

D-Lib Magazine

October 1999

Volume 5 Number 10

ISSN 1082-9873

Reference Linking in a Hybrid Library Environment

Part 3: Generalizing the SFX solution in the "SFX@Ghent & SFX@LANL" experiment

Herbert Van de Sompel
Los Alamos National Laboratory - Research Library
herbert.vandesompel@rug.ac.be

Patrick Hochstenbach
Automation Department of the Central Library
University of Ghent, Belgium
patrick.hochstenbach@rug.ac.be

Abstract

This is the third part of our papers about reference linking in a hybrid library environment. The [first part](#) described the state-of-the-art of reference linking and contrasted various approaches to the problem. It identified static and dynamic linking solutions, open and closed linking frameworks as well as just-in-case and just-in-time linking. The [second part](#) introduced SFX, a dynamic, just-in-time linking solution we built for our own purposes. However, we suggested that the underlying concepts were sufficiently generic to be applied in a wide range of digital libraries.

In this third part we show how this has been demonstrated conclusively in the "SFX@Ghent & SFX@LANL" experiment. In this experiment, local as well as remote distributed information resources of the digital library collections of the [Research Library of the Los Alamos National Laboratory](#) and the [University of Ghent Library](#) have been used as starting points for SFX-links into other parts of the collections. The SFX-framework has further been generalized in order to achieve a technology that can easily be transferred from one digital library environment to another and that minimizes the overhead in making the distributed information services that make up those libraries interoperable with SFX.

This third part starts with a presentation of the SFX problem statement in light of the recent discussions on reference linking. Next, it introduces the notion of global and local relevance of extended services as well as an architectural categorization of open linking frameworks, also referred to as frameworks that are supportive of selective resolution. Then, an in-depth description of the

generalized SFX solution is given.

Rephrasing the SFX problem statement

The problem statement

It is relevant to rephrase the SFX problem statement in the context of the meetings and the subsequent reports and publications on reference linking organized by the [Digital Library Federation \(DLF\)](#), the [National Information Standards Organization \(NISO\)](#), the [National Federation of Abstracting and Indexing Services \(NFAIS\)](#), and the [Society for Scholarly Publishing \(SSP\)](#) ([Caplan 1999a](#); [Caplan 1999b](#); [Caplan & Arms 1999](#); [Needleman 1999](#)).

The generic statement of the reference linking problem, as defined by the working group on reference linking was ([Caplan 1999a](#); [Caplan & Arms 1999](#)):

Given the information in a standard citation, how does one get to the thing to which it refers?

However, the working group concentrated on a specific variation on this:

Given the information in a citation to a journal article, how does a user get from the citation to an appropriate copy of the article?

The SFX research also addresses these problems, but only as an instance of a more general problem that can be formulated as:

Given bibliographic metadata, how does one present relevant extended services for it?

Bibliographic metadata as a starting point

Clearly, the SFX research is not only concerned about information in a standard citation. Its starting point is bibliographic metadata in general. As such, information entities originating from typical scholarly resources such as records from abstracting & indexing databases, OPAC systems and preprint archives can be used as a starting point in the SFX problem statement. This is also the case for citations to both journal articles and books found in journal articles or books. But even fractional bibliographic metadata such as an author name taken from an e-mail message is a valid starting point in the SFX problem statement.

Extended services as a goal

A similar generalization holds for the target of the problem statement since the SFX research is not only concerned about linking to the full-text that corresponds to a citation in a journal article. It aims at the presentation of a variety of extended services for whichever metadata is used as a starting point. Extended services are services that present an information entity in a digital library -- defined as the link-source -- in the context of the entire information environment ([Van de Sompel & Hochstenbach 1999a](#)). For instance, for a given link-source record from an abstracting & indexing database, extended services can -- amongst others -- be the presentation of:

- the full-text of the paper that is abstracted in the link-source;
- a record abstracting the same publication taken from another abstracting &

indexing database;

- citation information corresponding with the link-source;
- library holdings for the journal in which the article described by the link-source appeared.

Global and local relevance of extended services

The adjective *relevant* is of particular importance in the notion *relevant extended services* as used in the SFX problem statement. It actually has two meanings: relevance as a global notion and relevance as a local notion. In order to explain this, the following types of extended services are considered:

- *full_text*: a service providing the full-text that is referred to by a link-source;
- *review*: a service showing a book review for the item referred to by a link-source;
- *abstract*: a service that provides the abstract from an abstracting & indexing database for a link-source.

Relevant as a global notion must be interpreted as being opposed to irrelevant in every context. Certain aspects of extended services are independent of the context of an individual collection; they actually apply on a global level:

- *full_text*: If the publication year of an article is equal to or higher than that of the first electronic issue of the journal in which the article was published, a *full_text* service has global relevance. On the other hand, it never makes sense to present a *full_text* service for a link-source referring to a paper in a journal if the publication year of the paper is lower than the publication year of the first issue of the journal for which full-text is globally available. As such, the publication year of the first electronic issue is a constraint of global significance to the *full_text* service.
- *review*: It is always irrelevant to present a book *review* service if the link-source refers to a journal article. But, if the link-source describes a book, such a *review* service is globally relevant. In this case, the material type is a constraint of global significance for the *review* service.
- *abstract*: A constraint of global significance rules the relevance of an *abstract* service that looks up the abstract of a citation to a journal article in a particular abstracting & indexing database. Such a service is globally relevant if the journal in which the article is published is actually indexed in that abstracting & indexing database and is globally irrelevant otherwise.

Relevant as a local notion, refers to the fact that other aspects of extended services are dependent on the boundaries of a certain digital library collection. Local relevance has two manifestations:

- Relevance related to the content of a local collection:

While certain services are relevant in a global sense, they can become irrelevant if the digital library collection does not contain the information

resource(s) required to implement them. Even if a *full-text* service is globally relevant for a certain link-source, it might be considered to be irrelevant in the context of a certain digital library collection if the journal referred to by the link-source is not part of that collection. In the same way, an *abstract* service pointing to a particular abstracting & indexing database for a given link-source can be globally relevant, as described above. Still, such a service is of no local relevance if the user's digital library does not provide access to an implementation of that particular database, while it can be of local relevance if the digital library does.

- Relevance related to the implementation of a local collection:

The relevance of extended services will also depend on the technical implementation of the information resource(s) required to create the services. When a *full_text* service is globally relevant -- an electronic edition of an article exists -- as well as relevant in relation to the content of a certain collection -- the users of the digital library are authorized to access the electronic edition -- it can be regarded inappropriate to let the *full_text* service link to a full-text instance at a publisher's site, when the digital library holds an instance in its local storage. In the DLF reference linking discussion, this issue was given the name of "the Harvard problem" ([Caplan 1999a](#)). Similar problems occur in the broader scope of extended services. For instance, as shown before, an *abstract* service can be globally relevant -- the journal in which an article was published is abstracted in a particular abstracting & indexing database -- as well as relevant in relation to the content of the collection -- the local digital library does provide access to the particular database. Still, the service might be irrelevant in relation to the implementation, if the actual implementation of the database does not support a mechanism to link into it using the parameters required to do an *abstract* look-up.

Systems supportive of selective resolution

Both issues regarding the local relevance of extended services indicate the need for open linking solutions that take the context of the local collection into account when links are presented to a user ([Van de Sompel & Hochstenbach 1999a](#)). When addressing the Harvard problem the DLF reference linking discussions have referred to open linking solutions as being supportive of selective resolution ([Caplan & Arms 1999](#)). From the above, it can be seen that the problem of local relevance of extended services is actually a generalization of certain aspects of the Harvard problem. As such, when a framework is able to present an approach to deal with the broader problem, the approach will also contain valuable elements to address the narrower Harvard problem.

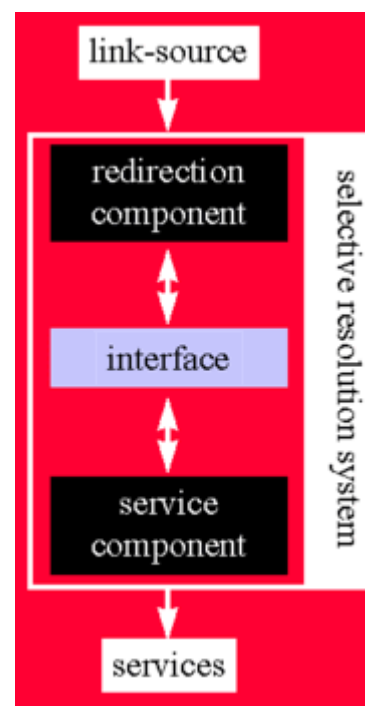


Figure 1: Systems supportive of selectiv

In relation to the Harvard problem, Caplan and Arms divide systems that support selective resolution into two categories:

- Systems with a non-local location database, to which institutions provide a profile describing their full-text collection. The profile controls the selection of links returned to users of that profile.
- Systems with an institutional location database describing the local full-text collection and a global location database as a fall back. In addition to that, there is a mechanism to pass resolution requests to the local resolver first and in the event of a local full-text instance not being found there, to the global resolver.

This categorization can further be generalized by:

- Broadening the scope of the services to be provided beyond the restriction to the full-text, taking into account all kinds of extended services.
- Identifying the crucial components of systems supportive of selective resolution (see [Figure 1](#)):
 - The redirection mechanism that brings metadata of the link-source for which extended services are requested from the information resource to which the link-source belongs to the service component. The redirection mechanism addresses the problem that has been referred to as grabbing the link-source ([Van de Sompel & Hochstenbach 1999a](#)).
 - The service component that takes metadata from whichever information resource in the digital library collection as an input, delivering extended services as an output. The service component is an extension of the location database referred to by Caplan and Arms.
- Recognizing that the order of the redirection is subject to variation:
 - Redirection of the link-source metadata to the local service component first, using a central service component as a means to complete the set of services that can be presented.
 - Redirection of the link-source metadata to the central service component, whose default services can be overwritten and/or completed after communication with the local service component.

CATEGORY			
Category 1		central	central
Category 2	a	central & local	local => central
	b	central & local	central => local

Category 3	local	local
	SERVICE COMPONENT	REDIRECTION ORDER

Table 1: categorization of systems supportive of selective resolution

The resulting categorization is represented in [Table 1](#), where 3 main categories of systems supporting selective resolution are shown, based on the nature of the service component and the redirection order:

- Category 1 only has a central service component and hence a central redirection mechanism. To some extent, this is the category under which the NCBI [LinkOut](#) solution resides. Still, since that solution is tied in with the PubMed database and cannot be used in connection with other resources, it can hardly be seen as a real service component in the sense described earlier.
- Category 2 has both a central and a local service component that contribute to the presentation of the services. Also, there is some form of communication between both. For this Category, it is possible to imagine both approaches regarding the redirection order mentioned above.
- Category 3 builds purely on a local service component and hence also needs a local redirection mechanism. The SFX implementations of both the Elektron and "SFX@Ghent & SFX@LANL" experiments fall within this Category.

The "SFX@Ghent & SFX@LANL" experiment

In the "SFX@Ghent & SFX@LANL" experiment (April 1999 - June 1999; henceforth referred to as Ghent&LANL), the [Library Without Walls](#) team of the [Research Library at the Los Alamos National Laboratory \(LANL\)](#) and the Automation Department of the [Central Library at the University of Ghent](#) have cooperated to illustrate the feasibility of the SFX approach as a means to provide extended services in a realistic and complex information environment.

The information environment in which Ghent&LANL has been conducted is dramatically different from the one of the first Elektron SFX experiment. To illustrate, [Table 2](#) presents an overview of the information resources used in Ghent&LANL. The rows show the names of the information resources used in the experiment, the columns refer to the digital library collection. For each resource/collection combination the table indicates:

- The Type of resource: OPAC system, abstracting & indexing database (A&I), full-text collection (FTXT) or web-service (WWW);
- The Authority running the resource;
- Whether within the digital library collection, the resource is used as a Source. If so, information entities from the resource can be link-sources for which extended services can be requested. If a resource is a Source, the authority running it has made it SFX-aware;
- Whether within the digital library collection, the resource is used as a

Target. If so, the resource is used to be linked into in order to provide extended services. If a resource is a Target, a link-to syntax has been developed by the authority running the resource, in order to allow for it to be the Target of dynamic SFX-links.

