

# Les archives du Web : gouvernance et identités

Francesca Musiani y Valérie Schafer

*La Gazette des archives*, ISSN 0016-5522, n. 245, 2017, p. 203-218

Los archivos digitales ya no son solo un soporte de conservación patrimonial, sino un patrimonio a conservar. Si el archivo de la web entra en el marco del depósito legal, aparecen cuestiones específicas relativas a la selección, recolección y explotación. La conservación es tarea de muchas instituciones, personas, y técnicas. Además de a los archivos y bibliotecas, también implica a al sector de la ingeniería, la investigación o las empresas. En Francia el Institut national de l'Audiovisuel (Ina) se encarga de la conservación de contenidos audiovisuales, mientras que l BnF se encarga del resto, incluyendo webs regionales o contenidos de autores franceses que utilizan dominios .com .org, etc. 4,5 millones de sitios son recolectados por la BnF anualmente, y realiza una recolección más selectiva sobre 20.0000 sitios, mientras que el Ina se centra en otros 14.000, que son recolectados en periodos más breves. El Ina utiliza un robot distinto al Heritrix, usando una tecnología específica que permite la captura de numerosos contenidos en audio y vídeo. El interfaz de consulta puede influir mucho en los resultados ofrecidos al usuario, como demuestran estudios sobre Wayback Machine de Internet Archive. La nueva interfaz de WM (2017) promete ser diferente, con una búsqueda *plein text* y un modo de presentación que prima los metadatos. El cambio en la interfaz de consulta del Ina también busca mejorar la experiencia, proponiendo una búsqueda uniforme. Hay que recordar que el archivo de urgencia no es algo nuevo. Las recolecciones de urgencia (de sucesos previstos, como unas elecciones, o imprevistos, como un atentado) han puesto de manifiesto la

diferencia de los métodos utilizados por la BnF y el Ina, como queda claro en sus diferentes aproximaciones a la recolección de Twitter. Al proporcionar a la vez de los datos y las herramientas para explotarlos, las instituciones de archivo asumen un papel central y un esfuerzo para comprender las aportaciones del *corpus* y de las herramientas. La gobernanza de los archivos web se explicita a la vez en las infraestructuras y en los artefactos técnicos (interfaces, bases de datos) y en las identidades de las instituciones, las misiones de las que se ocupan y las tradiciones de las que provienen. Los momentos de tensión que causan recolecciones de urgencia son una prueba para las instituciones que permite esclarecer sus técnicas.

Resumen elaborado por Antonio Rodríguez Vela