17. LinkOut: Linking to External Resources from Entrez Databases

Kathy Kwan

Created: October 9, 2002
Updated: April 14, 2003

Summary

The power of linking is one of the most important developments that the World Wide Web offers to the scientific and research community. By providing a convenient and effective means for sharing ideas, linking helps scientists and scholars promote their research goals.

LinkOut [http://www.ncbi.nlm.nih.gov/entrez/linkout/] is a powerful linking feature of the Entrez search and retrieval system (Chapter 15). It is designed to provide Entrez users with links from database records to a wide variety of relevant online resources, including full-text publications, biological databases, consumer health information, and research tools. (See Sample Links [http://www.ncbi.nlm.nih.gov/entrez/linkout/doc/linksexample.html] for examples of LinkOut resources.) The goal of LinkOut is to facilitate access to relevant online resources beyond the Entrez system to extend, clarify, or supplement information found in the Entrez databases. By branching out to relevant resources on the Web, LinkOut expands on the theme of Entrez as an information discovery system.

How Is LinkOut Represented in Entrez?

Any Entrez database record, e.g., a nucleotide sequence, a taxonomic record, a protein structure, or a PubMed abstract, can be linked to Web resources external to NCBI via LinkOut. The LinkOut homepage [http://ncbi.nlm.nih.gov/entrez/linkout/] contains up-to-date documentation about Link-Out. The page that lists LinkOut resources associated with a given record can be accessed in a variety of ways (Figures 1–3). In the case of PubMed, the full-text article and other resources related to the abstract being viewed may be accessed directly by icon buttons above the abstract (Figure 2). The LinkOut display can be customized by using the LinkOut preferences in the Cubby.

1: Curr Opin Cell Biol 2001 Aug;13(4):485-92

Related Articles   NEW   Links

ELSEVIER SCIENCE
FULL-TEXT ARTICLE

Cited in PMC

Books

LinkOut

Help

The role of phosphoinositides in membrane transport.

Simonsen A, Wurmser AE, Emr SD, Stenmark H.

Department of Biochemistry, Institute for Cancer Research, the Norwegian Radium Hospital, Montebello, N-0310, Oslo, Norway.

1: R14038. yf62fl1.r1 Soares...[gi:767114]

NEW   Links

IDENTIFIERS

Taxonomy

LinkOut

Help

```
dbEST Id:        184167
EST name:        yf62f11.r1
```

| Display | Summary ▼ | Sort ▼ | Save | Text | Clip Add | Order |

Show: 20   Summary | of 647191   Page 1 of 32360   Select page: **1** 2 3 4 5 6 7 8 9 10 »

Brief
Abstract
Citation
ASN.1
MEDLINE
XML
UI List

1: Schre...nzelmann K.   Related Articles, NEW   Links
Elect...   mouse trachea: no evidence for a contribution of luminal k+ conductance.
J Me...   2):143-51.
PMI...   n process]

Related Articles
Domain Links

2: Gallo...O'Connell MA, Keegan LP.   Related Articles, NEW   Links
Micr...   im mRNAs identified by DHPLC analysis.
Nucl...   15;30(18):3945-33.
PMID:12235378 [PubMed - in process]

Figure 1: The LinkOut display can be accessed by selecting LinkOut from a PubMed record (*top panel*), from other Entrez databases (*middle panel*), or from the Display list (*lower panel*).

1: Curr Opin Cell Biol 2001 Aug;13(4):485-92

Related Articles,   NEW   Links

ELSEVIER SCIENCE
FULL-TEXT ARTICLE

The role of phosphoinositides in membrane transport.

Simonsen A, Wurmser AE, Emr SD, Stenmark H.

Department of Biochemistry, Institute for Cancer Research, the Norwegian Radium Hospital, Montebello, N-0310, Oslo, Norway.

Figure 2: From PubMed, the links to the full text of research articles are also managed by LinkOut and can be accessed through an icon from PubMed Abstracts, highlighted here in *purple*, as well as from the associated list of LinkOut resources in Figure 3.

- Links to full-text and resource information are supplied by LinkOut providers.
- Links with an asterisk indicate the LinkOut provider requires a subscription, membership, or fee for access.

☐ **1:** Hotta SS. Cardiac rehabilitation progra...[PMID:1667265]   Related Articles,   NEW   Links

- LITERATURE:
    - Libraries

- MEDICAL:
    - Consumer health:
        - MEDLINEplus Health Information
          Heart Transplantation   Heart Valve Diseases   Rehabilitation
    - Disease organizations:
        - NLM Health Services Technology/Assessment Text
          Cardiac Rehabilitation Programs
    - Treatment guidelines:
        - NLM Health Services Technology/Assessment Text
          Cardiac Rehabilitation Programs

Figure 3: **Links to external resources are listed in the LinkOut Display of an Entrez record.**

How Does LinkOut Work?