RESOURCE		GHENT			LANL	
	Type	Authority	Source	Target	Authority	Source
Advance	OPAC	-	-	-	LANL	yes
Aleph 500	OPAC	Ghent	yes	yes	-	-
Amazon.com	WWW	Amazon	no	yes	Amazon	no
Antilope	OPAC	UA	no	yes	-	-
APS PROLA	FTXT	APS	yes	yes	APS	yes
the arXiv	FTXT	LANL	yes	yes	LANL	yes
BIOSIS	A&I	Ghent	yes	no	LANL	yes
Books in Print	A&I	Ghent	yes	yes	Ghent	yes
Compendex	A&I	Ghent	yes	no	LANL	yes
Current Contents	A&I	Ghent	yes	yes	Ghent	yes
EconLit	A&I	Ghent	yes	no	-	-
Genome base	A&I	NCBI	no	yes	NCBI	no
Inspec	A&I	-	-	-	LANL	yes
		SP	no	yes	SP	no
Ulrich's	A&I	Ghent	yes	yes	-	-

LiSa	A&I	Ghent	yes	yes	-	-
MathSci	A&I	Ghent	yes	no	-	-
Medline	A&I	Ghent	yes	no	-	-
		NCBI	no	yes	NCBI	no
SciSearch	A&I	LANL	yes	yes	LANL	yes
		LANL	no	yes	LANL	no
ScienceServer	FTXT	LANL	no	yes	LANL	no
Various	FTXT	various	no	yes	various	no
		Wiley	yes	yes	Wiley	yes

Table 2: information resources in Ghent&LANL

Some considerations regarding [Table 2](#):

- As can be seen, some resources are available in both digital library collections, but run on different technical implementations. This is the case for BIOSIS and Compendex, which in Ghent run on a [SilverPlatter](#) ERL platform, while LANL -- at the time of the experiment -- used a [Geac](#) Advance implementation.
- For the purpose of this experiment, Ghent and LANL share some of their resources. Ghent makes its SilverPlatter ERL version of Books in Print and Current Contents available for LANL, whereas LANL opens access to its Topic implementation of the [ISI](#) Science Citation Index (SciSearch) and its [ScienceServer](#) storing the full-text of all Elsevier journals.
- Ghent uses two Medline versions: a locally stored ERL version as Source and the NCBI PubMed version as Target. Similarly, LANL uses two Inspec versions: the local [Geac](#) Advance implementation as a Source and an ERL implementation run by [SilverPlatter](#) in Boston as a Target. Time constraints that prevented the development of appropriate link-to syntaxes for the local versions are the reason for this peculiarity.
- Of special importance is the fact that some journals from the [Wiley](#) InterScience collection as well as the complete [PROLA](#) archive of the [American Physical Society](#) are made SFX-aware ([Halstead 1999](#); [Spilka 1999](#)). Both Ghent and LANL can use citations in the full-text of these repositories as link-sources for SFX requests. Also, in the course of this experiment Wiley has implemented a link-to syntax that will be brought into production later in 1999. For the PROLA archive such a link-to syntax was already available.

- Some resource names require a little more explanation. [Aleph 500](#) is the Ghent Integrated Library System, [Advance](#) is the one for LANL, while [Antilope](#) is the Belgian Union Catalogue of Serials run by the University of Antwerp. The header "various" refers to a variety of full-text repositories to which dynamic links are available in this experiment. This is -- amongst others -- the case for [Academic Press](#), [Company of Biologists](#), [HighWire](#), [Springer](#), [American Chemical Society](#), etc. The [arXiv](#) is the Topic implementation of the Ginsparg arXiv e-print repository, developed by the [Library Without Walls](#) team of the LANL Library. It has also been made SFX-aware.
- As can be seen from careful exploration of [Table 2](#), from the point of view of each digital library collection, the SFX-aware information resources are highly distributed. Some resources are run by the institutional library automation team while others are run remotely, actually by three external authorities. From the point of view of Ghent these authorities are LANL, Wiley and the American Physical Society; from the point of view of LANL, they are Ghent, Wiley and the American Physical Society.

From the above, it can be concluded that from the point of view of the amount of resources that are involved, and given their distributed nature and the availability of multiple SFX service components, Ghent&LANL is a very realistic experiment.

The need for a generalization of the SFX components

Although the fundamental concepts of SFX -- dynamic linking, just-in time linking and conceptual services (see ([Van de Sompel & Hochstenbach 1999b](#))) -- have been left untouched for the Ghent&LANL experiment, the nature of its working environment and its goals have led to a strong generalization of the SFX components. The main impulses that inspired such a generalization and that distinguish the Ghent&LANL project from the Elektron experiment are:

- The extension of the digital library collection in which SFX was being tested beyond a well-controlled sub-collection of one institution. In Ghent&LANL, SFX is introduced in the realistic, complex and dissimilar digital libraries of two autonomous institutions each running their local SFX components;
- The extension of the scope of data for which extended services can be requested beyond the internally stored collections. Link-sources in Ghent&LANL also originate from resources held by external authorities;
- The extension of the datatypes for which extended services can be requested beyond abstracting & indexing databases and OPAC systems. Link-sources in Ghent&LANL can also be citations in journal articles;
- The accommodation of extended services linking into target resources, based on metadata in general, not only SICI-related metadata;
- The need for high transportability of the SFX solution between the digital library environments that are involved.

The redesign of the SFX solution for Ghent&LANL leads to an architecture with a clear separation between the redirection component and the service component. Both components obviously interoperate in order to achieve a functional system.

But the redirection component can potentially operate in an environment with non-SFX service components, while the SFX service component can equally function with another redirection mechanism, as long as that supports delivery of link-source metadata to the SFX service component. Several functional building blocks in both components have also been generalized in order to address the problems that arise from the complexity of the Ghent&LANL environment. The overall approach of the generalized solution is shown in [Figure 2](#) and will be explained in more detail in the remainder of this paper. Information resources that can interoperate with SFX -- from now on referred to as SFX-aware systems -- insert an SFX-button for each link-source in the result set of a query. The just-in-time approach of SFX requires the user to click such an SFX-button when requesting extended services for a specific link-source record. In response to this click, the local SFX redirection component will fetch link-source metadata -- usually -- from the origin resource using whichever protocol it takes to do so. Next, link-source metadata as well as information on its origin will be converted into an interfacing format. At this point, the local redirection mechanism has fulfilled its task and is able to deliver this information in a consistent representation to the local SFX service component.

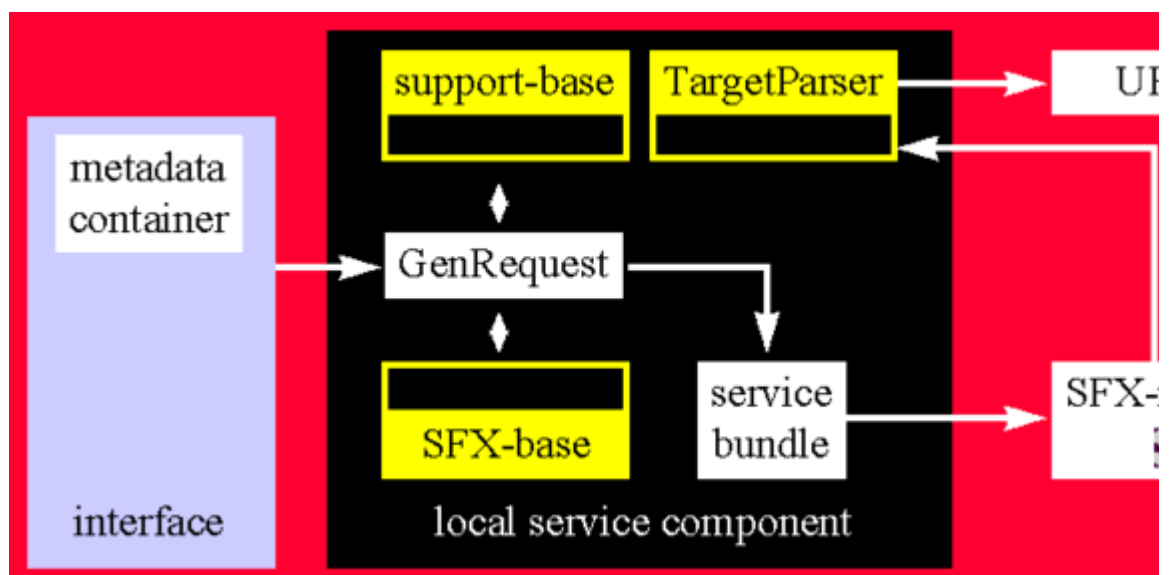
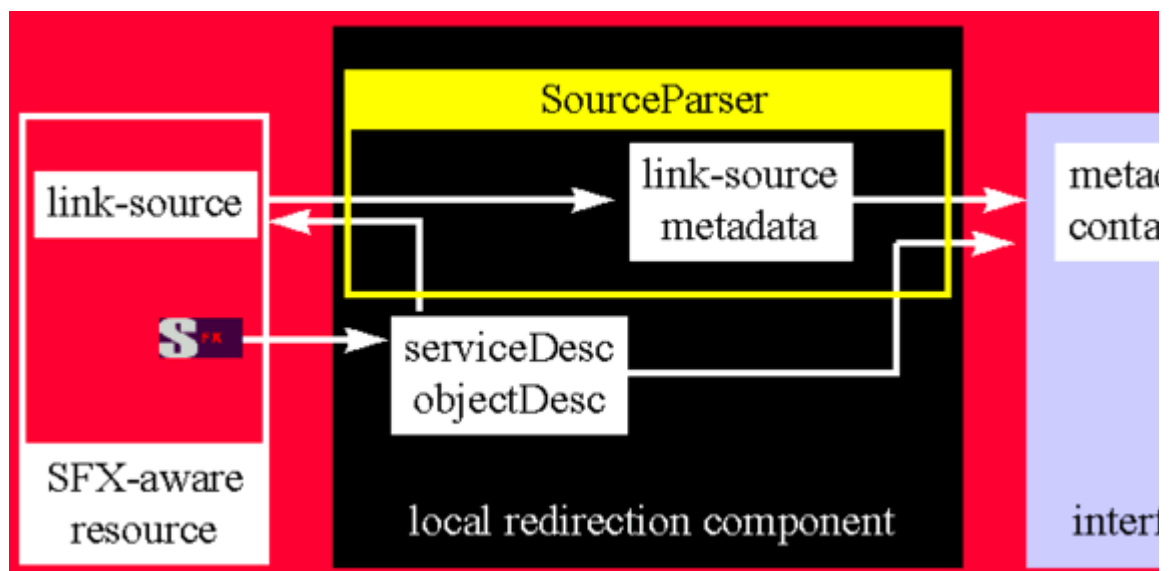


Figure 2: the local redirection and service components of the generalized SFX solution

The first task of the local service component is to parse the information, handed over by the local redirection component, into a normalized internal representation object. During this process, the original content can be enhanced and/or augmented. The resulting information object is then fed into the SFX evaluation process in which it will be compared to the SFX-database. The SFX-database is a special kind of linking database. Unlike traditional linking services, it does not contain any static links between "documents" (records/citations/full-text/etc.) of a collection. Rather, it contains a collection of conceptual services that express potential inter-relationships between documents at the level of the resource from which they originate. The SFX evaluation process determines the relevance of each of these conceptual services using the -- lack of -- content in the information object. Next, the resulting bundle of relevant services is sent back to the user in the SFX-menu-screen. Consistent with the just-in-time approach of SFX, only when the user decides to use a service from the bundle, will the service be resolved into a URL to which the user is being redirected.

The SFX mechanism for local redirection

The task of the local redirection mechanism is to transport link-source metadata to the local redirection component, that interfaces with the local service component. In order to be able to interoperate with the SFX redirection mechanism, information resources need to be enhanced by the authorities running them in order to make them SFX-aware. The aim of this is to create the ability for information resources to insert an SFX-button targeted at the local redirection component for each link-source in the result set of a query into the resource. In the context of Ghent&LANL, the following are important considerations with this regard:

- a. Many information resources that are involved in the experiment are also used in normal production at the very same time. This means that they are also approached by users that do not have access to an SFX service component. In order to prevent such a user from seeing an irrelevant SFX-button, an SFX-aware resource must be able to recognize whether the user has access to an SFX service component or not. Based on that information, the resource can insert an SFX-button or not.
- b. Some information resources are approached by users from both digital library environments, hence with access to different SFX service components. An SFX-aware resource must be able to target the SFX-button at the appropriate local redirection component, in order for it to be able to deliver the link-source metadata from the origin information resource to the doorstep of the appropriate service component. This means that an SFX-aware resource must be able to parameterize the target of an SFX-button.
- c. Upon receipt of a request for extended services from a user, the local redirection component must be able to fetch the link-source metadata from its origin resource. This means that the local redirection component has to be informed about the origin and the identity of the link-source in order to be able to take the appropriate steps. Given the amount, distribution and

diversity of the SFX-aware resources in Ghent&LANL, a consistent manner to communicate such information to the local service components is required.

- d. Link-source metadata must be fetched from a wide variety of distributed information resources that support different access protocols. In addition to that, those resources will respond by sending link-source metadata formatted according to different metadata schemes. In order for the local redirection component to be able to interface in a generic manner with the local service component, a unique metadata interchange format is desirable.

As will be shown, in the detailed description below, these issues are approached by:

- For (a) and (b): the CookiePusher mechanism;
- For (c): the consistent SFX-URL structure;
- For (d): the SourceParser solution.

Making information resources SFX-aware

The authorities running information resources need to enhance their systems in order to make them SFX-aware. The complexity of the Ghent&LANL environment has urged for a thoughtful exploration of ways to make resources SFX-aware, since only approaches that minimize the overhead in doing so for the authorities running the resources can be acceptable and workable. In the current implementation of the SFX redirection mechanism, they have to do this by:

- Installing the CookiePusher script delivered by the project managers of Ghent&LANL;
- Hyperlinking the SFX-buttons for link-sources using a URL that complies to a predefined format.

The CookiePusher

The CookiePusher script is a pragmatic solution introduced to dynamically notify an information resource about the existence and location of a local SFX redirection component in the environment of the user consulting the resource. The underlying idea is that an information resource could at any time access the location of a local redirection component, if its URL were written as a cookie in the browser of the user consulting the resource. The availability of this URL is essential, since the resource must be able to dynamically target the SFX-button at the appropriate local component. However, for reasons of security and privacy, such browser cookies can maximally be read within the Internet domain of the server that has set the cookie (see [Shishir 1996](#) pages 203-204). As such, it is impossible to set such a cookie so that it can be read by all information systems in a digital library collection when it consists of resources distributed over several domains, typically resources that are local and remote to the user's institution.

In order to solve this problem, the first step in connecting to a resource is to

request a server in the domain of the information resource to create an HTTP cookie. This detour is called the CookiePusher. The very simple CookiePusher script is installed in the domain of the information resource that has to be made SFX-aware. Rather than connecting immediately to the desired URL in the information resource, a connection is made to the resource's CookiePusher first, sending values for the two parameters of the CookiePusher script:

- **SFX_location**: the URL of the local redirection component of the SFX solution;
- **Redirect**: the desired URL in the resource.

