Pergamon

Library Collections, Acquisitions,
& Technical Services 26 (2002) 219–230

Library Collections, Acquisitions, & Technical Services

Metadata and reference linking

Miriam E. Blake*, Frances L. Knudson

Library Without Walls, Los Alamos National Laboratory, P. O. Box 1663, MS-P362, Los Alamos, NM 87545, USA

Abstract

Reference linking is a broad term that generally means linking from one information object to another. The specific types of linking which have been addressed in most detail in recent years are those having to do with the links between journal articles. These would include the links from citation metadata to the electronic full-text article and links from references following an article (the bibliography) directly to the referred citation and/or article. A basic concept is that there must be a way to identify the work to be 'linked-to.' A second concept is that in order to 'link-to' an outside system, there must be an identifiable syntax, which often includes an identifier, for creating a query into that system to find the correct article. In this paper we focus on experiences in linking from an A&I database record to full-text and linking from a bibliography to full-text. Accomplishing this required implementing a system that uses metadata to determine the identifiers and the required elements for various 'link-to' syntaxes across disparate systems.

Keywords: Metadata; Reference linking; SFX

1. Introduction

Since the late 1990s, significant attention has been given to defining and creating the architectural components of reference linking systems. By 1999, there had been some elucidation of the complexities involved [1, 2] and some steps made toward building a generalized infrastructure to replace the individualized constructs used in the earliest systems. Basic concepts, such as identifiers, resolvers, reference databases, and localization services [3] are all pieces of the emerging architectures. At the Los Alamos National Laboratory (LANL) Research Library, we have been building systems with elements of

* Corresponding author.
E-mail address: meblake@lanl.gov (M.E. Blake).

1464-9055/02/$ – see front matter
PII: S1464-9055(02)00253-1

reference linking since the mid 1990s. Our early systems, which linked from locally loaded A&I databases (such as INSPEC® and BIOSIS®) to full-text journal articles stored both locally and at remote websites, employed what has since been termed a "static" linking system [4]. We have since begun moving all of our linking efforts into SFX (marketed by Ex Libris Information Services Division) [5]. SFX is a dynamic, open-linking system [6] that includes an OpenURL resolver and a locally controlled linking system rules database. Through these efforts we have encountered a myriad of metadata-related issues that are intimately entwined with the emerging universe of reference linking.

2. Identifiers and 'link-to' URLs in reference linking

Identifiers are data strings that uniquely identify a specific information object, such as a specific journal article. These identifiers can either be 'dumb' –that is, strings with no immediately intelligible meaning – or they can be based on an algorithm that uses the metadata present in the citation. The Digital Object Identifier (DOI) [7] is essentially a 'dumb' identifier that has a resolver, or resolution system, behind it (the Corporation for National Research Initiatives, or CNRI Handle System) to connect it with the actual URL of the content. The DOI is now the primary identifier being assigned in the publishing community. The problem currently faced, however, is that DOIs (and similarly deployed identifiers, such as PubMed IDs) do not exist right now in the bulk of abstract and indexing (A&I) databases, electronic full-text journal metadata, etc.

Many information providers already have systems in place that can accommodate some level of linking (both in and out) regardless of the ubiquity of the DOI. Generally, most systems determine identifiers by using algorithms based on the bibliographic metadata found in the citation. The Serial Item Contribution Identifier (SICI) standard provides rules to calculate unique identifiers from journal citation metadata and is used in practice at the LANL Research Library in locally controlled A&I databases as well as in implementations of large-scale linking systems [8]. In bodies of information with large amounts of well-tagged metadata, it is fairly easy to automatically calculate metadata-based unique identifiers such as a SICI, but even then, inconsistencies in metadata exist between different sources for a variety of reasons (Lagoze offers some explanations for the diversity in metadata as part of the background to his paper [9]). Many ways of calculating and formatting metadata-based identifiers exist in addition to the SICI –no single standard has led the way. The 'link-to' syntax used by each system is equally disparate. The construct of the URL needed to query an information system to retrieve the item that should be 'linked-to' generally includes either an identifier or some amount of metadata, and sometimes both. These query strings are again non-standard (although it should be noted that the OpenURL specification provides an interface to transport identifiers and metadata in a standardized way and is currently being reviewed by a NISO Standards Committee [10]). Tables 1 and 2 list the variety of metadata currently required by different scientific, technical and medical (STM) publishers to 'link-to' their full-text journals.