Design Overview

The URLs to LinkOut resources are all provided by the person or organization that owns or created the resource. Links can be provided in any URL syntax, and providers of links may choose as much or as little access to their resource as they wish. Providers use one format to submit links to all Entrez databases.

LinkOut is in itself an Entrez database that holds all the linking information to external resources. The separation of the data records (e.g., PubMed abstracts) from the external linking information (e.g., URLs to journal articles on a publisher's Web site) enables both the external link providers and NCBI to manage linking in a flexible manner. This means that if links to external resources change, such as in the case of a Web site redesign, this will not affect the Entrez database records, and linking information can be updated as frequently as necessary.

The LinkOut database contains information on the relationship between a link and all of the applicable unique Entrez ID numbers (UIDs). By taking advantage of the interconnectivity among Entrez nodes, the linking information is presented seamlessly and efficiently.

LinkOut DTD and XML Files

LinkOut information is submitted in XML, defined by the LinkOut Document Type Definition (DTD [http://www.ncbi.nlm.nih.gov/entrez/linkout/doc/LinkOut.dtd]).

Linking information is supplied in two elements: the Provider element, which specifies information about a link provider; and the LinkSet element, which describes information about the link. Each element should be submitted to NCBI in a separate file. Identity files contain the Provider element, and Resource files contain the LinkSet element.

The Identity file is always called providerinfo.xml. It describes the identity of a provider, including an ID (ProviderId) and an abbreviated name (NameAbbr) assigned by NCBI, the provider's name, and other general information about the provider. There should be only one providerinfo.xml file for each provider (see Box 1 for an example of an Identity file).

The Resource file, which contains the LinkSet information, specifies a set of Entrez records with a valid Entrez query, a specific rule to build the link to an external resource, and description of the resource using the SubjectType, Attribute, and UrlName [http://www.ncbi.nlm.nih.gov/entrez/linkout/doc/subjecttypes.html] fields. There is no standard for naming the LinkSet files, except that they must use the .xml extension. There may be any number of LinkSet files associated with a ProviderId. (See Box 2 for an example of a Resource file.)

Terms used in SubjectType and Attribute elements are controlled to describe LinkOut resources in a systematic manner. This is because resources are presented to users by SubjectType on the LinkOut display page (and within the Cubby system), making it easier to browse and access available resources. Attributes can be used to describe the nature of a LinkOut resource (i.e., whether the resource requires a subscription or registration to access the content). A short text string may be used in the UrlName element to describe a resource. UrlName is typically used when the allowed SubjectType and Attribute terms cannot describe the resource adequately or when multiple links are available from one provider for a single Entrez record.

XML File Processing and Indexing

All links from Entrez are generated on a daily basis so that new or modified Entrez records will have accurate LinkOut resources connected to them. Once a day, all LinkOut files are parsed according to the LinkOut DTD, and the LinkOut database is rebuilt, relating the Entrez UIDs with the link information specified in the LinkSet XML files.

A LinkOut record consists of a link and the associated information, including its URL and all descriptive terms (SubjectType, Attribute, and UrlName) pertaining to the link. The Entrez UIDs applicable to the link are indexed to associate this information to the corresponding Entrez databases. As explained in Chapter 15, LinkOut information is interconnected with all related Entrez records.

LinkOut Filters

To facilitate search and retrieval of LinkOut resources, there are a number of filters in the LinkOut-enabled Entrez databases. These filters, although not part of the LinkOut database, use the result generated in the LinkOut indexing process.

The filters are all prefixed with **lo**. Filters are available for all allowable SubjectType and Attribute terms and the NameAbbr of a provider. Some examples include:

- **loprov** LinkOut Provider

- **loattr** LinkOut Attribute

- **losubj** LinkOut SubjectType

- **loall** all LinkOut resources in an Entrez database

To use these filters to retrieve a set of Entrez records with LinkOut resources, the filter term can be entered as a search. For example, in PubMed, searching

```
"loattrfull text online"[Filter]
```

will retrieve all records with LinkOut resources that have an attribute "full-text online". The **Pre-view/Index** section in PubMed can also be used to select LinkOut filters by first selecting **Filter** and then typing in "lo" and selecting **Index** to browse through all of the filters related to LinkOut.

Guides for LinkOut Providers

LinkOut resources should be directly relevant to specific subjects of the Entrez records to which they will be linked, thus providing further research resources for Entrez users. The information and its delivery system should be of high quality and must not, through typographic or factual errors, omissions, or other flaws or inconsistencies, mislead, hinder, or frustrate the research efforts of Entrez users. The resources should be easy to use and navigate. Resources from professional societies, government agencies, educational institutions, or individuals and organizations that have received grants from major funding organizations are preferred.