Upon receipt of these parameters, the CookiePusher will first read the URL of the local redirection component and will use it to set a cookie in the user's browser. Since the CookiePusher is in the domain of the resource, that cookie will be readable by the resource. Next, the CookiePusher will redirect the user to the desired URL in the resource.

As such, once the CookiePusher has been installed for a resource, the URL to connect to that resource will be changed to:

CookiePusher_URL?SFX_location= local_SFX& Redirect= service_URL

Where

- **CookiePusher_URL** is the URL of the CookiePusher script;
- **local_SFX** is the URL of the local SFX redirection component;
- **service_URL** is the desired URL in the information resource as used under normal -- non-SFX -- conditions. Such a URL can point at the initial search screen for an abstracting & indexing database, it can be a URL linking to an article at a publishers site, etc.;
- **local_SFX** and **service_URL** are URL-encoded.

For instance:

```
http://publish.aps.org/edaccess/prolatest/cookiepusher?  
SFX_location=http%3A%2F%2Ffisiserv.rug.ac.be%2Fcgi-bin%2Fsfx%2Fbin%2Fmenu.cgi  
&Redirect=http%3A%2F%2Fpublish.aps.org%2Fedaccess%2Fprolatest%2Ftext%2FPRD%2Fv!
```

is the URL used to connect to an item in the [APS/PROLA](#) domain. The APS/PROLA CookiePusher will read the location of the local redirection component from the SFX_location parameter and will use this to set a cookie named local_SFX with value:

```
http%3A%2F%2Ffisiserv.rug.ac.be%2Fcgi-bin%2Fsfx%2Fbin%2Fmenu.cgi
```

which is the encoded location of the Ghent local SFX redirection component. Next, it will redirect the user to the desired location in the APS/PROLA:

```
http://publish.aps.org/edaccess/prolatest/text/PRD/v52/i1/p15_1
```

From now on, at any point in the consultation, APS/PROLA will be able to read this cookie and use it to target -- in this case -- the Ghent redirection component.

The consistent SFX-URL structure

The essence of the detour made via the CookiePusher is the ability it creates for an information resource to know at any point whether the consulting user has access to a selective resolution system and, if so, what the location of its redirection component is. Based on that information, the resource can dynamically decide whether or not to insert an SFX-button for search results and if it does, which redirection component to target with the SFX-button. In order to make the many systems involved in the Ghent&LANL experiment interoperable with SFX, authorities running the systems have been asked to make the URL targeted by the SFX-button -- the SFX-URL -- compliant to the following format:

GENERAL	target? serviceDesc & objectDesc
DETAILED	local_SFX?vendorId=<theVendor>&databaseId=<theBase>&objectDesc=<theIdentifier>

Table 3: the syntax of the SFX-URL

In [Table 3](#)


- target is the URL of the local redirection component of the SFX solution;
- **serviceDesc** uniquely defines the origin resource. It contains information on the vendor of the resource and on the resource itself. It is of the form:

vendorId=<theVendor>&databaseId=<theBase>.


serviceDesc information will play a crucial role at later stages of the SFX local redirection mechanism, as well as in the SFX-base which is central to the SFX service component.

- **objectDesc** contains information that relates to the identity of the link source. Its syntax and content is extremely flexible and it will be defined by the authority running the resource, making it dependent on the vendor and his database implementation. objectDesc typically contains the unique record identifier for a link-source in its origin resource. Alternatively or in addition to that, it can contain SICI-like metadata. In some cases, it can even contain all metadata of the link-source.
- The parameter values **<theVendor>**, **<theBase>** and **<theIdentifier>** are URL-encoded.

[Figure 3](#) to [Figure 6](#) show examples of link-sources taken from Sources in the Ghent and/or LANL collections, mentioning their SFX-URL. For reasons of readability, the parameter values are not shown as being URL-encoded. Rather, it is mentioned that parts should be URL-encoded by enclosing them in a URLEncode function.

<p>☐ Record 2 of 33 in Biological Abstracts 1999/01-1999/03</p>	
TI	Long term outcome of patients with hairy cell leukemia treated with pentostatin.
AU	Ribeiro-Patricia; Bouaffia-Fadhela; Peaud-Pierre-Yves; Blanc-Michel; Salles-Bruno
AD	(a) Serv. Hematol., Cent. Hosp. Lyon-Sud, 165 Chemin du Grand Revoyet, 69495
SO	Cancer, Jan. 1, 1999, 85 (1) 65-71..
PV	1999
DT	Article-
IS	0008-543X
LA	English
AI	Y
ST	Hominidae-; Primates-; Mammalia-; Vertebrata-; Chordata-; Animalia-
RN	22910-25-1; PENTOSTATIN
AN	199900063465
UD	19990223
AV	

□ Record 2 of 33 in Biological Abstracts 1999/01-1999/03

TI Long term outcome of patients with **hairy cell leukemia** treated with pentostatin.
AU [Ribeiro-Patricia](#); [Bouaffia-Fadhela](#); [Peaud-Pierre-Yves](#); [Blanc-Michel](#); [Salles-Bruno](#)
AD {a} Serv. Hematol., Cent. Hosp. Lyon-Sud, 165 Chemin du Grand Revoyet, 69495
SO [Cancer](#)-.Jan. 1, 1999; 85 (1) 65-71..
PY 1999
DT Article-
IS 0008-543X
LA English
AI Y
ST Hominidae-; Primates-, Mammalia-, Vertebrata-, Chordata-, Animalia-
RN 53910-25-1: PENTOSTATIN
AN 199900063465
UD 19990223 .
AV 

SFX-URL for this link-source, pointing at the Ghent local redirection component:

<http://isiserv.rug.ac.be/cgi-bin/sfx/bin/menu.cgi?vendorId=ERL&databaseId=BX>

[&objectDesc=URLEncode\(BX02 A:199900063465 I:0008-543X V:00085 S:000001 P:000065 Y:1999\)](#)

In the serviceDesc part of the URL, ERL refers to the SilverPlatter ERL implementation of BIOSIS, while BX is the family name of BIOSIS databases in the ERL environment. The objectDesc component contains several information elements in a tagged and fixed length representation. BX02 is the volume of the BIOSIS database where the link-source originates, while 199900063465 is the accession number, a unique record number of the link-source in BIOSIS. Other elements in the objectDesc are ISSN number, volume, issue, starting page and publication year.

Figure 3: a link-source from the Ghent ERL implementation of BIOSIS and its SFX-URL

Record 3 of 91  

Article:  [Article](#)

Title: [The identification of cDNAs that affect the **mitosis**-to-**interphase** transition in *pombe*, including *sbp1*, which encodes a *spi1p*-GTP-binding protein.](#)

Author: [HE, XIANGWEI](#); [HAYASHI, NAOYUKI](#); [WALCOTT, NATHAN G.](#); [PATTERSON, THOMAS E.](#); [BISCHOFF, F. RALF](#); [NISHIMOTO, TAI](#)
[SHELLEY](#)

Source: [Genetics](#), Feb. 1998, v.148, no.2, p.645-656.

Other Links: 

Location **Holdings**

MAIN Holdings: v.109- (1985-) ; Missing: v. 134 no. 4 (1993) ; Last rec'd: VOL. 1: issue on display ; Shelved as: GENETICS.

WWW <http://www.genetics.org/> ; Holdings: v. 148, no.1- (Jan.1998-) ; Abstracts and no.2-v.147, no.4 (Feb.1980-Dec.1997)

SFX-URL for this link-source, pointing at the LANL local redirection component:

<http://vole.lanl.gov/cgi-bin/sfx/bin/menu.cgi?vendorId=ADVANCE&databaseId=Biosis>

&objectDesc=

[URLencode\(fetchId=21179970&objectId=PREV199800135979&SICI=0016-6731\(1998\)148:2<645:TIO](http://URLencode(fetchId=21179970&objectId=PREV199800135979&SICI=0016-6731(1998)148:2<645:TIO)

The serviceDesc part of this URL is self-explanatory. The objectDesc component is tagged and fields can lengths. The fetchId is the unique number of the link-source in the LANL implementation of BIOSIS, which is the objectId after "PREV" is the BIOSIS accession number which is comparable to the A field in the SilverPL [Figure 3](#). The SICI part contains a SICI for the link-source, from which ISSN, volume, issue, pagination a year can be derived.

Figure 4: a link-source from the LANL Advance implementation of BIOSIS and its SFX

WILEY
InterScience
PERSONAL HOMEPAGE JOURNAL FINDER SEARCH HELP CONTACT US LOGO
ALL JOURNALS PREVIOUS ARTICLE NEXT ARTICLE
Article Abstract
CANCER
Online ISSN: 1097-0142 Print ISSN: 0008-543X
Cancer
Volume 85, Issue 1, 1999. Pages: 65-71
ADD ART
Original Article
Long term outcome of patients with hairy cell leukemia treated with pentostatin
Patricia Ribeiro, M.D. ¹, Fadhela Bouaffia, M.D. ¹, Pierre-Yves Peaud, M.D. ², Michel Bla
References
1 Saven A, Piro L. Treatment of hairy cell leukemia. *Blood* 1992; **79**: 1111-20. [Medline](#)
2 Jaiyesimi I, Kantarjian H, Estey E. Advances in therapy for hairy cell leukemia. A review *Cancer* 1993; **72**: 5-16. [Medline](#)
3 Saven A, Piro L. The newer purine analogues for the treatment of hairy-cell leukemia. *A J Med* 1994; **330**: 691-7. [Medline](#)

SFX-URL for the third reference as a link-source, pointing at the Ghent local redirection component:

[http://isiserv.rug.ac.be/cgi-bin/sfx/bin/menu.cgi? vendorId=Wiley&databaseId=WIS](http://isiserv.rug.ac.be/cgi-bin/sfx/bin/menu.cgi?vendorId=Wiley&databaseId=WIS)

&objectDesc= URLencode(TYPE=JCIT& SNM=Saven&FNM=A&SNM=Piro&FNM=L&ATL= The ne purine analogues for the treatment of hairy-cell leukemia.&JTL=N Engl J Med &PYR=1994&VID=330&PPF=691&PPL=7)

The serviceDesc component now refers to the Wiley InterScience collection. The objectDesc is tagged with an indication on the material type of the reference -- journal citation in this case -- followed by a tag repetition of the full citation.

Figure 5: a link-source from Wiley InterScience and its SFX-URL

Preprints Retrieval Results - Netscape

File Edit View Go Communicator Help

Los Alamos Research Library
NATIONAL LABORATORY

New Search Up Comments
Mark All List Marks Help

Preprints

Preprints Retrieval Results

23 out of 100245 records matched the query below. 23 records displayed:
(fractal <in> tissuab <and> year >= 1991)<and>year <= 1999<and>physics <in> arc

Marks	Score	Title, Author, Eprint
<input type="checkbox"/>	0.97	Integers and Fractions Diptiman Sen (Indian Institute of Science, Bangalore, India) physics/9811004 (03 Nov 1998)
<input type="checkbox"/>	0.96	Irrational Numbers of Constant Type --- A New Characterization Manash Mukherjee and Gunther Karner physics/9706009 (04 Jun 1997)
<input type="checkbox"/>	0.95	Adaptive Ising Model and Bacterial Chemotaxis Yu Shi physics/9901073 (28 Jan 1999)

SFX-URL for the first link-source in the above result screen, pointing at the LANL local redirection component

[http://vole.lanl.gov/cgi-bin/sfx/bin/menu.cgi?vendorId=LANLTopic&databaseId=arXiv&objectDesc=URLencode\(fetchId=phys-9811004&objectId=physics/9811004\)](http://vole.lanl.gov/cgi-bin/sfx/bin/menu.cgi?vendorId=LANLTopic&databaseId=arXiv&objectDesc=URLencode(fetchId=phys-9811004&objectId=physics/9811004))

The serviceDesc refers to the LANL Topic implementation of the Ginsparg e-print archive. The fetchId is unique key for the record in that implementation, while the -- very similar -- objectId is the unique record number in Ginsparg's implementation of the archive. No further metadata is available in the objectDesc.

Figure 6: a link-source from the arXiv and its SFX-URL

Fetching link-source metadata from an SFX-aware information resource with SourceParsers

The CookiePusher mechanism enables a resource to insert an SFX-button for each of the link-sources that are transferred to a user consulting the resource. The structure of the SFX-URL targeted by these SFX-buttons has been made consistent across resources to be of the form `target?serviceDesc&objectDesc`. When a user requests extended services by clicking such an SFX-button, a request is sent to his local SFX redirection component, which will receive serviceDesc and objectDesc values as parameters for the target script. The local component holds a collection of SourceParser scripts with names corresponding to valid serviceDesc's (see [Table 4](#)). Having analyzed the serviceDesc information, the target script will launch the appropriate SourceParser. This

serviceDesc-specific SourceParser uniquely implements:

- The interpretation of the information contained in the objectDesc parameter based upon the syntax defined by the vendor (see examples in [Figure 3](#) to [Figure 6](#));
- The mechanism to fetch the link-source from its origin resource based on its origin and on the content of its objectDesc. [Table 4](#) shows those fetch mechanisms for the examples of [Figure 3](#) to [Figure 6](#). As can be seen, no real fetching is required for the [Wiley](#) citations, since these are completely transferred in the objectDesc part of the SFX-URL. The same technique is used for citations in the [PROLA](#) archive. Both the Ghent and LANL BIOSIS implementations deliver some -- SICI related -- metadata in the objectDesc. But since several extended services that SFX aims to deliver require more metadata, a fetch is required in order to obtain more complete information. Since the objectDesc for the [arXiv](#) only contains an identifier, a fetch is definitely required;
- The conversion of the fetched link-source metadata, that is expressed in the metadata scheme supported by the authority running the origin resource, into a metadata container compliant with the scheme of the unique metadata interchange format. This metadata container is the interface between the local redirection component and the local service component.

