

Metadata Provenance and Vulnerability

Timothy Robert Hart, Denise de Vries

Information Technology and Libraries, ISSN 2163-5226, Vol. 36, n. 4, 2017, p. 24-33

Los objetos nacidos digitales corren el riesgo de degradación, corrupción, pérdida de datos e inaccesibilidad. Para combatirlo se usa la preservación digital. Las dos estrategias principales son migración y emulación. La migración es más sencilla, pero la emulación es más segura. Los metadatos son clave en ambas estrategias. Por eso hay que asegurarse de que los metadatos sean procesados correctamente para todo el ciclo vital de los objetos digitales. Los metadatos más usados son administrativos, descriptivos, estructurales, técnicos, transformativos, y de preservación. Verificar y validar la autenticidad de los objetos digitales es importante para su preservación, conservación y archivo. Algunos tests, como el proporcionado por ExifTool son útiles para analizar su validez. El primer caso de estudio se refiere a un archivo JPEG. Los metadatos incluían fechas, metadatos técnicos, de creación, etc. Una vez subido a una red social, la mayoría de estos metadatos se pierden. El segundo caso cambia una imagen TIFF a una JPEG usando Photoshop. También hay una importante pérdida de datos. El tercer caso se centra en un PDF con metadatos sobre versión, autor, creador, etc. Algunos softwares permiten extraer metadatos, lo que permite a los usuarios importar PDF y automáticamente genera los metadatos bibliográficos apropiados. Pero es un método inconsistente que produce fallos en la recuperación de metadatos. El cuarto caso examina los metadatos de un archivo DOCX. Los archivos de Microsoft Office contienen un archivo XML externo con información sobre formateo de datos, información del usuario, historial de edición, etc. Cada elemento del documento tiene

sus propios metadatos. El siguiente paso en el test consiste en convertir el documento en PDF, borrándose casi todos los metadatos. Esto demuestra la pérdida de información que supone un cambio de formato. Para preservar los metadatos es necesaria la intervención humana, una tarea laboriosa y que implica errores. Es conveniente usar metadatos genéricos y un entrenamiento específico. Todos los tipos de metadatos deben ser capturados para permitir a los usuarios de cualquier disciplina que se beneficien de unos datos históricos durante muchos años. Dados los continuos cambios tecnológicos, hay que estar preparados, observar las vulnerabilidades de los metadatos y actuar para salvaguardar la historia digital.

Resumen elaborado por Antonio Rodríguez Vela