoza, España)* (pp. 668-673). Zaragoza, España: Servicio de Publicaciones Universidad de Zaragoza.
- García-Holgado, A., & García-Peñalvo, F. J. (2017b). *Gestión del conocimiento abierto mediante ecosistemas tecnológicos basados en soluciones Open Source*. Paper presented at the Ecosistemas del Conocimiento Abierto (ECA 2017), Salamanca, España.
- García-Holgado, A., & García-Peñalvo, F. J. (2017c). A metamodel proposal for developing learning ecosystems. In P. Zaphiris & A. Ioannou (Eds.), *Learning and Collaboration Technologies. Novel Learning Ecosystems. 4th International Conference, LCT 2017. Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9–14, 2017. Proceedings, Part I* (pp. 100-109). Switzerland: Springer International Publishing.
- García-Holgado, A., & García-Peñalvo, F. J. (2017d). Preliminary validation of the metamodel for developing learning ecosystems. In J. M. Doderó, M. S. Ibarra Sáiz, & I. Ruiz Rube (Eds.), *Fifth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'17) (Cádiz, Spain, October 18-20, 2017)* (Article 91). New York, NY, USA: ACM.
- García-Holgado, A., García-Peñalvo, F. J., & Rodríguez-Conde, M. J. (2015). Definition of a Technological Ecosystem for Scientific Knowledge Management in a PhD Programme. In G. R. Alves & M. C. Felgueiras (Eds.), *Proceedings of the Third International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'15) (Porto, Portugal, October 7-9, 2015)* (pp. 695-700). New York, NY, USA: ACM.
- García-Peñalvo, F. J. (2017a). Mitos y Realidades del Acceso Abierto. *Education in the Knowledge Society*, 18(1), 7-20. doi:10.14201/eks2017181720
- García-Peñalvo, F. J. (2017b). Publishing in Open Access. *Journal of Information Technology Research*, 10(3), vi-viii.
- García-Peñalvo, F. J., García de Figuerola, C., & Merlo-Vega, J. A. (2010). Open knowledge: Challenges and facts. *Online Information Review*, 34(4), 520-539. doi:10.1108/14684521011072963
- García-Peñalvo, F. J., Merlo-Vega, J. A., Ferreras-Fernández, T., Casaus-Peña, A., Albás-Aso, L., & Atienza-Díaz, M. L. (2010). Qualified Dublin Core Metadata Best Practices for GREDOS. *Journal of Library Metadata*, 10(1), 13-36. doi:10.1080/19386380903546976
- García-Peñalvo, F. J., & Tena-Espinoza-de-los-Monteros, M. A. (2017). *Investigación y Ciencia Abierta*. Salamanca, Spain: GRIAL Research Group. Retrieved from <https://goo.gl/4EqCCC>. doi:10.5281/zenodo.1035416

- González-Pérez, L. I., Ramírez-Montoya, M. S., & García-Peñalvo, F. J. (2016a). Discovery Tools for Open Access Repositories: A Literature Mapping. In F. J. García-Peñalvo (Ed.), *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'16) (Salamanca, Spain, November 2-4, 2016)* (pp. 299-305). New York, NY, USA: ACM.
- González-Pérez, L. I., Ramírez-Montoya, M. S., & García-Peñalvo, F. J. (2016b). Open access to educational resources in energy and sustainability: Usability evaluation prototype for repositories. In F. J. García-Peñalvo (Ed.), *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'16) (Salamanca, Spain, November 2-4, 2016)* (pp. 1103-1108). New York, NY, USA: ACM.
- González-Pérez, L. I., Ramírez-Montoya, M. S., & García-Peñalvo, F. J. (2017a). *Identidad digital 2.0: Posibilidades de la gestión y visibilidad científica a través de repositorios institucionales de acceso abierto*. Paper presented at the Ecosistemas del Conocimiento Abierto (ECA 2017), Salamanca, España.
- González-Pérez, L. I., Ramírez-Montoya, M. S., & García-Peñalvo, F. J. (2017b). Usability evaluation focused on user experience of repositories related to energy sustainability: A Literature Mapping. In J. M. Doderó, M. S. Ibarra Sáiz, & I. Ruiz Rube (Eds.), *Fifth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'17) (Cádiz, Spain, October 18-20, 2017)* (Article 35). New York, NY, USA: ACM.
- González-Pérez, L. I., Ramírez-Montoya, M. S., & García-Peñalvo, F. J. (2018). User experience in Institutional Repositories: A Systematic Literature Review. *International Journal of Human Capital and Information Technology Professionals (IJHCITP)*, 9(1).
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques* (3rd ed.). Waltham, MA, USA: Morgan Kaufmann.
- Rodrigues, E., Bollini, A., Cabezas, A., Castelli, D., Carr, L., Chan, L., . . . Yamaji, K. (2017). *Next generation repositories*. Göttingen, Germany: Confederation of Open Access Repositories (COAR). Retrieved from <https://goo.gl/zWeaaG>
- Shneiderman, B. (2008). Science 2.0. *Science*, 319(5868), 1349-1350. doi:10.1126/science.1153539
- Tena-Espinoza-de-los-Monteros, M. A., García-Holgado, A., Merlo-Vega, J. A., & García-Peñalvo, F. J. (2017). Diseño de un plan de visibilidad científica e identidad digital para los investigadores de la Universidad de Guadalajara (México). *Ibersid: Revista de sistemas de información y documentación*, 11(1), 83-92.
- Vázquez-Ingelmo, A., Cruz-Benito, J., & García-Peñalvo, F. J. (2017). Improving the OEEU's data-driven technological ecosystem's interoperability with GraphQL. In J. M. Doderó, M. S. Ibarra Sáiz, & I. Ruiz Rube (Eds.), *Fifth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'17) (Cádiz, Spain, October 18-20, 2017)* (pp. Article 89). New York, NY, USA: ACM.
- Vázquez-Ingelmo, A., Cruz-Benito, J., García-Peñalvo, F. J., & Martín-González, M. (2018). Scaffolding the OEEU's Data-Driven Ecosystem to Analyze the Employability of Spanish Graduates. In F. J. García-Peñalvo (Ed.), *Global Implications of Emerging Technology Trends*. Hershey PA, USA: IGI Global.