RESOURCE	serviceDesc		SourceParser	Fetch protocol	Fetch key
the arXiv	LANLTopic	arXiv	S::LANLTopic:arXiv	HTTP	f
BIOSIS	ERL	BX	S::ERL::BX	Z39.50	A
BIOSIS	ADVANCE	Biosis	S::ADVANCE::Biosis	Z39.50	f
Wiley	Wiley	WIS	S::Wiley::WIS	none	r

Table 4: Some SFX-aware resources with their serviceDesc, Fetch protocol and Fetch key

The SFX service component

The task of the local SFX service component starts at the point where the local redirection mechanism hands over the metadata container that contains, in a consistent representation:

- link-source metadata that became available through the local redirection mechanism;
- information on the origin of the link-source, basically serviceDesc information.

It is the task of the SFX service component to deliver extended services based on

this information. The following are important considerations regarding the SFX service component in Ghent&LANL:

- a. The amount and quality of link-source metadata that becomes available in the metadata container is dependent on the type of resource from which its link-source originated and on the amount of information that the authority running the origin resource allows and/or supports to be fetched. In some cases such metadata can be corrupt or lack information that is essential for the SFX evaluation process to adequately perform its task;
- b. The SFX service component must be easily transportable between different digital library environments and remain easily manageable;
- c. The SFX service component must ultimately deliver service links in a just-in-time manner.

As can be seen from a detailed description of the SFX service component, these problems have been approached by:

- For (a): the GenericRequest object;
- For (b): a generalization of the implementation of the SFX-database, that explicitly reflects the notion of global and local relevance of conceptual services as well as the notion of global and local Thresholds;
- For (c): the TargetParser solution.

The GenericRequest object

The service component will take the metadata container delivered by the local redirection mechanism as input and turn it into a normalized internal representation, called the GenericRequest object. [Table 5](#) shows a representation of the GenericRequest object for the third citation in [Figure 5](#). The GenericRequest object is an intelligent object, that is able to self-check the validity of its information elements based on pre-configured rules. It can also augment/enhance its content using information from a supporting database. For instance, the citation of [Figure 5](#) does not contain an ISSN number nor a journal title, but rather an abbreviated journal title. In this case, the GenericRequest object augments its content, by adding the missing information via communication with a supporting database. Obviously, the GenericRequest object also contains a normalized version of the link-source metadata, as well as information about its origin.

At the time of the experiment, interoperability between the SFX local service component and non-SFX local redirection mechanisms was not an issue, since none were existing. As such, for reasons of simplicity, the metadata scheme of the GenericRequest object has fulfilled the role of interfacing metadata scheme between the local redirection and the local service component in Ghent&LANL.

```
<perldata>
<hash>
<item key="rec$vendorId">Wiley</item>
<item key="rec$databaseId">WIS</item>
```

```

<item key="rec$dbId">Wiley::WIS</item>
<item key="objectType">JOURNAL</item>
<item key="@abbrevTitle">
<array>
<item key="0">N ENGL J MED</item>
</array>
</item>
<item key="journalTitle">NEW ENGLAND JOURNAL OF MEDICINE</item>
<item key="ISSN">0028-4793</item>
<item key="year">1994</item>
<item key="volume">330</item>
<item key="startPage">691</item>
<item key="endPage">7</item>
<item key="@authLast">
<array>
<item key="0">Saven</item>
<item key="1">Piro</item>
</array>
</item>
<item key="@authInit">
<array>
<item key="0">A</item>
<item key="1">L</item>
</array>
</item>
<item key="articleTitle">The newer purine analogues for the treatment of hair
</hash></perldata>

```

Table 5: Representation of an augmented GenericRequest object for the link-sourc

The SFX linking service and the SFX-base

As a result of the above, an instance of the GenericRequest object for the link-source for which extended services have been requested has become available to the SFX service component. It will be the task of this component to deliver the extended services to the user that has requested them. In this sense, the SFX service component is a linking service that, given a certain input "document", outputs "documents" related to the input. The SFX linking service is special, however, since it does not store static relationships between individual documents. Rather, it stores relationships between the resources from which the

documents originate. In SFX, these relationships are called conceptual services and they are stored in the SFX-base. The SFX evaluation process will determine the relevance of each of these conceptual services based upon the information and origin of a link-source.

The requirement imposed on the Ghent&LANL implementation of the SFX service component to be easily transportable between different digital library environments has led to an important generalization of the design of the SFX-base. This has been achieved by explicitly reflecting the notion of global and local relevance of services in the implementation. A synthesized representation of the lay-out of the Ghent&LANL SFX-base is given in [Figure 7](#).



Figure 7: Simplified lay-out of the SFX-base

Splitting the Colli table

As in the Elektron version of the SFX-base, the Source table contains the information resources that can be origins for link-sources. They are SFX-aware resources. In the Elektron version, the Colli contained conceptual services, directly coupled with the Target resources. (see [Table 2](#) in [\(Van de Sompel & Hochstenbach 1999b\)](#)). Such a set-up was not adequately generic and, in the current design, this Colli has been split. One table has kept the name Colli, the other has been named the Target table. The Target table contains those resources into which linking is possible. The Colli table that connects the Source and Target tables now expresses the type of service that relates Source with Target resources. [Table 6](#) shows the type of services implemented in Ghent&LANL.

COLLI SERVICES	FUNCTION
----------------	----------

<i>abstract</i>	look-up of abstract information in an abstracting & indexing database for the item by the GenericRequest object
<i>author</i>	look-up of references by an author of the item represented by the GenericRequest abstracting & indexing database
<i>cited_author</i>	look-up of citations to work by an author mentioned in the GenericRequest object
<i>cited_reference</i>	look-up of works citing the item represented by the GenericRequest object
<i>full_text</i>	link to the full-text of the item represented by the GenericRequest object
<i>genome</i>	look-up of sequence information found in the GenericRequest object
<i>holding</i>	holdings look-up in an OPAC system for the item represented by the GenericRequest object
<i>review</i>	look-up of a book review for the item represented by the GenericRequest object

Table 6: Services in the Colli and their function

Taking advantage of the global relevance of conceptual services

It is not a coincidence that the resources shown as Source and/or Target carry their globally common names rather than those of their local implementations in Ghent or LANL. This is actually a reflection of the conclusion that services relating Source and Target resources have global relevance. It is globally relevant to deliver an *abstract* service that, given a link-source from BIOSIS shows the corresponding abstract from Medline. Such a conceptual service can be imagined regardless of the implementations of each of these resources in a specific digital library. Therefore, the Ghent&LANL SFX-base expresses the relationships between Sources and Targets at the level of global relevance: there is an *abstract* service connecting BIOSIS and Medline, regardless of their local implementations. A very limited number of examples of how such services of global relevance connect Source and Target is shown in [Table 7](#).

COLLI		
SOURCE	SERVICE	TARGET
APS/PROLA	<i>abstract</i>	Inspec
the arXiv	<i>author</i>	Inspec

BIOSIS	<i>abstract</i>	Medline
BIOSIS	<i>genome</i>	Genome Base
Current Contents	<i>abstract</i>	LiSa
EconLit	<i>review</i>	Books in Print
Inspec	<i>full_text</i>	Springer
Wiley	<i>abstract</i>	Medline
Wiley	<i>cited_reference</i>	Science Cit. Base

Table 7: Examples of service relationships between Sources and Targets

Localization of services of global relevance

While the services shown in [Table 7](#) are of global relevance, they do not take into account issues of relevance in relation to the local digital library collection. This localization of services of global relevance is achieved by:

- The introduction of fields referring to the local implementations, next to the globally common names.

As shown in [Table 8](#) and [Table 9](#), a key reflecting the serviceDesc values of the local implementations of resources -- found in the rec\$dbId field of the GenericRequest object -- is added next to the global common name of the Sources. In the same way, at the Target side, the name of a local TargetParser is added next to the global name of which the local Target is an implementation. The TargetParser procedure implements the link-to syntax into the local implementation of the Target resource. It can be seen from [Table 8](#) and [Table 9](#) that Ghent and LANL use a different SourceParser for BIOSIS, which reflects that they have a different implementation. However, they share a TargetParser to provide the *abstract* service into Medline, since both have chosen the PubMed implementation as a Target to achieve this.

- Deactivating services of global relevance when they are not of local relevance.

When the Source or Target resource required to implement a certain service is not available in the digital library collection, when the local implementation of the Target resource does not support the link mechanism required to implement the service, or when local librarians decide the service to be of no use to their end-users, its flag will be set to inactive. The service will no longer be taken into account in the SFX

evaluation process deciding on the local relevance of conceptual services. In [Table 8](#) this is the case for services with Inspec as a Source since Ghent does not have an Inspec implementation in its collection. In [Table 9](#), this is the case for services with LiSa as a Target, since LANL does not have access to a LiSa implementation.

SOURCE		COLLI	TARGET	
local	global		global	
S::APS::PROLA	APS/PROLA	<i>abstract</i>	Inspec	T::ERL::E
S::LANLTopic:arXiv	the arXiv	<i>author</i>	Inspec	T::ERL::E
S::ERL::BX	BIOSIS	<i>abstract</i>	Medline	T::NCBI::Pub
S::ERL::BX	BIOSIS	<i>genome</i>	Genome Base	T::NCBI::Gen
S::ERL::CCO	Current Contents	<i>abstract</i>	LiSa	T::ERL::E
S::ERL::EC	EconLit	<i>review</i>	Books in Print	T::ERL::E
inactive	Inspec	<i>full_text</i>	Springer	T::Springer::L
S::Wiley::WIS	Wiley	<i>abstract</i>	Medline	T::NCBI::Pub
S::Wiley::WIS	Wiley	<i>cited_reference</i>	Science Cit. Base	T::CIC15:SciSe

Table 8: Localization of services from [Table 7](#) for Ghent

Source		Colli	Target	
local	global		global	
S::APS::PROLA	APS/PROLA	<i>abstract</i>	Inspec	T::ERL::E
S::LANLTopic:arXiv	the arXiv	<i>author</i>	Inspec	T::ERL::E
S::Advance::Biosis	BIOSIS	<i>abstract</i>	Medline	T::NCBI::Pub
S::Advance::Biosis	BIOSIS	<i>genome</i>	Genome Base	T::NCBI::Gen

S::ERL::CCO	Current Contents	<i>abstract</i>	LiSa	ina
inactive	EconLit	<i>review</i>	Books in Print	T::ERL::E
S::Advance::Inspec	Inspec	<i>full_text</i>	Springer LINK	T::Springer::L
S::Wiley::WIS	Wiley	<i>abstract</i>	Medline	T::NCBI::Pub
S::Wiley::WIS	Wiley	<i>cited_reference</i>	Science Cit. Base	T::CIC15:SciSe

Table 9: Localization of services from [Table 7](#) for LANL

Global and local Thresholds

The relationships between Source and Target resources expressed by a service connection in the Colli is made subject to restrictions called Thresholds. These Thresholds are the way to fine-tune conceptual services in order to minimize the presentation of services that are considered not to be appropriate to be presented. In order to illustrate this concept, two types of Thresholds are described:

- Thresholds expressed in terms of boundaries for the metadata elements that make up the GenericRequest object structure. Technically, these Thresholds are expressed as conditional statements using field names of the GenericRequest object. Such Thresholds are in many cases very simple, but they can as well be scripts of whichever degree of complexity. For instance:
 - *cited_author*: In order for a *cited_author* service to be relevant, the lowest Threshold that has to be passed is the existence of an author name in the GenericRequest object. Such a Threshold could be expressed as `$GenericRequestObject->need('authLast')`.
 - *book_review*: A *book_review* service is only relevant for link-sources that describe books:
`$GenericRequestObject->need('objectType', 'eq', 'BOOK')`.
 - *genome*: A *genome* service is only relevant if the link-source contains genome sequence identifiers:
`$GenericRequestObject->need('genID')`
 - *abstract*: The Threshold for an *abstract* service might express that the link-source should describe a journal article and should at least have year, volume and issue information:
`$GenericRequestObject->need('objectType', 'eq', 'JOURNAL')`
`&& $GenericRequestObject->need('year') &&`
`$GenericRequestObject->need('volume') &&`
`$GenericRequestObject->need('issue')`.
- objectLookup Thresholds: The *abstract* service is clearly also subject to another type of boundary requiring that the Target resource into which the abstract service intends to link, actually abstracts the journal in which the item referred to by the GenericRequest object was published. This requirement explains the existence of the Objects table in [Figure 7](#) and of a

special objectLookup threshold. This type of Threshold will also be required to determine whether a journal in which the item referred to by link-source was published is part of a specific full-text repository in order to decide on the relevance of a *full_text* service into the repository.

Just as with the conceptual services, there is a global and a local component to these Thresholds. The global objectLookup Threshold for a *full_text* service linking into the [Springer](#) full-text collection, will learn whether a certain journal is a Springer e-journal or not. The local part of this Threshold will learn whether the journal is part of the actual digital library collection. In the same context, a global Threshold can express the fact that a journal is available in electronic form since 1996, while the local component might show that the local subscription only starts in 1998. In the same way, the BIOSIS-*abstract*-Medline service is subject to a global objectLookup Threshold. But there is also a global Threshold expressing that the publication year of the GenericRequest object has to be greater than 1965, reflecting the full coverage of Medline. Still, the local Threshold component for this service might be set to a more recent year if the local Medline implementation stores less data.

The SFX evaluation process

In order to present extended services for a given GenericRequest object the SFX evaluation process will determine the relevance of each of the conceptual services stored in the SFX-base using the content, or lack thereof, in the GenericRequest object. There are two phases to this evaluation process.