We have been using SFX in production at LANL since November 2000. In brief, SFX is a suite of software and a "rules" database that helps glue together the many pieces needed

Table 1
Metadata required for full-text linking—Academic-Dekker

Metadata required to link	Academic	Allen	ACS	ACM	AIP	AMS	APS	ADS	Cambridge	Catchword	CSIRO	Dekker
ISSN	y	y								y	y	y
E-ISSN		y								y		
Volume	y	y	y	y	y	y	y	y	y	y	y	y
Issue	y	y	y	y	y	y			y	y	y	
Starting page	y	y	y	y	y		y			y		
Ending page										y		
Date-year				y		y		y		y		
Date-month												
First author last name				y								
Article title												
Journal title												
DOI												y
Short journal name	y		y	y	y	y	y	y	y	y	y	y
Host URL	1	3	3	1	2	1	3	2	1	2	1	
Article level (= abstract level)	y	y	y	y	y	TOC level	y		TOC level		TOC level	Journal home page

for creating working reference links and other types of linking services based on a local library profile. SFX is an early contender in an emerging field of "local resolution servers" [11] that employ the OpenURL framework [12]. The OpenURL is used in conjunction with source-specific translators to create identifiers (if no DOI or other identifier is already present) from available metadata. The associated rules database reads the OpenURL and

Table 2
Metadata required for full-text linking—Elsevier-Wiley

Metadata required to link	Elsevier	Highwire	JSTOR	Project Muse	Oxford	Royal Society of Chemistry	Science Server	SIAM	Springer	Synergy	Wiley
ISSN	y		y		y		y				
E-ISSN											
Volume	y	y	y	y	y	y	y	y	y	y	
Issue	y	y	y	y	y	y	y	y	y		
Starting page	y	y	y		y		y	y	y	y	
Ending page											
Date-year	y	y	y			y			y		
Date-month			y								
First author last name	y		y								
Article title			y								
Journal title	y										
DOI	y										y
Short journal name		y		y	y	y		y	y	y	y
Host URL		5	2	1	2	1		1	1	2	2
Article level (= abstract level)		y	y	y	y	TOC level	y		y	y	Journal home page

creates a query syntax into the target system using rules supplied by the library for its local subscriptions and services. In practice, SFX does the following:

1. Creates a unique OpenURL hook for each citation or bibliography reference. This is displayed as a link or "SFX Button" to authorized users of a given source system
2. Assembles a metadata string (possibly including an identifier) "on-the-fly" based on rules for that particular source, when the user clicks on the link
3. Sends the metadata string via the OpenURL to the SFX server and compares the metadata elements to existing rules and thresholds in the rules database
4. Returns a menu of "appropriate" links [13], including a link to the full-text if the rules have been met, to the user. For example, a full-text link can be presented if the source metadata contains an ISSN that exists in the rules database and the volume and date are within the threshold information for that ISSN.
5. Ultimately enables the user to click on a link in the menu and access the target via the URL that was constructed via the 'link-to' syntax information in the rules database using the descriptive metadata from the source

The SFX workflow shows clearly that metadata is an important element in both identifying the object to be retrieved from the target as well as in constructing the 'link-to' syntax in the target URL. This is likely to continue to be the case for SFX and any other local resolution linking systems that emerge, since metadata, however different in its content and markup, is the one thing available in all text-based systems.

3. Metadata and reference linking in practice

At LANL, our main focus to date has been linking from A&I database records to full-text articles. Using our locally loaded databases as sources in our SFX implementation required an examination of what metadata we would pass on the OpenURL and use for SFX services. Additionally, we set up 57 publishers (4120 journal titles) as targets to be 'linked-to', and became very familiar with the oddities of the different 'link-to' syntaxes required. For this paper, we also examined the variations in how papers are cited and the resulting linking problems. While our expertise lies with the STM world, humanities and social science literature will experience these same metadata problems and may introduce discipline specific issues.