Participation in LinkOut is voluntary. Providers need to submit two types of files to describe the LinkOut resources, Identity files and Resource files (see Boxes 1 and 2). These files include the necessary information for the Entrez system to construct an appropriate URL to access specific resources.

A list of Frequently Asked Questions [http://www.ncbi.nlm.nih.gov/entrez/linkout/doc/linkoutfaq.html] is available to address questions that potential LinkOut providers may have. Current lists of LinkOut providers [http://www.ncbi.nlm.nih.gov/entrez/journals/active_providers.html] can also be browsed.

Submission Procedures

Step 1. Initial Contact. A prospective provider can write to linkout@ncbi.nlm.nih.gov, indicating interest in creating links from Entrez records to the providers' Web-accessible online resources. Please include the name, email address, and phone number of an individual who will act as a designated contact. The email should also include a LinkOut Identity file (providerinfo.xml) based on the specifications described above.

Step 2. File Evaluation. NCBI staff will evaluate the resources before a ProviderId and NameAbbr are assigned. NCBI will also provide assistance with setting up an appropriate Resource file to describe the LinkOut resources.

Step 3. File Submission. An FTP account will be assigned to a provider for submission. Files must have been validated by the LinkOut Validation [http://www.ncbi.nlm.nih.gov/entrez/ linkout/validate.html] utility before uploading. Providers may transfer new versions of current files or add new Resource files at their own discretion. Providers are responsible for keeping their files current and valid. Links in Entrez databases are regenerated each day based on the files in each provider's directory; therefore, providers must delete obsolete files from their holdings directory.

Step 4. Representation in Entrez. Once a provider's LinkOut files are processed, the resources described in the file will be available in the LinkOut display of a relevant Entrez record as described in the above section, *How Is LinkOut Represented in Entrez?*. In PubMed, publishers of the abstract can choose to display a "button" on the Abstract and Citation displays of the PubMed record by adding the parameter "holding=NameAbbr" to the basic PubMed URL. For example, to activate the icon of WebDatabase Co, the URL would be provided as:

```
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=WebDB
```

Furthermore, multiple NameAbbr parameters may be used in a URL to activate more than one icon. For example, to display icons for both WebDB and MyDB, the following URL should be provided:

```
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=WebDB,MyDB
```

A provider's icon can be activated if the provider is selected from the LinkOut Preferences in Cubby.

All access restrictions will still apply. For example, if access to a database is limited by the user IP address, access will only be allowed via computers within an approved IP range; if access is password protected, the password must still be entered.

Detailed Guides

Interested parties can consult the following guides for more details:

- LinkOut and Non-Bibliographic Resources [http://www.ncbi.nlm.nih.gov/entrez/linkout/doc/ nonbiblinkout.html] – written for providers of general LinkOut resources in all Entrez databases.

- LinkOut and Publisher Holdings [http://www.ncbi.nlm.nih.gov/entrez/linkout/doc/publinkout. html] – written for publishers and others that provide full-text links to PubMed records.

- LinkOut and Library Holdings [http://www.ncbi.nlm.nih.gov/entrez/linkout/doc/liblinkout. html] – written for libraries to indicate information about their electronic full-text subscription and print holdings.

Auxiliary Tools

A number of tools are available to facilitate participation in LinkOut:

- Library LinkOut Files Submission Utility [http://www.ncbi.nlm.nih.gov/entrez/linkout/doc/lbsub-i.html] utility developed for Libraries to generate and manage their LinkOut files. Libraries simply check off their electronic journal collections from a list of journals that participate in LinkOut. With this utility, libraries can provide correct holdings information easily, and staff do not have to construct LinkOut files by hand.

- LinkOut File Validation [http://www.ncbi.nlm.nih.gov/entrez/linkout/validate.html] utility to be used by providers of links to parse their LinkOut files, ensuring the accuracy of the files before submission. Besides validating the file syntax against the LinkOut DTD, this tool will ensure that only allowable SubjectType and Attribute terms have been provided.

Additional tools are being developed to assist other groups of providers. Interested parties can subscribe to announcement lists described in *Communicating with LinkOut Providers* (next section) to be informed of new developments.

Communicating with LinkOut Providers

LinkOut resource providers can communicate with NCBI's LinkOut team in a number of ways. Users and providers can write to linkout@ncbi.nlm.nih.gov to ask questions about LinkOut. There are also three announcement lists where development related to LinkOut will be communicated to link providers:

1. Linkout-news [http://www.ncbi.nlm.nih.gov/mailman/listinfo/linkout-news] is for general announcements on LinkOut.

2. Library-linkout [http://www.ncbi.nlm.nih.gov/mailman/listinfo/library-linkout] is for announcements on development related to library LinkOut participants.

