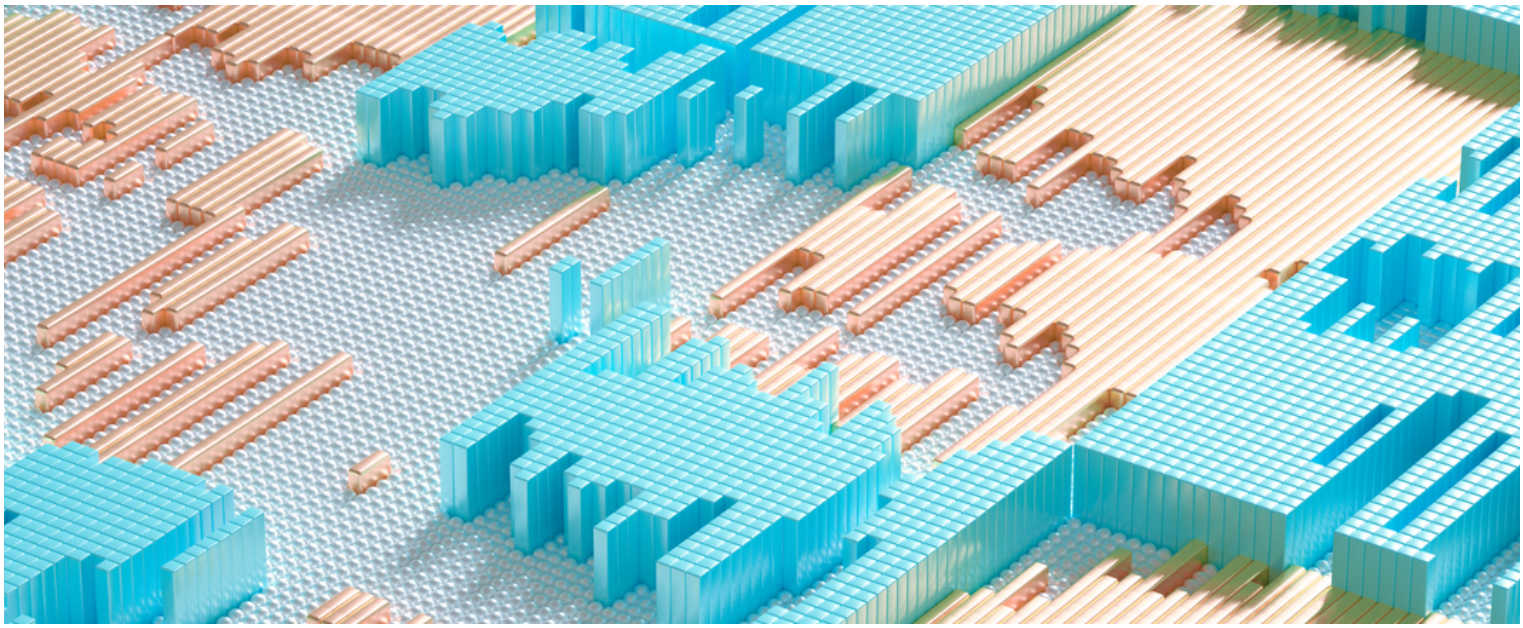
Mar 2020

International: Metadata and standards

Metadata, sometimes called 'data about data,' has been more formally defined as 'structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource[1].' Dr. Victoria Uren, Lecturer at Aston University, provides an overview of what metadata is and how it can be used.



Pierell / Essentials collection / istockphoto.com

We are all familiar with metadata in our daily lives. A library catalogue which locates books is a classic example.

The 'structured' part of the definition is important. The thing that turns data into metadata, rather than just more data, is its structure. Communities of users share understanding of metadata through standards, including vocabularies and formalised syntax for marking up metadata, to allow it to be shared between systems. To illustrate, take the number 0141364726. To a UK reader, this number may appear at first glance to be a phone number, because 0141 is the dialing code for Glasgow. It gets closer to becoming metadata if we label it with the tag International Standard Book Number ('ISBN'). ISBN 0141364726 identifies the book, 'Diary of a Wimpy Kid: Old School' by Jeff Kinney. The label has this meaning because ISBN numbers are defined in an international standard.

Labels alone do not make a metadata system. Our ISBN number was useful because it let us get to the title and the author of the book. For this we need a vocabulary to let us describe the different things in our metadata system. In the context of books, this would be provided by something like the MAchine-Readable Cataloging ('MARC') 21[2] format for bibliographic data maintained by the Library of Congress. This defines the kinds of data to be expected in a bibliographic record and gives each an agreed label.

020 is the MARC label for an ISBN number, and 210 is the abbreviated title, the one you are likely to see in the catalogue, and so on. This system of standardised labels provides users with a vocabulary that links bits of data to a shared understanding of things in the world.

Libraries and booksellers around the world buy bibliometric metadata in formats such as MARC 21 records in order to populate their catalogues, saving many hours of tedious, error prone work creating records by hand. This readily available source of structured metadata is one reason why some online retailers started life as bookshops, as publishers' metadata could kick start the product catalogue.

Metadata based knowledge exchange formats have had a significant role in business information systems. A historical example, electronic data interchange ('EDI'), goes back to the 1960s and for decades facilitated the transfer of business documents via Telex. The development of the Extensible Markup Language (XML) by the World Wide Web Consortium in 1998 provided a standard syntax for metadata which was adopted by a wide range of user communities to structure data for sharing. For example, the Universal Business Language (UBL)[3] provides templates for standard electronic business documents, like invoices and purchase orders. Transforming these to metadata allows them to be directly exchanged between the information systems of partners in supply chains.

As we have seen, metadata acquires meaning based on the shared understanding, embedded in standards. This idea was picked up and extended in the 2000s by Tim Berners Lee and a community of artificial intelligence researchers, who called their idea 'the Semantic Web.' The underlying rationale of the Semantic Web is that the labels on items of metadata provide semantics, formal meaning, to items of data, and that this can support algorithms to reason using the metadata. Semantic web metadata is joined together into networks of linked data items.

It is expected that metadata will also become a necessary component of the Internet of Things ('IoT'), as devices will need to share information in mutually understandable ways. For example, fitness trackers currently transmit metadata to the cloud for review on smart phones. In the IoT, it could exchange metadata with equipment in the gym giving a more detailed picture of your training, or even setting up the equipment automatically for a tailored workout.

Adherence to metadata standards is not necessarily a legal requirement, but it is always good data management practise. Standards ensure metadata is compatible with different information systems, which facilitates cooperation. However, where metadata relates to personal information, as in the example of fitness data, it is covered by General Data Protection Regulation (Regulation (EU) 2016/679) ('GDPR') and other similar legislation.

Dr. Victoria Uren Lecturer
v.uren@aston.ac.uk
Aston University, Birmingham

---

1. See: https://www.niso.org/publications/understanding-metadata-2017
2. See: https://www.loc.gov/marc/bibliographic/
3. See: http://ubl.xml.org/

RELATED CONTENT

NEWS POST

USA: ABA sends letter to OMB calling for AI guidance

NEWS POST

USA: ITI issues comments on draft memorandum on guidance for AI application regulation

NEWS POST

India: Lower House publishes question on online video streaming regulation

NEWS POST

Pakistan: PTA invites AIC for consultation on citizen protection rules

NEWS POST

Australia: OAIC releases submission on IoT voluntary code of practice