Phase 1: Selection of active services with the origin resource of the GenericRequest object as a Source

The interface between the redirection component and the service component delivers both link-source metadata and information on the origin of the link-source. The latter is stored in the rec\$dbId field of the GenericRequest object that is created by the service component. During the evaluation phase, the value of this field becomes the key for a lookup in the local component of the Source table of the SFX-base. The global common name of the resource is detected there, next to this key which refers to the local implementation of a resource. This global name is now connected via services of the Colli to various global names of Target resources, as already shown in [Table 7](#). Hence, the result of this lookup is a bundle of services that might be relevant for the current GenericRequest object, as judged upon by its origin. Inactivation of certain services during the localization of the SFX-base guarantees that the resulting bundle already reflects the local digital library situation.

In [Table 8](#) and [Table 9](#) the mechanism is shown in bold for a GenericRequest object representing an item originating from the Ghent implementation of BIOSIS. Its rec\$dbId value is ERL::BX. In this phase of the evaluation process, S::ERL::BX -- a prefix S is added as a means to refer to Source -- becomes the key for a look-up in the local component of the Source table. There, BIOSIS is detected as the global common name of the resource. Several services are leading out from BIOSIS into Target resources. For instance, *abstract* connects BIOSIS with Medline & *genome* connects BIOSIS with Genome Base. These are the services that remain as potentially relevant.

Phase 2: Filtering out selected active services by comparing the content of the GenericRequest object with the Thresholds

Phase 1 of the SFX evaluation process filters out services of the SFX-base that do not have BIOSIS as an origin. For each of the resulting services, the information in the GenericRequest object will be matched against the Thresholds -- global and local -- that are connected to these services. The *genome* service connecting BIOSIS and Genome Base will be filtered out if the GenericRequest object does not contain an entry for the genID parameter. The *abstract* service connecting BIOSIS and Medline will be filtered out if an objectLookup for the ISSN value in the GenericRequest object learns that the journal is not abstracted in Medline. Or it could be filtered out if the GenericRequest object does not contain a value for year, volume or issue.

Again, since some Thresholds express the local situation, and since these Thresholds can overrule the global ones, the result of this filtering process will reflect the situation of the local digital library collection. Those services remaining from Phase 1, for which at least one of the Threshold evaluations fails will be rejected as not being relevant. The ones that make it through the complete evaluation process will be presented to the user in the SFX-menu-screen as locally relevant extended services for the current GenericRequest object, hence for the link-source for which the whole process has been initiated by clicking the SFX-button

Resolving locally relevant extended services into URLs with TargetParsers

Consistent with the just-in-time linking philosophy of SFX, the bundle of relevant services that is obtained as a result of the SFX evaluation process described above, is not resolved into URLs at the moment of their presentation to the user that launched a request for extended services. Rather, for each menu-item in the SFX-menu-screen, the following elements are sent as parameters for a script that will be initiated when a user selects a menu-item:

- The identifier of the GenericRequest object;
- A name referring to the service and its Target, that is represented by the menu-item;
- For some services, overwriteable metadata elements from the GenericRequest object (see SFX-menu-screens in [Figure 8](#), [Figure 10](#), [Figure 12](#) and [Figure 14](#)).

When the user clicks a menu-item, the appropriate TargetParser script corresponding to the chosen service and Target is launched. These TargetParsers implement resource-specific link-to syntaxes. They take data from the GenericRequest object as input and compute the URL to which the user will be redirected.

Comments

The impact of the redesign of the SFX service component

The new design of the SFX service component, reflecting aspects of global and local relevance, has a considerable impact on the transportability and manageability of the SFX service component. Once an SFX-base containing conceptual services of global relevance with appropriate global Thresholds has been compiled, the localization of the set-up requires minimal efforts. As an illustration of this, it is interesting to look once again at the

BIOSIS-*abstract*-Medline example. This service was initially localized in Ghent by filling out the names of the local SourceParser for BIOSIS and the local TargetParser for Medline. Both parsers implement the desired connection with the [SilverPlatter](#) ERL platform that locally hosts the databases. LANL has a different implementation of BIOSIS, currently running on a [Geac](#) Advance system, while no local Medline is available. Therefore, LANL chose to use the PubMed implementation as a Target. When transporting the SFX-base -- that had been localized for Ghent first -- over to LANL, very limited editing of the SFX-base had to be done to activate the BIOSIS-*abstract*-Medline service in the new environment: the global service and its global Thresholds remained valid. The parser values for the Los Alamos version of BIOSIS and Medline were used to overwrite the Ghent values, as shown in [Table 9](#). The Threshold indicating that Medline is only available from 1985 onwards in the Ghent collection, was overwritten by a 1965 Threshold for LANL. In this case, the local and global Threshold are equal. The elegance of the PubMed [Entrez](#) link-to mechanism and the availability of the complete Medline collection caused Ghent to reconsider the Target -- hence TargetParser -- to be used in favor of the PubMed implementation. Upon this decision, again, very limited editing of the Ghent SFX-base had to be done.

In Ghent&LANL, most of the TargetParsers are implemented as Perl scripts. Towards the end of the experiment, advantage has been taken of the launch of a preliminary version of the S-Link-S Calculator ([Openly Inc. 1999](#)). This Calculator is designed to compute URLs based on input metadata and XML templates that describe link-to syntaxes in an S-Link-S compliant manner ([Hellman 1998](#)). As such, SFX TargetParser scripts that perform the computation of the URLs can eventually be replaced by templates describing the link-to syntax and that are used as input for the S-Link-S Calculator. The experiment ended with a hybrid solution, in which the SFX service component was adapted to dynamically choose between the two mechanism that became available to compute URLs: the TargetParsers and S-Link-S templates. TargetParsers can be shared between different digital library implementations, since many will link into the same resources or family of resources. Again, this reduces the overhead in running the solution. But the tie-in between the SFX and the S-Link-S work, will eventually further diminish the administration of the SFX solution as publishers start and contribute link-to templates and corresponding metadata to the S-Link-S framework. Sharing of TargetParsers will then be replaced by using S-Link-S templates from the framework that has already been set up by Eric Hellman to collect them.

Also, SourceParsers are easily transferable between different digital library implementations, requiring little or no enhancements to be made. For instance, OPAC systems worldwide support the [Z39.50](#) protocol and respond to requests with [MARC](#) formatted records. Making abstraction of the regrettable idiosyncrasies of Z39.50 and MARC implementations, SourceParsers for such OPAC systems can be reused with little editing, apart from the adaptation of Z39.50 parameters such as host, target, port etc... to the local situation. Also, Z39.50 can be used to fetch link-sources from all implementations of databases on a [SilverPlatter](#) ERL platform, and -- again -- only the Z39.50 parameters will be different. All implementations of MathSci on such ERL platforms can use the same parsing procedures, making the parsing part of the SourceParser even universal. This is also the case for the SourceParser used for the APS [PROLA](#) archive and the [Wiley](#) journals, since there is only one implementation of those, worldwide. This approach opens attractive possibilities of sharing SourceParser on a large scale, reducing the overhead in running the solution. Furthermore, it

allows information vendors to provide SourceParsers for their resources, keeping full control of the amount of information that they allow to be fetched.

The impact of a service component building on conceptual services

The consequences of the introduction -- in the Elektron experiment -- of a linking service built upon a database of conceptual services, have only now become fully evident, due to the complex and distributed nature of the environment in which Ghent&LANL was conducted. Actually, the more resources are added to the environment, the more the elegance and feasibility of the SFX service component become apparent. The introduction of a new SFX-aware resource in an environment requires very limited editing of the SFX-base, in order for all the existing conceptual services that had already been registered in the SFX-base to become immediately available for the new resource too. The dynamic manner in which the SFX system brings up a list of extended services for link-sources of a newly added resource can only be called remarkable. Even the designers of the system found it increasingly difficult to manually predict the outcome of a request for extended services, even when knowing the system and its underlying database in and out, and when studying the content of the link-source record or its GenericRequest object in detail. As an illustration of this, it is noteworthy to mention how SFX delivers a PubMed link for a citation in the Information Science journal JASIS, where [Wiley](#) themselves do not have such a link (see Screencam 3 of the examples). This is odd, since Wiley does insert static PubMed links for citations in their journals. Probably they only do so for those with a biomedical subject, since the cost/benefit balance of sending all their citations into the NCBI [PubRef](#) process would be negative. The dynamic and conceptual SFX approach does not require such precomputational processes and is able to recognize the validity of presenting a PubMed link for a citation in JASIS on the fly. Another remarkable and appealing example is where full-text links presented by SFX lead from a citation in a Wiley journal to an article in another Wiley journal. At the time of the experiment, Wiley did not offer such a linking service within its own collection, even if they fully control the data to implement it. These examples do not only illustrate the strength of the SFX solution, but more importantly, they are a strong indication of the problems of scale of static linking solutions.

The SFX redirection mechanism and namespace specific identifiers

Important efforts are under way to enable reference linking using [DOIs](#) ([Paskin 1999a](#)). Publishers will contribute metadata of their publications along with the corresponding DOIs to the [DOI](#) metadatabase. Other publishers can then match references in their own publications to the DOI metadatabase enabling them to insert the DOI of the work represented by the reference next to the reference. Currently, due to lack of support of the [handle](#) protocol in mainstream browsers, such a DOI -- say 10.1000/123456789 -- is being hyperlinked as <http://dx.doi.org/10.1000/123456789> and the DOI handle proxy will resolve this link into a single URL, being the one of the publication at the publisher's site. ([Paskin 1999b](#)). This hyperlink is a perfect example of a closed link that does not take into account the local context in which the link will be used ([Van de Sompel & Hochstenbach 1999a](#)). This closed link causes the Harvard problem to arise, since it does not take into account the possibility of storage of the same work at another, preferred location. Also, this mechanism does not allow other, locally relevant extended services to be provided for the reference at hand, since their provision requires the full metadata -- not only the identifier -- of the reference.

The local redirection approach presented by the SFX work does present a pragmatic way to open such a closed linking framework. DOIs can be carried in an SFX-URL pointing at the preferred SFX-server. For instance, for a citation in a [Wiley](#) InterScience article that has a static link to DOI 10.1000/123456789 the link can be dynamically rewritten if the existence of a service component is detected. In the case of SFX as a service component, it can become:

<http://isiserv.rug.ac.be/cgi-bin/sfx/bin/menu.cgi?>

[vendorID=Wiley&databaseId=WIS&nameSpace=DOI&objectDesc=URLEncode\(DOI=10.1000/123456789\)](http://isiserv.rug.ac.be/cgi-bin/sfx/bin/menu.cgi?vendorID=Wiley&databaseId=WIS&nameSpace=DOI&objectDesc=URLEncode(DOI=10.1000/123456789))

In essence, such a pragmatic mechanism can redirect the identifier to the redirection component of a selective resolution system, that can decide what to do with it, based on the local context. In the case of SFX, receipt of the above URL causes the launch of a SourceParser. As shown before, typically, this will be the SourceParser corresponding to the serviceDesc part of the URL. Still, upon detection of the "nameSpace" parameter, this default could be overwritten to become the namespace-specific SourceParser, in this example the one for the DOI namespace. This DOI SourceParser would do a so-called reverse look-up in the DOI metadatabase, using the DOI value as a key to fetch the corresponding metadata. Both the fetched metadata and the DOI can then be used in a process to determine the relevant extended services for the citation, including a link to the most appropriate full-text instance. As will be discussed in the next section, for other types of service components, redirection of the DOI without metadata-fetch could be sufficient.

The same mechanism can be used to open the links that are connected to citations carrying identifiers originating from other namespaces such as [PubMed](#) and [Astrophysics Data System](#). This can easily be seen by taking the following citation from a Wiley journal that has a static link to PubMed:

Rainer RO, Geisinger KR. Beyond sensitivity and specificity. Am J Clin Pathol 1995; 103: 541-2. [Medline](#)

The Medline hyperlink points at:

<http://www4.ncbi.nlm.nih.gov:80/htbin-post/Entrez/query?uid=95259660&form=6&db=m&D>

and uses the PubMed ID as a look-up key into the NCBI PubMed. This hyperlink can be pointed at a local resolution solution if it is dynamically rewritten as:

<http://isiserv.rug.ac.be/cgi-bin/sfx/bin/menu.cgi?>

[vendorID=Wiley&databaseId=WIS&nameSpace=Medline&objectDesc=URLEncode\(Medline:95259660\)](http://isiserv.rug.ac.be/cgi-bin/sfx/bin/menu.cgi?vendorID=Wiley&databaseId=WIS&nameSpace=Medline&objectDesc=URLEncode(Medline:95259660))

In a similar manner as in the [DOI](#) example, this URL can deliver the PubMed ID to a local redirection component. In the case of SFX, this would cause the Medline-namespace SourceParser to be launched, which would fetch the corresponding record from the Medline database. This SourceParser can actually fetch the metadata either from the PubMed implementation of Medline (using the HTTP protocol and the [Entrez](#) link-to syntax) or from a local implementation of Medline if one exists. Once that metadata has been turned into a GenericRequest object, the SFX evaluation process can deliver locally relevant

extended services. In this way, the initial service of [Wiley](#) is augmented with locally relevant services. Also, the quality of the metadata resulting after such a fetch will be higher than that in the original citation, as can be seen from the fact that no issue number is available in the citation of this example, as is common in the medical literature. The fetched record, however, will contain the missing issue number and much more valuable information.

This consideration illustrates that link-source metadata does not necessarily have to be fetched from its origin resource. In both examples, for a link-source originating in a Wiley journal, the link-source metadata is fetched from the authoritative resource for the namespace of which the link-source carries an identifier.