Creating 'link-to' queries requires two types of information: metadata that describes the 'link-to' item and threshold information. Threshold information in a localized system contains library-specific full-text holdings information, including beginning/ending subscription dates and volumes. Tables 1 and 2 indicate the major pieces of metadata that are used in full-text linking algorithms. The majority of publishers require only the ISSN, volume, issue/number, and starting page. But as is evident in the table, some require additional information such as the first author's last name and the title of the article. Another frequent variation occurs in pubishers' requirements for the short journal name, which usually appears in the journal's URL. For example, Academic Press uses two letter abbreviations for short journal names while the American Chemical Society uses CODENs for their short journal

Table 3
Examples of database records and calculated full-text URLs

1. Springer
Photorefractive materials: properties and applications
Buse K, Kratzig E, Ringhofer KH
Applied Physics B-Lasers and Optics, May 2001; v. 72(#6) pp. 633–633
http://link.springer-ny.com/link/service/journals/00340/bibs/1072006/10720633.htm
2. ACM
Lazy rewriting on eager machinery
Fokkink W, Kamperman J, Walters P
ACM Transactions On Programming Languages And Systems, Jan 2000; v. 22(#1) pp. 45–86
http://www.acm.org/pubs/citations/journals/toplas/2000-22-1/p45-fokkink/
3. American Physical Society
Accelerated universe from gravity leaking to extra dimensions.
Deffayet, C.; Dvali G.; Gabadadze, G.
Physical Review D; 15 Feb. 2002; vol. 65, no. 4, p. 044023/1-9
http://publish.aps.org/abstract/PRD/v65/e044023
4. Highwire Press
Information storage and retrieval through quantum phase.
Ahn, J.; Weinacht, T.C.; Bucksbaum, P.H.
Science; 21 Jan. 2000; vol. 287, no. 5452, p. 463–5
http://www.sciencemag.org/cgi/content/abstract/287/5452/463

names and Highwire Press employs several kinds of short journal names. Another area of variation is with the host URL data. Tables 1 and 2 indicate the number of URL structures used by publishers. Although the majority of publishers have one URL structure, others have multiple with Highwire having the highest number of URL types. Table 3 contains bibliographic records and the calculated URLs for several different publishers.

Below we examine most of the metadata pieces, pointing out weaknesses or difficulties that have been encountered. Examples have been extracted from A&I databases locally loaded at LANL. These include BIOSIS®, DOE Energy, Engineering Index®, INSPEC®, and SciSearch® at LANL.

4. Examination of metadata

4.1. ISSNs

Any serials cataloger could easily produce a thesis on problems with ISSNs. Tables 1 and 2 indicate that most algorithmic linking syntaxes rely on ISSNs. A frequent problem in linking from a citation database record to a full-text article is incorrect ISSNs. For example, the journal *Hyperfine Interactions* displays an ISSN of 0304–3834 on its physical cover, which is the ISSN that most OPAC records contain. Ulrich's and Kluwer (the publisher) note an ISSN of 0304–3843, which is the ISSN used by most A&I databases. The transposition of the last two digits of the ISSN probably started as a simple typographical error but is capable of causing confusion and breaking a linking system.

Other causes of ISSN problems are journal title changes, splits, mergers, and combina-

tions. When a journal splits, new titles are created with enumeration beginning with volume 1 but the ISSN will not change. There are still journals without ISSNs or journals that have started publishing with "ISSN pending." In the paper world, catalogers can wait to fix the record. In the electronic world, this is not acceptable.

Some publishers are slow to change to a new ISSN after a title change, and some A&I databases are extremely slow in catching ISSN changes. This delay can break a linking system that relies on ISSNs. *Fusion technology*, 0748–1896, changed its title to *Fusion science and technology* in 2001. The ISSN changed at some date to 1536–1055 but there are A&I database records for *Fusion science and technology* in 2001 using the old ISSN. Another example is the title, *Meteoretics*, 0026–1114. The title changed in 1996 to *Meteoretics and planetary science*. A new ISSN, 1086–9379, was obtained but there are A&I database records for 1996 and 1997 under the old ISSN. The full-text linking algorithm does not work for this two-year span as a result. Isolating these types of ISSN problems is very time consuming.

Electronic ISSNs seem to be gaining in popularity and some linking schemes use a journal's electronic ISSN. Many A&I databases, however, do not seem to handle both a print ISSN and an electronic ISSN, requiring any linking system to check for both.

Threshold data can be manipulated to handle some of these ISSN problems. Starting years and ending years can be altered in attempts to fix the problem. For the title, *Meteoretics*, we could alter the ending year and volume for the ISSN 0026–1114 but determining the effect of this change in all A&I databases is time consuming and prone to errors. This also makes the threshold data in the linking system rules database different from the OPAC holdings.

