Since starting my work here at OpenUCT in mid-November 2013, I’ve been using the word ‘metadata’ quite a lot (sometimes also spelled ‘meatdata’ at first attempt). Say this word to a given member of the general public and you will usually elicit the following response: a slight glazing or crossing of the eyes, accompanied by dull incomprehension, and a tentative “What’s that?” This is not to say that these people don’t know what metadata is or don’t use it themselves on a daily basis; there simply comes a point where conceptions and definitions diverge from everyday experience, and seem to enter the magical realm of fairy dust and crop circles.

The ubiquitous definition of metadata – “data about data” – is perhaps the cause of much of the trouble. How can something be defined by itself? This is indeed flirting with the mystical. A more sensible approach might be to look at it etymologically: the Greek root “metá” meaning “beyond”, “upon” or “after” (thank you Wikipedia). Data after data; descriptive data. Thus, you could say that metadata is information that takes as a precondition for its existence, the existence of other information: you cannot have metadata without something to describe. I suppose, from a linguistic perspective, then all data is metadata, in as much as all data is an attempt to describe perception, but let’s not complicate matters. Alternately, if one is typing too fast, it could simply be data about hamburgers.

But realistically, what is metadata? The most oft cited example would have to be from my own field of library science: catalogue entries. These are highly structured records (data sources) that describe the physical – and sometimes intellectual – content of items (other data sources), usually books. Another example, also from library science, would be the Dewey Decimal System, which ascribes a data point to each item, which itself contains a whole lot of data about that item. Then of course, each book itself contains metadata of its own, in the form of a verso (the back side of the title page, where publishers put all of the fine print about copyright and other things which people who actually read the book don’t really care about), an index, and even the blurb on the back cover, a kind of abstract. So, we can see that metadata itself is hardly a novel concept, but you might still be wondering how it applies to you. (Assuming you’ve never been to a library, and used the catalogue and Dewey to look for a book, which is the worst possible thing to assume about anyone. I’m sorry. I didn’t mean it.)

The following explanation might clarify things (again, Wikipedia): by describing the contents and context of data, the quality of the original data is greatly increased. Think about it: if a library containing millions of books (UCT’s library, for example, houses over 1.2 million volumes, and provides electronic access to over 72 000 journal titles) did not have a catalogue, and some kind of classification scheme like Dewey to organise them, it would be more-or-less useless. Nobody (except, of course, for the librarian) would have any idea of how or where to find anything. The value of the library’s collection resides in the content of the books that it contains, but if you have no way of accessing that content, it may as well not even be there. By using metadata to describe data, to situate it either in a context or a physical (or cyber) location, you can help people (or, perhaps more crucially, Google) to make decisions about whether or how the data may be useful to them. Metadata is about making your data maximally visible and usable.

Let me give you a more personal example. I’m a musician – at least by training, if not so much by disposition – and quite deeply passionate about collecting and listening to music. Over the years, I’ve amassed quite an extensive collection of recordings (I won’t even touch on all of the copyright issues), which I organise, much to the dismay of a number of my friends, using iTunes. Now, while I won’t comment on the mentality behind the Gracenote approach to metadata and the album information available for download, I find iTunes a great tool for creating and managing metadata (think “album title”, “artist”, “composer”, etc.), and for providing access to all of my music. I’ve developed a pretty sophisticated metadata protocol within my iTunes environment, and frankly, I’m pretty darn proud of it. All of this was done years before I followed that incomprehensible fascination for library science, and so, when it came time to talk and learn about these issues in my studies, I felt like I’d been doing it for years. I immediately understood the value and importance of good quality metadata, and a rigorous, consistent protocol.
Of course, with the so-called “information explosion” and the exponential increase in the world’s data volumes, the issue of metadata has come into stark relief. It is no longer simply about making data visible and usable, but also shareable and re-usable, and this is where standards and protocols come into play. If you want someone on the other side of the world to be able to access and use your data, obviously it needs to be described to some minimum standard of comprehensiveness and intelligibility. In an age where no one can afford to waste effort, and sharing has become the new norm, it makes sense to capitalize on the efforts of others.

And if you’re still not convinced that metadata applies to you, think about the last time you went to see a movie, and you looked-up a review on IMDB or Rotten Tomatoes: that’s metadata. What about the last time you went out for a meal, choosing something from the menu because the description sounded tasty. Technically, that’s metadata. Or if you were at Spur, maybe it was meatdata.

PS. Just in case you’re interested, here’s a bit of metadata for you: you’ve just read “Metadata: not a four-letter word”, a 985 word blog post about meta/meatdata, written by Kyle Rother, formatted in HTML, available on the OpenUCT blog since 24 February 2014, and licensed under CC-BY-SA. Sorry, I’ve been thinking about metadata quite a lot.

(In the library image by YLev available from Flickr under a CC-BY license; Folders vs metadata image by john-norris available from Flickr under a CC-BY-SA license)

by Kyle Rother
Blog post

Add comment

Your name

Subject

Comment *

Text format Filtered HTML ▼

Web page addresses and e-mail addresses turn into links automatically.
Allowed HTML tags: <a> <em> <strong> <cite> <blockquote> <code> <ul> <ol> <li> <dl> <dt> <dd>

Lines and paragraphs break automatically.