Identifiers, metadata and service components

It is interesting to further reflect on the nature of the service component and the requirements it imposes on the redirection mechanism of a selective resolution solution. To commence this consideration, service components that only aim at the delivery of the appropriate full-text instance for a given link-source are considered. Such a service component could operate solely on an underlying database of identifiers, meaning it could be a traditional linking service building on static links between documents. For such service components, it would be sufficient if the redirection mechanism would transfer identifiers only, without bothering about the associated metadata. Such a service component might be sufficient to address the Harvard problem. It might contain a repository of identifiers and locations of full-text for which a local or preferred warehouse other than the default one exists. However, such a repository of identifiers could quickly become very large and difficult to maintain. It can even be considered awkward to maintain such an institutional repository if no full-text is stored locally, but is only being accessed from preferred external aggregators. This consideration points at the desirability of a service component of a different, more abstract nature. Such a service component can build on the logic underlying the distribution of the collection rather than on individual identifiers of material in the collection. Under normal conditions, such logic might tell that all journals of a certain publisher are accessed at a certain warehouse, that certain ISSN numbers have to be accessed from another one, and that an ISSN number has to be accessed in one repository before a certain date and at another one after that date. This level of abstraction drastically reduces the amount of information to be maintained in the service component and hence makes it more scaleable. But it requires metadata of link-sources to act as operators, not only identifiers. Adding to this the fact that the identifiers required to make such a scenario work for link-sources originating from full-text repositories, abstracting & indexing databases, OPAC systems and e-print archives are not available and will most probably not become universally available any time soon, it must be concluded that service components will have to be able to operate on the basis of metadata in general with identifiers being a special instance of metadata. This imposes a requirement on the redirection mechanism to be able to deliver link-source metadata, not only identifiers.

Adding to the task of the service component the delivery of other extended services, it becomes hard to imagine that a scaleable solution in a highly distributed environment could build on an architecture with a static linking database. Several illustrations of this consideration have resulted from the Ghent&LANL experiment. A more abstract and dynamic service component is required, which will perform some rule-based decision making, that tends

towards the evaluation of the relevance of conceptual services as introduced in the SFX work. As will be clear from the SFX experiments, this type of service component requires the availability of link-source metadata in order to be able to function. Again, this imposes a requirement on the redirection mechanism to be able to deliver link-source metadata. This does not mean that identifiers are irrelevant to this type of solution. Quite to the contrary, since it has been shown that identifiers -- from whichever namespace -- are a welcome tool to enable the local redirection component to adequately deliver high-quality metadata.

The general conclusion of the above is that, realistically, identifiers will not be sufficient to address the problem of delivering extended services in a distributed digital library collection. Metadata is required for scaleable service components to be able to perform their tasks. Moving from left to right on the scale of service components ranging from traditional static linking systems to dynamic linking systems building on conceptual services, the data required for the service components to adequately do their jobs ranges from identifiers to full metadata.

Illustrations of project results

The concrete results of the project are illustrated by Lotus Screencam movies that show how a Ghent and a LANL user navigates in his institutional SFX-aware digital library collection. The Screencams are provided as stand-alone executables that can only be run on WinTel computers. Since the Screencams are large files, their size is mentioned. The Screencams do not contain audio. In addition to these Screencams, some examples are also given by means of screendumps.

Screencam 1: Ghent implementation of BIOSIS (Lotus Screencam executable for WinTel computer; no audio; size 52 Mb)

The user starts from the Ghent implementation of BIOSIS, and requests the services of the Ghent SFX solution. Services are shown linking from BIOSIS into the Ghent OPAC, into full-text collections, into the LANL implementation of the Science Citation Database, into PubMed and into Ulrich's Serials Directory. At a certain point the user links out to a paper in Cancer, a Wiley InterScience journal that is SFX-aware. Upon request, the user receives similar services from the Ghent SFX solution for citations in that paper. They link him to the LANL OPAC, to full-text at other publishers sites, to PubMed and to the LANL implementation of the Science Citation Database.

Screencam 2: LANL implementation of BIOSIS (Lotus Screencam executable for WinTel computer; no audio; size 46 Mb)

The user starts from the LANL implementation of BIOSIS and requests services from the LANL SFX solution. He uses links into the Ghent implementation of Current Contents, into several full-text collections and into the LANL implementation of the Science Citation Database. From the Citation Database, again, he requests SFX-services that lead him into more full-text collections as well as into PubMed. Next, the user returns to his result set in BIOSIS, where he requests services for other records. These lead him into the Journal Citation Reports and to more full-text. For one of the BIOSIS records, the *genome* service appears leading the user to the Genome database.

ScreenCam 3: Ghent implementation of Current Contents (Lotus ScreenCam executable for WinTel computer; no audio; size 36 Mb)

The user starts from the Ghent implementation of Current Contents, and requests the services of the Ghent SFX solution. The user links out to a paper in JASIS, an SFX-aware Wiley InterScience journal, and -- upon request -- is receiving SFX-services from the Ghent SFX server for citations found in the JASIS paper. Shown are links to PubMed, LISA and to the LANL implementation of the Science Citation Database. For another record from Current Contents, the user links into the Science Citation Database. There, he requests extended services for several records in the result set, which link him to full-text collections etc.

ScreenCam 4: Ghent Aleph 500 OPAC (Lotus ScreenCam executable for WinTel computer; no audio; size 13 Mb)

The user starts from the Ghent OPAC. He requests SFX-services for a record describing a book and links into Amazon.com. For a journal, he links to the full text collection and to the Journal Citation Reports.

ScreenCam 5: LANL Advance OPAC (Lotus ScreenCam executable for WinTel computer; no audio; size = 21 Mb)

The user starts from the LANL OPAC and requests services from the LANL SFX solution. For a book record, he links into Amazon.com. For the OPAC record of the journal Cancer he links to the full-text repository and browses towards a paper. In that paper, citations have SFX-buttons since the journal is SFX-aware. The user requests SFX-services for some citations, that bring him to PubMed and to the LANL implementation of the Science Citation Database. For records resulting from the latter service, the user again requests SFX-services that take him to the Journal Citation Reports, the LANL OPAC and to various full-text collections. One of those is -- again -- the Wiley InterScience collection.

ScreenCam 6: LANL Inspec (Lotus ScreenCam executable for WinTel computer; no audio; size = 29 Mb)

The user starts from the LANL Inspec and requests services from the LANL SFX solution. Services bring him to full-text collections, the LANL OPAC, the Journal Citation Reports. At a certain point the user goes out to the PROLA archive and finds SFX-aware citations for which he requests extended services. Due to the current implementation of the SFX-URL for PROLA little metadata makes it into a GenericRequest object and as such little services result. Back in the Inspec, the user links to the LANL implementation of the Science Citation Database, from which he further links to the LANL OPAC and to full-text collections.

Screenam 7: the arXiv, consulted by a LANL user (Lotus Screenam executable for WinTel computer; no audio; size = 19 Mb)

The user searches the Topic interface for the arXiv e-print repository and requests SFX-services for search results. These link him into the Inspec database, searching for references by the authors of the e-prints.

Screeendump example 1:

Figure 8 shows how the link-source record from Figure 3, originating from the Ghent implementation of BIOSIS yields a variety of extended services, amongst other *full_text*, *holding*, *abstract*, *author*, *cited_author* and *cited_reference* that are presented in the SFX-menu-screen. From this menu-screen, the user chooses full-text link to Wiley, which leads him into an article of the journal Cancer that is SFX-aware (Figure 9). Figure 10 shows how the same user has requested extended services for the third citations in that Wiley article and has received a Ghent SFX-menu-screen. For this citation similar services are available, all linking dynamically from the external Wiley resource back into the Ghent digital library collection. As can be seen link to the on-line version of the New England Journal of Medicine has become available. The user chooses use the *cited_reference* service that is also available for this citation. This leads him into the SFX-aware implementation of the Science Citation Database (Figure 11) from where he can request extended service again.

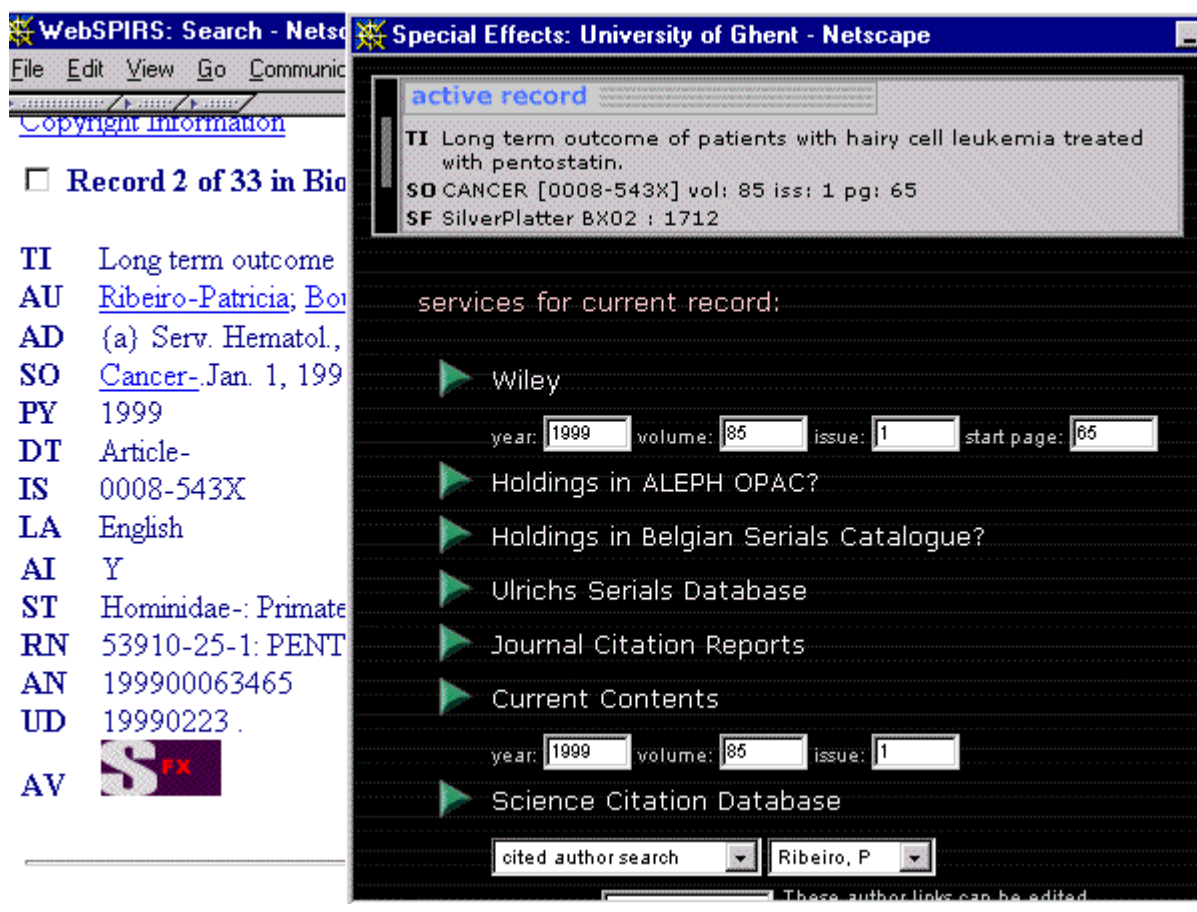


Figure 8: Ghent SFX-menu-screen for link-source from Ghent BIOSIS (record from Figure 3)





Figure 9: Ghent user follows Wiley *full_text* service from the SFX-menu of [Figure 8](#)

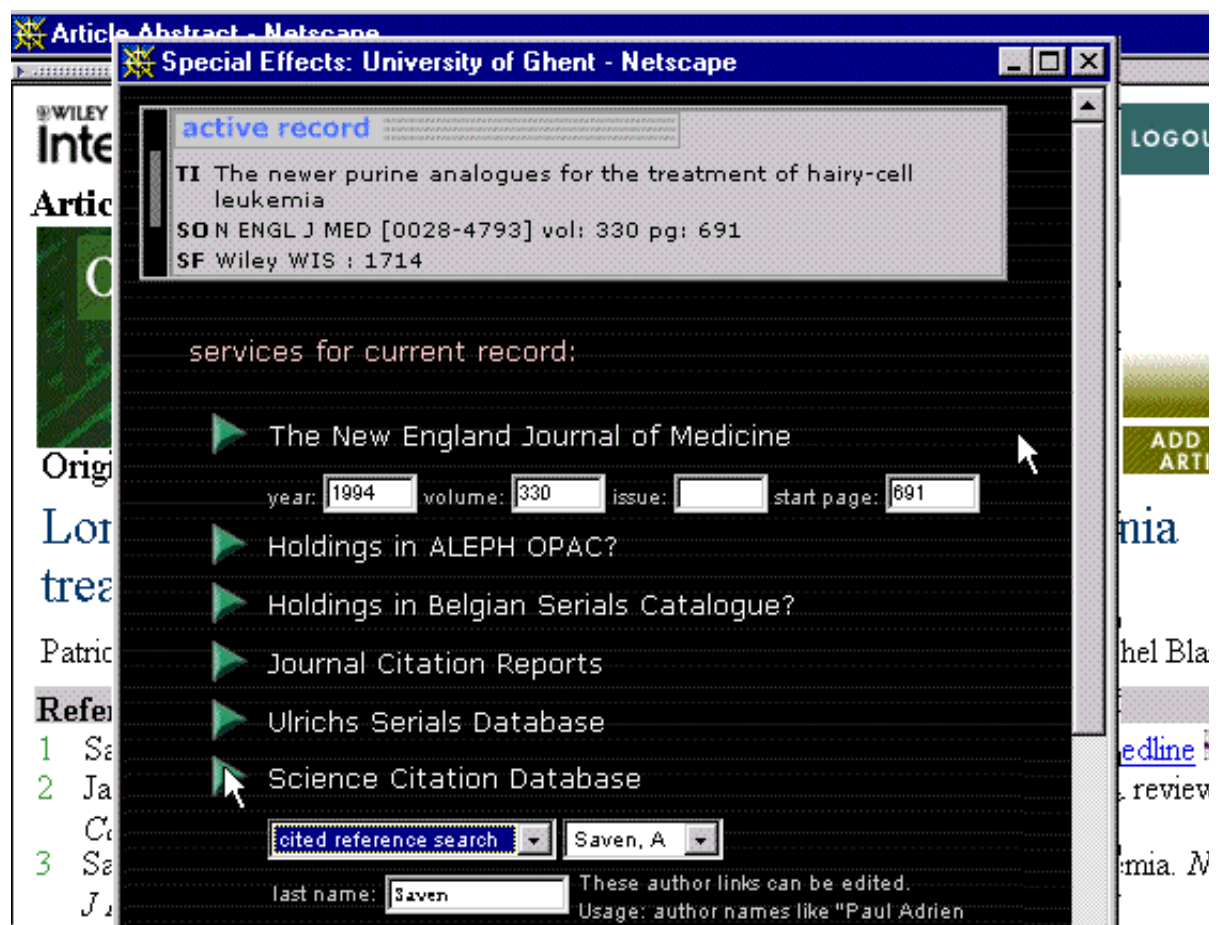


Figure 10: Ghent SFX-menu-screen for link-source from Wiley InterScience (third citation from [F](#)