4.2. Volume & issue

An interesting pattern has been detected involving volume designation and journal titles that have multiple parts. Some A&I databases consider the part designation to belong to the title; others include the part designation as part of the volume. And, of course, there is inconsistency even within single databases. The three examples below are source fields from two different databases. The alpha characters in the volume designation can break a linking syntax or at least require special handling that makes the algorithms more complicated.

INSPEC: Applied Physics A (Materials Science Processing); Dec. 2000; vol.A71, no.6, p.689–93
SciSearch: Applied Physics A: Materials Science and Processing; Dec 2000; v.71, no.6, p.689–693

INSPEC: Indian Journal of Physics, Part A; July 2000; vol.74A, no.4, p.371–4
SciSearch: Indian Journal of Physics, Part A; Jul 2000; vol.74, no.4, p.371–374

Engineering Index: Sensors and Actuators, A: Physical; Aug 2000; v.85, no.1, p.54–59
INSPEC: Sensors and Actuators A (Physical); 25 Aug. 2000; vol.A85, no.1–3, p.54–9

The difficulty here is determining if all alpha characters in the volume can be ignored. For these three titles, ignoring the alpha character will allow the full-text linking to work but it could break it for other journal titles.

Double volumes are also treated with inconsistency. Some A&I databases only include the first volume; others include the range; one includes SISI. The example below contains several inconsistencies.

SciSearch: Materials Science And Engineering A-Structural Materials Properties Microstructure And Processing, Jul 15, 2001v. 309(SISI) pp. 328–330
DOE Energy: Materials Science and Engineering A; Jul 15 2001; v.309–310, p.328–330
INSPEC: Materials Science & Engineering A (Structural Materials: Properties, Microstructure and Processing); 15 July 2001; vol.A309-A310, p.328–30

Double issues are common in the STM world. Again with a trial and error approach, the appropriate issue form can be determined, however, it might differ among different publishers or aggregators.

Engineering Index: Journal of Low Temperature Physics; 2000; v.119, no.3, p.337–342
INSPEC: Journal of Low Temperature Physics; May 2000; vol.119, no.3–4, p.337–42

This citation shows different handling of parts, double volumes, etc.

Engineering Index: Physica B: Condensed Matter; 2000; v.284 (III), p.1944–1945
SciSearch: Physica B, Jul 2000 v. 284(pt.2) pp. 1944–1945
INSPEC: Physica B; July 2000; vol.284–288, p.1944–5

Difficulty arises in determining if a second or subsequent issue can be ignored. This might make the full-text linking successful in some A&I databases and for some publishers, but break the linking mechanism for other databases and publishers.

4.3. Page numbers

Frequently encountered problems in page numbers include typographical errors, reversing digits, adding digits or losing digits. The following examples all have extra digits.

INSPEC: Biofizika; 1982; vol.27, no.6, p.10004 (should be p.1004)
INSPEC: Computers and Structures; 1984; vol.18, no.6, p.10005–8 (should be p.1005–8)
Engineering Index: J Phys E Sci Instrum; Dec 1970; v.3, no.12, p.10006–08 (should be p. 1006–08)

The next two examples are of database records missing leading digits in the page numbers.

SciSearch: Journal Of Geophysical Research, v. 87(#B12) pp. 69–82 1982
SciSearch: Journal Of Biological Chemistry, v. 258(#24) pp. 5037–5045 1983

The last example could be corrected from the database side. For specific ISSNs, after a certain issue date, “10000” can be added to correctly reflect the right starting page number; however, determining the specific journals affected is a difficult process

4.4. Date

The date is fairly straightforward, right? Not always so! Most matching algorithms ignore month, season, day, and simply match on the year. Determining if the year is correct can be tricky. One whole issue of *Macromolecules* was given the date of 1885 instead of 1995 in one of the A&I databases and so full-text linking for that issue in that particular database does not work. Some A&I database producers have processes to correct data but these are usually slow with minimal feedback to the users. If the A&I database producers correct the error, then the corrections must filter down to all of the redistributors, etc.

