

Open linking

A discussion paper for the ELAG 2004 workshop.

Peter Noerr

What is linking?

Linking in the bibliographic sense may refer to a number of things:

- linking from a bibliographic citation to the complete record,
- Linking from the bibliographic citation to the full text of the item,
- Linking from the full text of an article to “associated” items,
- Linking from a citation to a particular copy of the item,
- Linking from a citation to other FRBR elements (e.g. manifestation to work),
- Linking from a bib citation to holdings information
- Linking from a citation to an identifier (and vice versa)
-

All of these links relate a starting piece of bibliographic information to other pieces of information, related in one of a number of different ways (part-whole, alternative form, delivery location, etc.) and the links may be designated and processed in many different ways.

Open linking generally refers to linking which takes place using the OpenURL as the linking mechanism and a resolver of some sort to process the information and create the (or more than one) link for the user.

However open linking may also refer to linking done using open source software and systems, or to broadly based linking offering an “open” set of options to the user.

What is it useful for?

Linking provides a mechanism to lead the user from where they are now to either:

- A more complete form of the same item,
- The object to which the item where they now are refers,
- A place where the object can be obtained,
- An interesting, somehow related, item,
- A process which uses the item as its starting point to discover other items.

Thus links can be used to narrow or broaden the users focus, depending on how they are created and how they are processed. Some links are single minded and have only one purpose (e.g. to take the user to where a copy of an article may be obtained), others offer the user a range of options, some of which are links and some of which are not (e.g. the menu of services from an OpenURL resolver), but all of which are deemed by the administrator of the system to be of potential use to the user.

An important classification of links is whether they are manually or automatically created. Most are manually created even though processed by machine. A person decides that a full text link should be made available or a 'place hold/reservation' link should be on monograph entries in a result list, or a DOI should be added to a record.

Another important point about links is that they usually appear as a result of searches conducted by some other means. They are rarely (except in the case of the Web or a hyperlinked document) the sole method of discovery. They co-exist with other methods. Even their resolution may involve a variety of processes.

They are both a method of discovery and a method of delivery. They are definitely a product of a 'relationship' view of the world for they are the embodiment of those relationships.

How is it achieved?

At its crudest linking is achieved by embedding a pointer within one item which points to another. And this was the original form of links. It is still the form of linking used within the Web. Links there are generally one-to-one, single purpose, static, and absolute. They are basically fairly crude compared to bibliographic links. Of course there are quite a few of them!

An interesting question is to compare Web linking structures with those used within the bibliographic field and attempt to determine if the Web linking structure can represent the same structures as bibliographic linking, and, conversely, if bibliographic linking could be scaled up to handle Web volumes.

The biggest problems with the simple static linking just described is that it is static and that it is context free. It can be argued that this is a big advantage in some circumstances, but for most real world situations an amount of context is needed. This lead to attempts to add context to the processing of the link, usually the user's context as that is the biggest variable and the one of most interest.

In the information world this lead to the OpenURL resolver. Note that an OpenURL is only a format for specifying various pieces of bibliographic and other information. It does not "DO" anything. The "doing" is the property of the resolver system. The original resolver (and still over 90% of them) determined where the user belonged (to what

organisation) and used that to determine which copy of a journal article to deliver to him/her. This is the so called ‘appropriate copy’ problem (or the ‘Harvard’ problem). It is a serious real world problem, in that not having such a system costs a lot of money in wasted subscriptions to journals when the already subscribed to copy is not used by the user, but another one is purchased.

This Open(URL) linking is achieved through a resolver which provide one or more services to the user. A generator is also necessary to create the OpenURL in the first place and this may be pre-done or it may be real time, but this is essentially a mechanical process. (It is worth noting that this generation process is actually one of the biggest challenges facing the adoption of OpenURL technologies in the library market. The problem occurs in all circumstances, but is worst when metasearch engines are involved. Discussion on this problem and what can be done about it are worthy of time within this workshop – please bring ideas of what the problem might be to the workshop.)

The interesting part of the OpenURL process takes place at the resolver, and it is there where discussion of possible, probable, and interesting linking functions can be very productive. With the imminent introduction of V1.0 of the OpenURL standard (NISO Z39.88) all the organisations that would play in this market have an opportunity to produce new unique offerings and capture market share, and drive the direction of linking for the next few years. Most of these players do not have many ideas about where linking (more correctly link resolution) can lead. This workshop could educate some of them as to what practitioners are conceiving as the possibilities.

What is its future?

In bibliographic and information terms the immediate future lies with the direction and shape of OpenURL resolver services. These are in their infancy. Now is the time to influence them.

Other linking methods and systems abound. The use of DOIs can provide very useful linking. The Crossref service is a linking mechanism which indirectly provides access to article full text. Reference databases within library portals or web sites (www.lij.org for example) provide a manually created set of links. The creation and use of these is in its infancy. They are labour intensive to create and maintain, Automation may help them grow more important. OAI harvesting projects create vast stores of information in a very unstructured form (as do ‘open publishing’ projects such as the various Dspace projects where original research papers are available usually free of charge). These are ripe for some techniques to internally organize them as well as link them collection to collection – is keyword indexing really the answer?

In a broader context the whole semantic web area is one of linking, whether by explicit embedded links or by catalogue references or co-occurrences. This is an area where linking is obviously very important and things are just getting started. Here the idea of context has a somewhat different slant to that of the OpenURL. Semantics play a much

bigger part (as they should with the name), and the use of ontologies for various search and linking functions are under investigation. What those uses should (or could) be is open for anyone to suggest. If the suggestion is reasonable then possibly the EU will fund a pilot!

Linking is not in its infancy, but it is in the middle of a period of serious growth and is poised to move in some new directions. A few bright ideas now could influence those directions.

Homework

Instead of feed you references (which you won't read). I suggest you go to one of the bibliographic metasearchers and try searching for "open linking" and other suitable keywords. (If you do do this I would be very interested in the keywords you use and the results you get.)

As a quick and easy web search; go to my company's demo of a public metasearch engine at www.museseek.com. This will let you search all the major search engines at once. There is little processing of the results and none of our linking features are implemented on this site, but it will let you find a lot of articles and web sites. You could try one of our competitor's systems as well and see how they compare. (You have to find out who they are. Hint – look at the delegate list for this conference.)