3. Tax-linkout [http://www.ncbi.nlm.nih.gov/mailman/listinfo/tax-linkout] is for announcements relevant to linking to taxonomic resources on the Web.

Box 1: Example of an identity file.

```
<?xml version="1.0"?>
<!DOCTYPE Provider PUBLIC "-//NLM//DTD LinkOut 1.0//EN" "LinkOut.dtd">
<Provider>
    <ProviderId>777</ProviderId>
    <Name>WebDatabase Co.</Name>
    <NameAbbr>WebDB</NameAbbr>
    <SubjectType>gene/protein/disease-specific</SubjectType>
    <Attribute>registration required</Attribute>
    <Url>http://www.webdatabase.com</Url>
    <IconUrl>http://www.webdatabase.com/images/webdb.gif</IconUrl>
    <Brief>On-line publisher of biomedical databases and other Web resources</Brief>
</Provider>
```

Identity File Elements

Provider: root element of the identity file.

ProviderId: unique ID assigned by NCBI.

Name: full name of the resource provider.

NameAbbr: short, one-word name of the provider assigned by NCBI. May only include alpha and numeric characters; spaces and special characters such as hyphens are not allowed.

SubjectType, Attribute: descriptions of the resources and relationship of the provider to the resources listed in the resource file. SubjectType and Attribute values appearing in the identity file will apply to all of the resources listed by that provider.

Url: URL of the provider's Web site, used in the LinkOut Providers list in Cubby.

IconUrl: logo of the provider, used to display the link from Entrez records.

Brief: short (up to 256 characters) description of the provider.

Box 2: Example of a resource file.

```
<?xml version="1.0"?>
<!DOCTYPE LinkSet PUBLIC "-//NLM//DTD LinkOut 1.0//EN" "LinkOut.dtd"
[<!ENTITY icon.url "http://www.webdatabase.com/images/webdb.gif">
<!ENTITY base.url "http://www.webdatabase.com/cgi-bin/elegans?">]>
<LinkSet>
    <Link>
        <LinkId>1</LinkId>
        <ProviderId>777</ProviderId>
        <IconUrl>&icon.url;</IconUrl>
        <ObjectSelector>
            <Database>Nucleotide</Database>
            <ObjectList>
                <Query>Caenorhabditis elegans [orgn]</Query>
            </ObjectList>
        </ObjectSelector>
        <ObjectUrl>
            <Base>&base.url;</Base>
            <Rule>an_lookup=&lo.pacc;</Rule>
            <UrlName>Caenorhabditis elegans</UrlName>
            <SubjectType>organism-specific</SubjectType>
        </ObjectUrl>
    </Link>
</LinkSet>
```

Resource File Elements

LinkSet: the root element of the resource file.

Link: an element that describes a specific set of resources grouped together by access characteristics or for convenience. A resource file may have multiple Link elements.

LinkId: an identifier assigned by the provider for its own reference. It may be any character string. Each Link should have a unique LinkId within each LinkSet or file.

ProviderId: the identifier number assigned to the provider by NCBI and listed in the providerinfo.xml file.

IconUrl: the URL to the icon that will be displayed on the PubMed Citation and Abstract Displays.

ObjectSelector: an element containing sub-elements in which providers will specify which Entrez records are being linked from by a <Link> element.

Database: a sub-element of <ObjectSelector>. Databases available for linking include: PubMed, Protein, Nucleotide, Genome, Structure, PopSet, Taxonomy, and OMIM.

ObjectList: a sub-element of <ObjectSelector> containing either the <Query> or <ObjectID> that specifies the Entrez records from which the resource will be linked.

Query: a sub-element of <ObjectList> that contains any valid Entrez search, used to select the Entrez records being linked from.

ObjId: a sub-element of <ObjectList> that contains an Entrez record unique identifier (UID).

ObjUrl: an element that contains the necessary information for the Entrez system to construct URLs to link to the provider's resources.

Base: a sub-element of <ObjUrl> that is the base of the URL for the provider's records.

Rule: a sub-element of <ObjUrl> that specifies the construction of the remainder of the URL, based upon the specification of systems where the resources reside.

UrlName: a short (two- or three-word) description of the link. This may be used when multiple links are available for a single Entrez record. This may also be used if the allowed terms in SubjectType and Attribute cannot meet the need of a provider.

SubjectType, Attribute: sub-elements of <ObjectUrl>, used to describe the subject(s) of the provider's resources, barriers (if any) to using the resources, and relationship of the provider to the resources listed in the resource file. The SubjectType(s) and Attribute(s) will be applied to the resources provided within a <Link>.