4.5. Author

Formulating an algorithm that can accurately parse author names when there are so many possibilities is very challenging. The traditional practice of inverting author names is no longer practiced by some databases. Some databases have switched to direct order. Some databases invert author names except for Asian names, which are presented in direct order. This practice increases the inconsistency of author names in even one database. Inconsistencies are also introduced for romanized forms of foreign names. Another area of inconsistency is whether a database packs author initials and surnames. Most linking syntaxes do not require author names, probably due to the wide amount of inconsistency. However, ACM and JSTOR require the last name of the first author for full-text linking. The examples below highlight some of the variations in author names.

INSPEC: Le Boudec, J.-Y.; Hebuterne, G.
SciSearch: LeBoudec JY, Hebuterne G

INSPEC: Woei-Shyan Lee; Chi-Feng Lin
SciSearch: Lee WS, Lin CF

DOE Energy: Abd El-Salam, F.; Abd El-Khalek, A.M.; Nada, R.H.; Mostafa, M.T.
INSPEC: El-Salam, F.Abd.; Mostafa, M.T.; Nada, R.H.; El-Khalek, A.M.Abd.
SciSearch: El-Salam FA, Mostafa MT, Nada RH, El-Khalek AMA

4.6. Title

Accommodating differences in how different databases handle titles is another challenge to linking. Some A&I databases omit leading articles in titles; others retain leading articles. Just this simple omission could break a SICI based linking system. The treatment of mathematical notation, chemical formulas, subscripts, superscripts and other special characters differs from database to database. For example, JSTOR ignores mathematical formulas and symbols that occur in article titles when computing title codes for their SICIs. These two titles demonstrate very different treatments of subscripts and superscripts.

INSPEC: Zeros of J{sub 1}{sup 2}({zeta})-J{sub 0}({zeta})J{sub 2}({zeta}) = 0 with an application to swirling flow in a tube

SciSearch: THE ZEROS OF J1 (2)(ZETA)-J0(ZETA)J2(ZETA) = 0 WITH A APPLICATION TO SWIRLING FLOW IN A TUBE

5. Linking from a bibliography to full-text

The next linking path to discuss is from a bibliography to full-text, an area which LANL is just beginning to explore. We identified three articles and studied their citation history. An original article was identified in SciSearch® @ LANL. Citing articles in electronic form were located and the specific citation was extracted. Citations that were identical were eliminated. We then examined the different forms of citations to determine if links to the full-text would be successful.

The first original article is:

Generalized Privacy Amplification, Bennett CH, Brassard G, Crepeau C, Maurer UM, IEEE Transactions On Information Theory Nov 1995 v. 41(#6/pt.2) pp. 1915–1923.

Eight different cited forms of this article are given below:

1. Bennett C H, Bassard G, Crepeau C and Maurer U 1995 IEEE Trans. Inform. Theory 41 1915
2. C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, IEEE Trans. Inf. Theory 41, 1915 (1995)
3. C. H. Bennett et al., IEEE Trans. Inf. Theory 41, 1915 (1995)
4. C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, "Generalized privacy amplification," IEEE Trans. Inform. Theory, vol. 41, pp. 1915–1923, Nov. 1995
5. BENNETT, C. H., BRASSARD, G., CREPEAU, C., and MAURER, U. M., 1995, IEEE Trans. Info. Theory, 41, 1915
6. C. H. Bennett, G. Brassard, C. Crepeau, and U. Maurer, Generalized privacy amplication, IEEE Trans. Inform. Theory, 41 (1995), pp. 1915–1923
7. C. H. Bennett, G. Brassard, C. Crepeau, and U. M. Maurer. Generalized privacy amplication. IEEE Trans. Info. Theory, vol. 41, no. 6, pp. 1915–1923, 1995
8. C.H. Bennett, G. Brassard, C. Crepeau, U.M. Maurer, IEEE .Trans. Inf. Theory 41 1995 1915

The first author is easily identified in all of these examples. Placement of the date is extremely varied but the date could be parsed if one knows the citing journal's bibliography style. The journal title is fairly consistent, with the abbreviation variations for the word *information* within reason. The labeling of the volume, issue and start page makes these citations much more parsable. Removing those identified pieces make the date easier to identify. A full-text link to citation no. 7 can be made with no modification to the current algorithm for IEEE journals. At LANL, the Science Server parser is used for IEEE journals. Volume, issue, and page are clearly identified, enabling the date to be parsed correctly. If we assume consistency with a known bibliography style, the date is extractable. Then, if we remove the use of the issue in the matching algorithm (and assume continuous paging), full-text links for the remaining citation forms could be formed but we are making lots of large assumptions.