Article Abstract - Netscape

Special Effects: SciSearch® at LANL Retrieval Results - Netscape

Los Alamos Research Library
NATIONAL LABORATORY

Home Gen Search Cited Search Alerts
Up Mark All List Marks Cited S

SciSearch® at LANL Cited Reference R

30 out of 3114509 records matched the query below. The 30 most relevant the work below are shown:

"SAVEN_A_1994_N330_691"
(Years searched = 1996 - 1999)


Marks	Title, Author, Source
<input type="checkbox"/>	 Treatment of a de novo fludarabine resistant-bryostatin 1 followed by fludarabine Mohammad RM, Limvarapuss C, Hamdy N, D AIKatib AM INTERNATIONAL JOURNAL OF ONCOLOGY

Figure 11: Ghent user follows the *cited_reference* service from the SFX-menu of [Figure 10](#)

Screeendump example 2:

In [Figure 12](#), the record from [Figure 4](#) originating from the LANL implementation of BIOSIS is used as a link-source. The resulting SFX-menu-screen looks a little different, as an illustration of the fact that another SFX system is being consulted to provide extended services. The LANL localization can also be derived from the OPAC link, that now leads into the Los Alamos Advance catalogue. Another service appears in this screen too: it provides a look-up of sequence information for genome identifiers that were found in the link-source metadata. This service leads the user to the NCBI Genome database using the Entrez link-to syntax ([Figure](#)

LAN Special Effects: Los Alamos National Laboratory - Netscape

active record

TI The identification of cDNAs that affect the mitosis-to-interphase transition in *Schizosaccharomyces pombe*, including sbp1, which encodes a sp11p-GTP-binding protein.
 SO Genetics; [0016-6731]
 yr: 1998 vol: 148 iss: 2 pg: 645

services for current record:

- Highwire
- Holdings in Advance OPAC?
- PubMed Genome Database
genome id: D42063
- Journal Citation Reports
- Current Contents
- Science Citation Database

year: 1998 volume: 148 issue: 2 start page: 645

cited author search HE, XIANGWEI

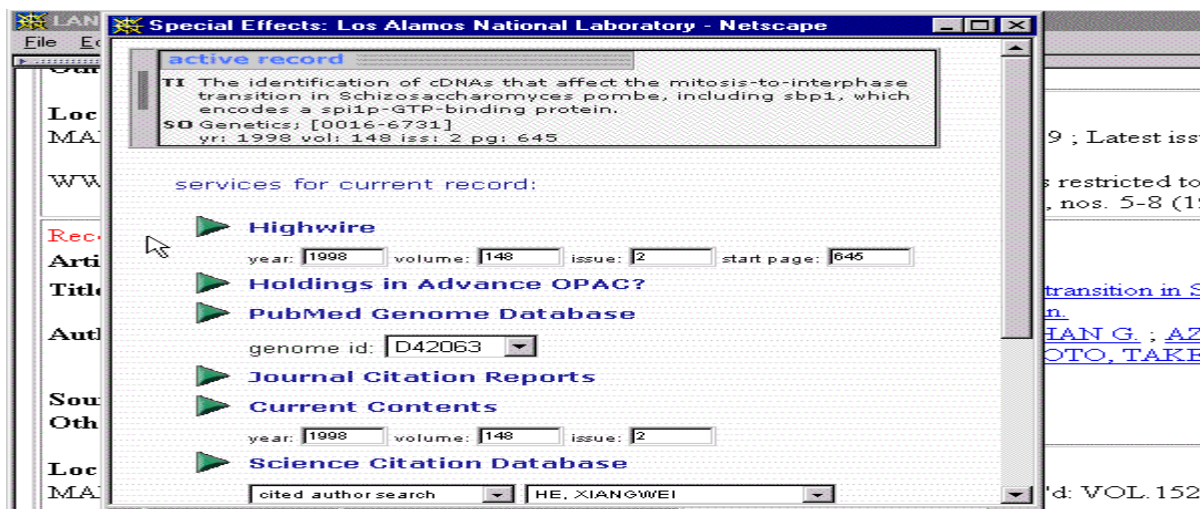


Figure 12: LANL SFX-menu-screen for link-source from LANL BIOSIS (record from [Figure 11](#))

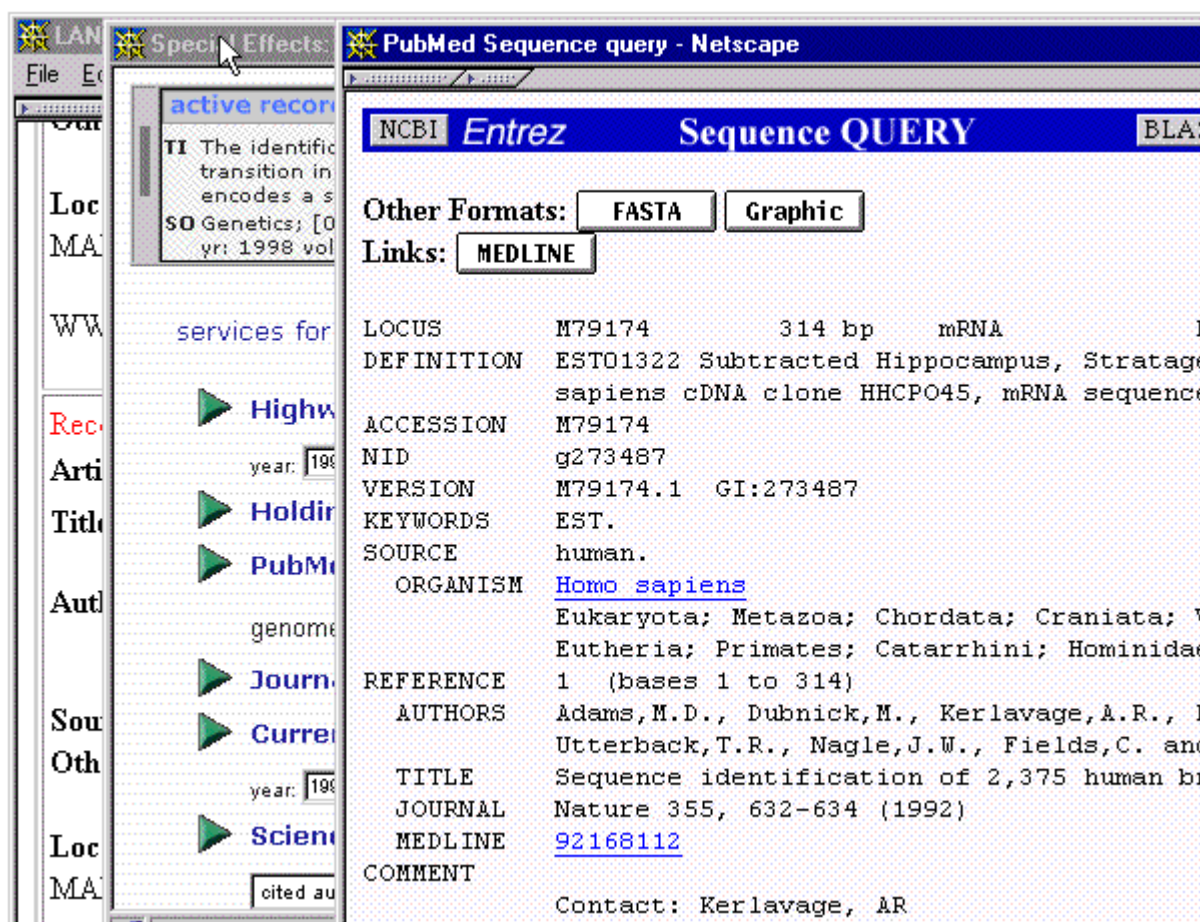


Figure 13: LANL user follows the *genome* service from the SFX-menu of [Figure 12](#)

(after selection of identifier M79174 from the pop-down)

Screeendump example 3:

[Figure 14](#) shows the LANL SFX-screen for the first record from the Topic implementation of the [arXiv](#) e-repository shown in [Figure 6](#). Here, an *author* service is available, that provides the option to look-up rec

the Inspec database, that abstract publications authored by the e-print authors as a means to support the evaluation of the reliability of the non-peer-reviewed e-print. The author that is being looked-up has 160 references in the Inspec database (Figure 15).

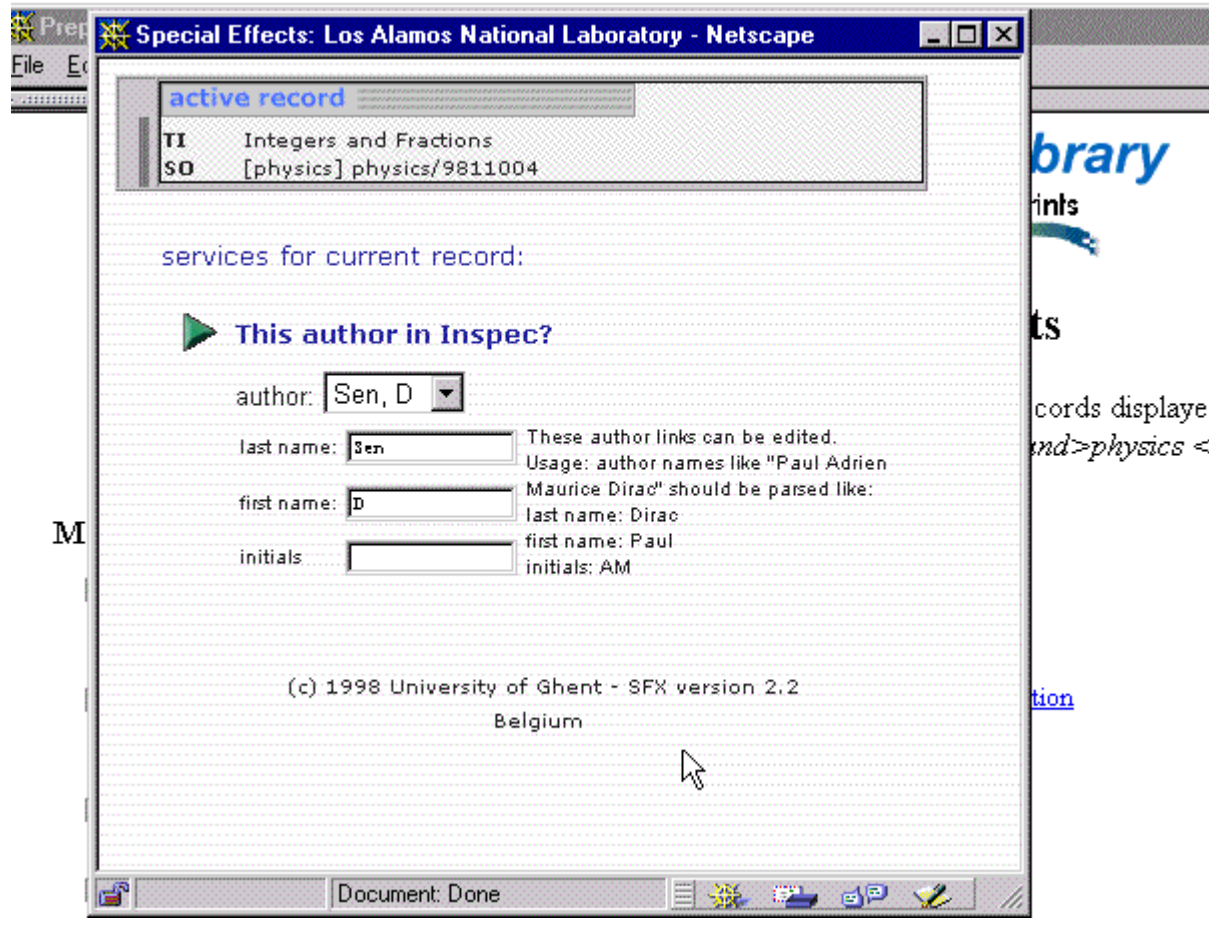


Figure 14: LANL SFX-menu-screen for link-source from the arXiv (first record from Figure 15)

