

Manifestations of Metadata Structures in Research Datasets and Their Ontic Implications

Erik Radio, y otros

Journal of Library Metadata, ISSN: 1938-6389, Vol. 58, n. 3-4, 2017, p. 161-182

La proliferación de conjuntos de datos de investigación y su disponibilidad en varios repositorios requiere metadatos que proporcionen contexto suficiente y claridad organizativa como para permitir su reutilización. Las “estructuras de metadatos” hacen referencia a los tipos de datos subyacentes que definen la ejecución técnica de un estándar de metadatos y la “estructura interna” a cómo los datos están organizados. La elección de un modelo de metadatos tiene un efecto significativo en la consecuente interpretación y uso de los registros como representación. Las estructuras de datos informan a los modelos de datos más allá de la simple descripción. La estructura plana es uno de los tipos de datos más básicos, consistente normalmente en un conjunto desordenado de relaciones clave: valor. Su ejemplo más claro es Dublin Core. La estructura en árbol consiste en una colección de nodos y ramas que surgen de una única raíz. Es común en metadatos bibliográficos y estructurales porque permite relaciones entre valores más explícitas. Su forma es reconocible en el modelo XML, como el usado por MODS. La estructura en gráficos consiste en vértices y bordes y pueden ser directos o indirectos. Se suele asociar con el lenguaje RDF. Estos estándares describen diferentes tipos de conjuntos de datos. Los utilizados por las Ciencias de la vida se caracterizan por varios ejes de variación, reflejados en su mayor parte por metadatos desestructurados y falta de estándares. En Ciencias sociales hay una larga tradición de reutilización, en parte gracias al estándar de la Data Documentation Initiative, de tipo árbol, que sirve para describir datos de encuestas y estadísticas. Para Humanidades se crean datos muy heterogéneos, que en muchos casos ni tan siquiera se puede definir. Los Sistemas de Información Geográfica han permitido a muchas disciplinas utilizar datos geoespaciales, que no se basan en materias, sino en contenido. Es común a todas las disciplinas la necesidad de describir no solo los datos generados, recolectados y usados, sino también el software utilizado, ya que esto permite la reproducción y la reutilización. Hay varias dificultades a la hora de describir un software: diferentes escenarios, complejidad o dificultan en las interacciones.

Resumen elaborado por Antonio Rodríguez Vela