A second article studied is:

Kidera A. 1995. Enhanced conformational sampling in Monte Carlo simulations of proteins: Applications to a constrained peptide. *Proc Natl Acad Sci USA 9*2:9886–9889.

Nine citations are displayed below:

1. A. Kidera, Proc. Natl. Acad. Sci. U.S.A. 92, 9886 (1995)
2. Kidera, A. Proc. Natl. Acad. Sci. U.S.A. 1995, 92, 9886
3. A. Kidera, Proc. Natl. Acad. Sci. USA 92, 9886 (1995)
4. A. Kidera, A. Proc Natl Acad Sci USA 1995, 92, 9886
5. A. Kidera, Proc. Natl. Acad. Sci. USA 92 1995 9886
6. Kidera A. 1995. Enhanced conformational sampling in Monte Carlo simulations of proteins: applications to a constrained peptide. Proc. Natl. Acad. Sci. USA 92:9886–89
7. Kidera A. (1995). Enhanced conformational sampling in Monte Carlo simulations of proteins: application to a constrained peptide. Proc. Natl Acad. Sci. USA, 92, 9886–9889
8. Kidera A. Enhanced conformational sampling in Monte Carlo simulations of proteins: Application to a constrained peptide. Proc Natl Acad Sci USA 1995; 92: 9886–9889
9. Kidera A. (1995) Enhanced conformational sampling in Monte Carlo simulations of proteins: application to a constrained peptide Proc. Natl Acad. Sci. USA 92 9886–9889

The journal abbreviations are incredibly similar, especially numbers 1–5; for these, one only needs to know the order of the elements. Trying to deduce date versus volume versus page number could be very tricky. Punctuation and placement are clues in numbers 6–9, while numbers 7 and 8 have the date in parentheses. However, the current URL structure for *Proceedings of the National Academy of Science,* a Highwire Press title, requires an issue number, so without a lookup table to match pages to issues a full-text link cannot be built for any of these citations.

A third article citation from scisearch® at lanl is:

Scalar glueball mixing and decay: art. no. 014022. Burakovsky L, Page PR, Physical Review D v. 5901(#1) pp. 4022-& JAN 1, 1999.

Six citations are given below:

1. L. Burakovsky and P.R. Page, Phys. Rev. D 59, 014022 (1999)
2. L. Burakovsky, P.R. Page, Phys. Rev. D 59 1999 014022
3. L. Burakovsky, P. R. Page, Phys Rev D59 (1999) 014022; erratum ibid. 079902
4. L. Burakovsky, P.R. Page, Phys. Rev. D59 (1999) 014022;
5. L. Burakovsky, P.R. Page, Phys. Rev. D 59 1999 014022, 079902 E
6. M.Strohmeier-Presicek,T.Gutsche,R.Vinh Mau,Amand Faessler,phys.rev.d 60,054010 (1999);l.burakovsky, P.r.page,phys.rev.d 59,014022 (1999)

Note how differently the database and the cited references treat the article metadata. This example also deals with a journal part. Smart parsing would be required to determine that this is Physical Review D. Punctuation exists in most cases to help determine the date; however, for numbers 3 and 4, handling multiple citations, as when the erratum is included, would be difficult. Citation number 6 is a multiple citation, which would require parsing on a

semi-colon to handle correctly. All of these citations include the required metadata to build a full-text link.

We have only looked at three examples. Reference librarians might say we chose very basic, well-formulated citations. The examples are straightforward but do point out numerous difficulties in linking from a bibliography reference to the full-text, which require extremely capable rule-based parsers.