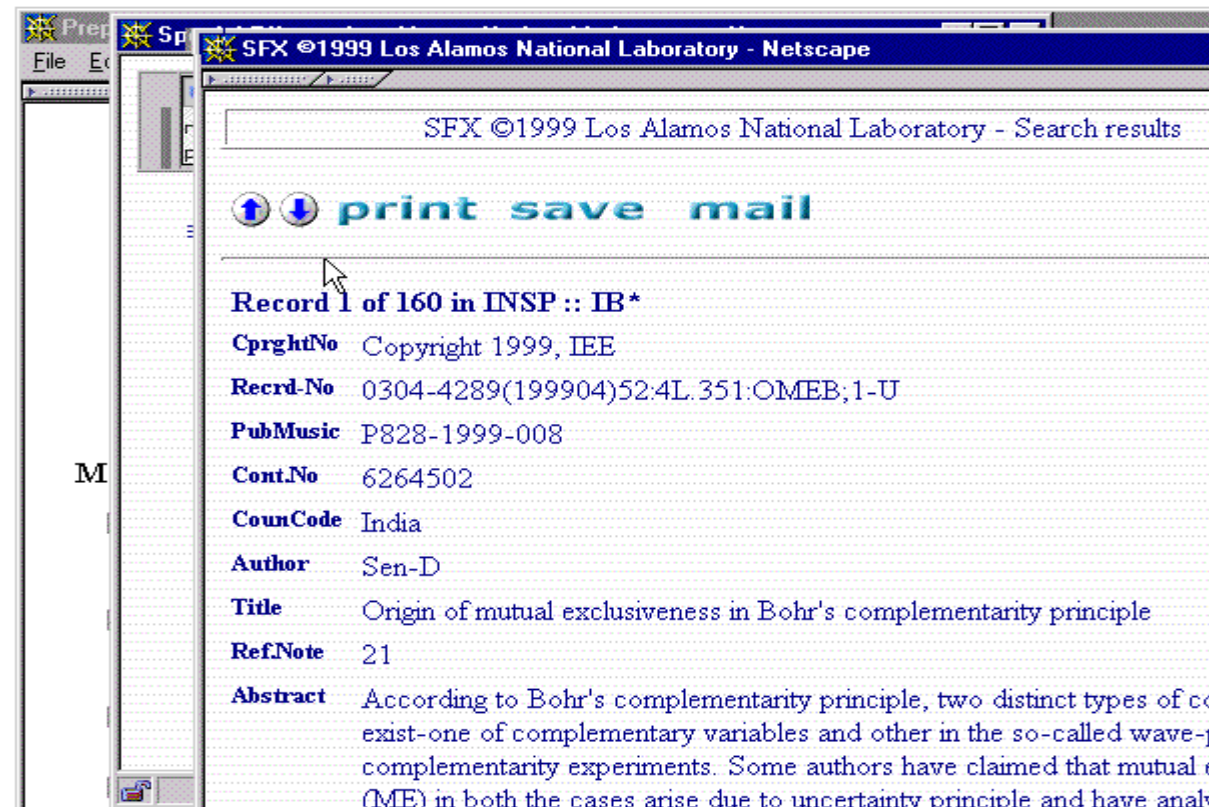




Figure 15: LANL user follows the Inspec *author* service from the SFX-menu of [Figure 14](#)

Conclusion

The "SFX@Ghent & SFX@LANL" experiment has led to important generalizations of two components that had been introduced in the Elektron experiment and that are essential for systems supportive of selective resolution: the redirection mechanism and the service component. Although both have been discussed in relation to one another, it has also been shown that they can be separate components that exchange information in a unique metadata interchange format. For the experiment, this format was internally defined and inspired on the structure of the GenericRequest object, since -- by lack of non-SFX local redirection components -- interoperability at this level was not an issue. If more local redirection solutions and more local service solutions become available, standardization of this interchange format will become important. Also, a standardization of a local redirection URL -- like the SFX-URL -- and of vendorId, databaseId and nameSpace values is required. Such a standardization will also be valuable for other ongoing work in the digital library environment.

In essence, the SFX redirection mechanism can be combined with a service component of a very different architecture, even one that builds on a static linking database of identifiers. The current implementation of the SFX local redirection mechanism builds on the CookiePusher mechanism, a consistent SFX-URL and SourceParsers. Each of these building blocks can be replaced by more robust alternatives, as long as they perform the same function. The investment required to make systems SFX-aware using the CookiePusher and the SFX-URL has been minimized. Still, it will be easier to implement SFX-awareness in resources that deliver information in a dynamic rather than in a static manner. It has been shown that the SFX local redirection mechanism can be used to redirect namespace-specific identifiers to a local service component. This leads to the capability of the SFX-URL to open closed linking frameworks, which is seen as a powerful illustration of the feasibility of the approach.

The SFX service component can also operate with a different redirection method, as long as that supports delivery of link-source metadata and its origin to the service component. The Ghent&LANL information environment, with its many resources and different technologies running those resources, has led to a design in which the SFX linking service has become a totally neutral module in the digital library that can potentially interoperate with every other system in the environment. Its redesign, reflecting the notions of global and local relevance of services, has led to an important reduction in the overhead of running the solution. In addition to that, the possibility to share SourceParsers, TargetParsers and S-Link-S templates further diminishes the administrative overhead.

As far as can be verified, Ghent&LANL has been the first experiment in which bi-directional context-sensitive linking between distributed resources under

control of different authorities has been demonstrated in the scholarly communication environment. As can be seen from the examples, the net result of making systems SFX-aware and delivering extended services for link-sources originating from those systems via the SFX-menu-screen, is a fully hypertextual scholarly information environment in which jumping between related distributed resources becomes possible, hopefully in the way Gardner had imagined it ([Gardner 1990](#)). As with the most renowned hypertext system -- the World Wide Web -- the ease with which this navigation occurs, can lead to getting lost in the information space. At this point, this feature is seen as a compliment to the solution, since no comparable navigational capability has been demonstrated before.

Future directions

The SFX solution will be brought into production in both Ghent and Los Alamos around January 2000. This will hopefully lead to user feedback that has so far been limited and has definitely not been investigated in a systematic manner. It will also lead to more involvement of librarians, which can only result in improvement of user aspects in the system. The current SFX solution is also ready to be beta-tested in digital libraries that are run by staff with reasonable technical skills. However, in order to be able to successfully coordinate such beta-testing, some central resources would be required. SFX is also being tested as a tool to integrate the subversive scholarly communication mechanisms ([Okerson & O'Donnell 1995](#)) with the established ones. This is, amongst others, done in the [UPS](#) protoproto work, that aims at building a multidisciplinary digital library service for major e-print initiatives ([Van de Sompel 1999](#)). In that project, the SFX work is combined with the Smart Objects Dumb Archives research ([Maly, Nelson & Zubair 1999](#)). Work is also underway to investigate the feasibility of making the [SLAC/SPIRES HEP system](#) SFX-aware. The JISC/NSF project ([Harnad 1999](#)) that looks into inter- and intralinking citations in the [arXiv](#) e-prints also intends to experiment with SFX as a building block. Discussions with scholarly publishers about interoperating with SFX are also under way. It is believed that a wider distribution of the solution is feasible, but will probably require commercial support or alternative financial investments.

The current design of the SFX-base in Ghent&LANL, implementing the notions of global and local relevance, suggests the possibility of an architectural redesign that builds on a division of the set-up in a central component, that describes the global level, and a local component for the localization. Such a redesign would bring the SFX solution into category 2 of the categorization of systems supportive of selective resolution introduced in [Table 1](#). This could considerably lower the redundancy of information in various implementations of SFX-bases and hence could make their local administration easier.

Research is on its way to build a recommendation system based on user activities in an SFX-aware environment. The fact that relevant user traversals in both external and internal information resources can be tracked leads to a wealth of log data that can be exploited by the recommendation system. Further research is also required in the area of the insertion of SFX-buttons for citations in closed information containers like PDF files and Word documents.

References

Caplan, Priscilla. 1999a. A model for reference linking. Report of the working group of the reference linking workshop; May 1999.

[\[http://www.lib.uchicago.edu/Annex/pcaplan/reflink.html\]](http://www.lib.uchicago.edu/Annex/pcaplan/reflink.html).

Caplan, Priscilla. 1999b. Report of the second workshop on linkage from citations to journal literature; June 9th 1999, Boston.

[\[http://www.niso.org/linkrept.html\]](http://www.niso.org/linkrept.html).

Caplan, Priscilla and William Y. Arms. 1999. Reference linking for journal articles. *D-Lib Magazine* 5, no. 7/8.

[\[http://www.dlib.org/dlib/july99/caplan/07caplan.html\]](http://www.dlib.org/dlib/july99/caplan/07caplan.html).

Gardner, William. 1990. The electronic archive: scientific publishing for the 1990s. *Psychological Science* 1, no. 6.

Halstead, Amy. 1999. PROLA: More Than Just a Pretty Acronym. *APS News* 8, no. 8. [\[http://www.aps.org/apsnews/0899/089914.html\]](http://www.aps.org/apsnews/0899/089914.html).

Harnad, Stevan. 1999. Integrating and navigating eprint archives through citation-linking (NSF / JISC-eLib Collaborative Project).

[\[http://www.princeton.edu/~harnad/citation.html\]](http://www.princeton.edu/~harnad/citation.html).

Hellman, Eric. 1998. Scholarly Link Specification Framework (SLinkS).

[\[http://www.openly.com/SLinkS/\]](http://www.openly.com/SLinkS/).

Maly, Kurt, Michael Nelson, and Mohammad Zubair. 1999. Smart Objects, Dumb Archives: A User-Centric, Layered Digital Library Framework. *D-Lib Magazine* 5, no. 3. [\[http://www.dlib.org/dlib/march99/maly/03maly.html\]](http://www.dlib.org/dlib/march99/maly/03maly.html).

Needleman, Mark. 1999. Meeting report of the NISO linking workshop; February 11th 1999, Washington DC. [\[http://www.niso.org/linkrpt.html\]](http://www.niso.org/linkrpt.html).

Okerson, Ann and James O'Donnell. 1995. Scholarly Journals at the Crossroads; A Subversive Proposal for Electronic Publishing. Washington, DC: Association of Research Libraries. [\[http://www.arl.org/scomm/subversive/toc.html\]](http://www.arl.org/scomm/subversive/toc.html).

Openly Inc. 1999. S-Link-S Calculator. June 1999.

[\[http://www.openly.com/SLinkS/Calculator/\]](http://www.openly.com/SLinkS/Calculator/).

Paskin, Norman. 1999a. DOIs used for reference linking. Washington & Geneva. [\[http://dx.doi.org/10.1000/143\]](http://dx.doi.org/10.1000/143).

Paskin, Norman. 1999b. DOI: Current Status and Outlook. *D-Lib Magazine* 5, no. 5. [\[http://www.dlib.org/dlib/may99/05paskin.html\]](http://www.dlib.org/dlib/may99/05paskin.html).

Shishir, Gunavaram. 1996. CGI Programming on the World Wide Web. Sebastopol, CA.: O'Reilly and Associates, Inc.

Spilka, Susan. 1999. Wiley InterScience Update. June 1999.

[\[http://www.wiley.com/about/corpnews/wisupdate.html\]](http://www.wiley.com/about/corpnews/wisupdate.html).

Van de Sompel, Herbert. 1999. the Universal Preprint Service initiative. July 1999. [\[http://vole.lanl.gov/ups/\]](http://vole.lanl.gov/ups/).

Van de Sompel, Herbert and Patrick Hochstenbach. 1999a. Reference linking in a hybrid library environment. Part 1: Frameworks for linking. *D-Lib Magazine* 5, no. 4.

[\[http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt1.html\]](http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt1.html).

Van de Sompel, Herbert and Patrick Hochstenbach. 1999b. Reference linking in a hybrid library environment. Part 2: SFX, a generic linking solution. *D-Lib Magazine* 5, no. 4.

[http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt2.html].

Acknowledgments

The authors wish to thank the following parties for their active cooperation in the experiment:

- Lieve Rottiers at the [Ghent Library Automation team](#) for art work
- Miriam Blake, Johan Bollen, Doug Chafe, Mariella Di Giacomo, Frances Knudson, Dan Mahoney and Mark Martinez at the [LANL Library Without Walls](#) team
- Abe Lederman at the LANL CIC-15 group
- Mark Doyle at the American Physical Society / [PROLA](#) archive
- Andy Stevens, Andy Townsend and Craig Van Dyck at [Wiley Interscience](#)
- Denis Lynch, Jenny Walker and Andrew Wilkins at [SilverPlatter](#)
- Oren Beit-Arie and Yohanan Spruch at [ExLibris](#)
- Eric Hellman at [Openly](#)

Herbert Van de Sompel wishes to thank:

- the [Belgian Science Foundation](#) for a special PhD grant
- Rebecca Graham, Deanna Marcum and Don Waters at the [Council on Library & Information Resources](#) and the [Digital Library Federation](#) for a travel grant and active support
- Donna Berg and Rick Luce at the [LANL Research Library](#)
- Paul Ginsparg at the LANL e-print [arXiv](#)
- William Y. Arms at [D-Lib Magazine](#) & Cornell University
- Jennifer De Beer for proofreading

Copyright © 1999 Herbert Van de Sompel and Patrick Hochstenbach

(At the request of the authors, reference to "Orion ScienceServer" has been changed to "ScienceServer" and a link to Science Server has been added; in the acknowledgement to Mark Doyle, "American Institute of Physics" has been corrected to read "American Physical Society"; and in the abstract, "digital library collections of the Los Alamos National Library Research Library" has been corrected to read "digital library collections of the Research Library of the Los Alamos National Laboratory". Corrections made 10/19/99, 1:34 pm.)

DOI: 10.1045/october99-van_de_sompel