The Need for Metadata

Ed Simons

Radboud University Nijmegen
e.simons@bz.ru.nl

Metadata allow us to describe and classify research information in a systematic way, and as such they are indispensable for searching and finding academic publications and other results of research.

In order to make full use of the information discovery potential of the Internet, the *formal* and *content* metadata commonly used in repositories should be supplemented with the *context* metadata as stored in a CRIS. Whereas the "traditional"repository metadata stem from the "Gutenberg Era" - the past few centuries in which print on paper was the main research information and communication tool - and focus on (the description of) the individual publication, the context metadata used in a CRIS are more appropriate for our age of electronic and networked information supply.

The reason for this is that *context* metadata present the user with and opens up an entrance to the information network of which the individual publication is (only) a part. By adding context metadata to the publications in a repository, the user at harvesting time not only sees a list of single publications but is shown an additional set of "hub-like", networked information objects (e.g. colleague researchers of the author(s) of the publication, the research project the publication results from etc...) which can lead to the discovery of other, related publications. This not only and obviously brings substantial added value for the user but at the same time fully corresponds to our culture of gathering information in the "Berners-Lee Era" we currently live in, notably by "hubbing" from link to link on the Internet.

So adding context metadata to the metadata set of the repositories should have a high priority. Now that the repository and CRIS communities - up to now isolated from each other - recently have started to discover and talk with each other, and gradually are moving towards cooperation and integration, the time seems right to start up the process of defining an optimal and standardized list of context metadata to be added to the "traditional" repository metadata. This in my view, as the above implies, should be an activity carried out jointly by the two communities.

Not only the addition of context metadata reveal the information network "around" a publication, there is another kind of information network, even more directly related to the content of the publications, which is based on the mutual citations within the publications included in the network. To put it briefly, these networks can be characterized as *citation networks,* and at the moment the (way to extract and register) metadata from publications are a hot topic in the repository world.

CRIS (will) play a cardinal role in both the storage and supply of context metadata as well as metadata on citations and as such in the discovery and presentation of both c*ontext based publication networks* and *citation based networks*. On the one hand the storage of context metadata, as implied in the above, is a kind of trade mark or key aspect of a CRIS and on the other hand experiments within the framework of CRIS are going on to automatically, through

web services, gather the needed citation metadata from the databases of the big and settled players in this field like ISI and scopus.

All this could, in the near future, lead to a CRIS-driven repository model, where all the metadata used in the repository is stored in a CRIS and automatically supplied to the repository.

For this to work is not only and perhaps not in the first place a matter of technology, but also and perhaps even more so, a matter of an optimal integration of the repository and the CRIS communities within a university.