6. Using metadata beyond reference linking

It is important to mention in this context that in addition to full-text reference linking, SFX and other linking servers can provide additional links to other library-defined services called "extended services." These services are almost exclusively dependent on metadata retrieval. For example, SFX can offer an author search out into another database, but only if the author name is available as part of the metadata the rules database examines. It is not uncommon for an OpenURL to contain a simple identifier that is then used by the linking system to actually fetch additional metadata that can then be used for "extended services" linking. A related initiative is the CrossRef Metadata Database [14]. CrossRef is a reference database where publishers deposit minimal sets of metadata for an article and its corresponding DOI. CrossRef can be queried using a DOI and retrieve metadata and vice-versa. Since DOIs are meant to facilitate cross-publisher reference linking, it is important that this mechanism be available. In a larger context, CrossRef can also be used by linking services such as SFX to retrieve some metadata for "extended services" links and appropriate copy linking [15]. But as is the case with other types of linking, the quality and quantity of metadata (both sent as a query into CrossRef for a DOI lookup, and returned from CrossRef from a DOI query) affects what metadata and how it is used by the local linking system.

7. Conclusions

None of the problems highlighted are insurmountable. Metadata inconsistencies continue to be a prominent problem when it comes to automating information resources. In the world of linking, metadata consistency simply rises in importance. Some, possibly idealistic, goals we can hope for in the future would include:

- Increased consistency in metadata within a single database and across databases. This would result in a higher success rate of linking and would allow the algorithms to be simpler. Simpler algorithms are easier to maintain and modify.
- Increased communication between primary publishers and secondary publishers. Metadata corrections and updates need to be better coordinated.
- Increased awareness of bibliographic/citation standards by authors. Increased submission of publications with bibliographical references reflecting the accepted standards.
- Increased outreach by librarians to authors emphasizing and promoting the importance of citation standards for electronic document retrieval.

Today's systems rely heavily on metadata for linking and standardization may lead toward more useful systems. Several new mechanisms, such as the OpenURL and the DOI, are becoming available and will lessen the impact of metadata inconsistencies on linking systems. Many content providers are realizing the importance of having well-tagged metadata and identifiable 'link-to' syntaxes. In the new paradigm of the digital library, metadata becomes more than a descriptive resource; it becomes a tool in and of itself.

References

[1] Needleman, M. Meeting Report of the NISO Linking Workshop. February 11, 1999. http://www.niso.org/news/events_workshops/linkrpt.html.
[2] Caplan, P., & Arms, W. Y. Reference linking for journal articles. *D-Lib Magazine*, July/August 1999. http://www.dlib.org/dlib/july99/caplan/07caplan.html.
[3] Caplan, P. A lesson in linking. Library Journal netConnect, Fall 2001. http://libraryjournal.reviewsnews.com/index.asp?layout = article&articleid = CA177643.
[4] Van de Sompel, H., & Hochstenbach, P. Reference linking in a hybrid library environment, Part 1: Frameworks for linking. *D-Lib Magazine*, April 1999. http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt1.html.
[5] Ex Libris Information Services Division's SFX. http://www.sfxit.com.
[6] Van de Sompel, H., & Hochstenbach, P. Reference linking in a hybrid library environment, Part 2:SFX, a Generic Linking Solution. *D-Lib Magazine*, April 1999. http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt2.html.
[7] The Digital Object Identifier System, run by the International DOI Foundation. http://www.doi.org.
[8] Hellman, E. S. Scholarly Link Specification Framework (S-Link-S). 1999. http://www.openly.com/SLinkS/SlinkS.html.
[9] Lagoze, C. The Warwick Framework: a container architecture for diverse sets of metadata. *D-Lib Magazine*, July 1996. http://www.dlib.org/dlib/july96/lagoze/07lagoze.html.
[10] NISO Committee AX OpenURL website. http://library.caltech.edu/openurl/.
[11] NISO. NISO/DLF/CrossRef Workshop on Localization in Reference Linking: Meeting Report. July 24, 2000. http://www.niso.org/news/events_workshops/CNRI-mtg.html.
[12] Van de Sompel, H., & Beit-Arie, O. Open linking in the scholarly information environment using the openURL framework. *D-Lib Magazine*, March 2001. http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html.
[13] Caplan, P., & Flecker, D. Choosing the appropriate copy. Digital Library Federation Architecture Committee Report. September 1999. http://www.niso.org/news/reports/DLFarch.html.
[14] For further information about CrossRef, see the FAQ at http://www.crossref.org/faqs.htm. The question "What are the components of CrossRef's reference linking system?" provides a brief definition of the CrossRef Metadata Database.
[15] Beit-Arie, O, et al. Linking to the appropriate copy: Report of the DOI-based prototype. *D-Lib Magazine*. September 2001. http://www.dlib.org/dlib/september01/caplan.htm/. LA-UR-02–1261.