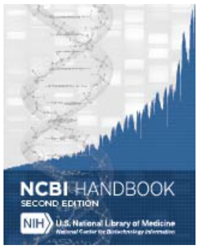
NLM Citation: Kwan YK. LinkOut: Linking to External Resources from NCBI Databases. 2013 Nov 14. In: The NCBI Handbook [Internet]. 2nd edition. Bethesda (MD): National Center for Biotechnology Information (US); 2013-.
Bookshelf URL: https://www.ncbi.nlm.nih.gov/books/


NCBI HANDBOOK
SECOND EDITION
NIH U.S. National Library of Medicine
National Center for Biotechnology Information

LinkOut: Linking to External Resources from NCBI Databases

Y. Kathy Kwan[1]

Created: November 14, 2013.

Scope

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 is a powerful linking feature of the NCBI Entrez search and retrieval system. It is designed to provide 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 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 NCBI databases. By branching out to relevant resources on the Web, LinkOut expands on the theme of Entrez as an information discovery system.

LinkOut is not just a list of links to Web sites. Two unique aspects of LinkOut set it apart from linking features in other information retrieval systems.

1. Specificity—LinkOut links users to resources that are specific to the subject of an Entrez record, e.g., linking to the full-text article of a PubMed citation, not the table of contents of the journal; to a specific section on Ginkgo Biloba, not just the searching interface for the USDA/NRCS PLANTS Database.
2. Voluntary participation—Participation in LinkOut is free and voluntary. Links are provided by external parties that create the link format, URL, and functionality. Resources reside on the provider sites, and they determine who may access their content. It is a unique collaboration with no parallel in similar data retrieval systems where links are typically created by the retrieval systems.

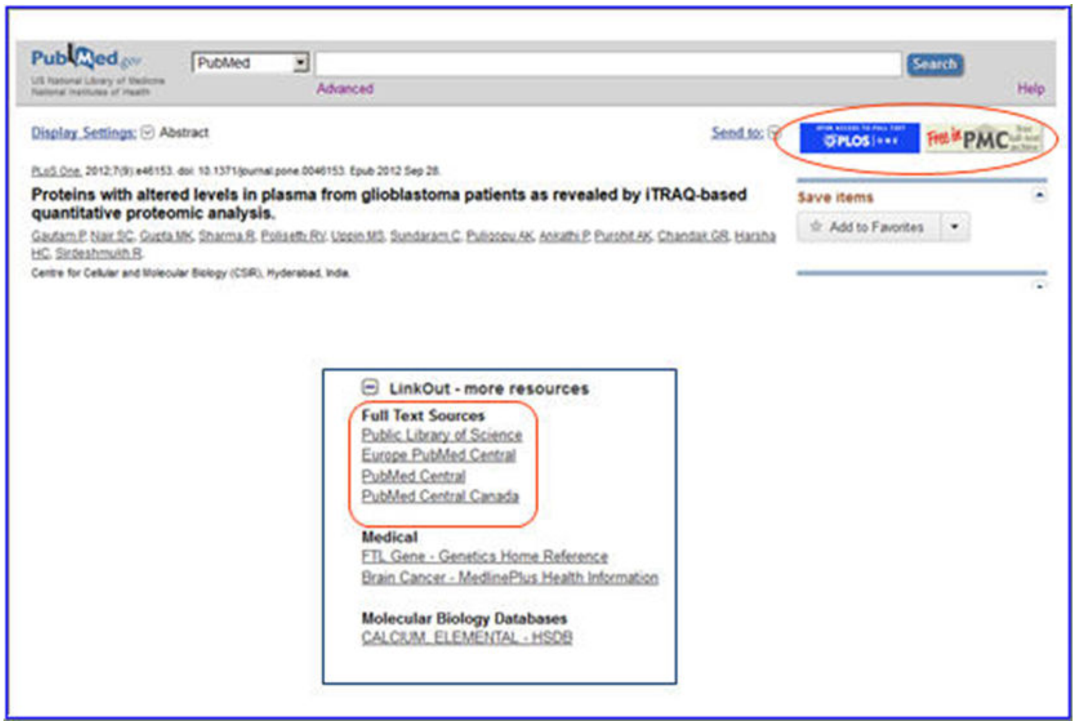
History

LinkOut was developed in 1999 during the reengineering of the NCBI Entrez system. The system was designed to allow third party providers to send links to their resources to be used by the Entrez databases.

All Entrez databases can enable LinkOut. See the current list of databases available for linking.

LinkOut has connected NCBI users to resources at over 3700 third party sites in more than 70 countries, for over 99 million NCBI database records. This greatly enhances the utility of the databases. Full text links in PubMed citations are an essential feature of the database, as shown in Figure 1.

Author Affiliation: 1 NCBI.


PubMed.gov
US National Library of Medicine
National Institutes of Health
PubMed
Search
Advanced
Help
Display Settings: Abstract
Send to:
PLOS
Free in PMC
Save items
Add to Favorites

PLoS One. 2012;7(9):e46153. doi: 10.1371/journal.pone.0046153. Epub 2012 Sep 28.

Proteins with altered levels in plasma from glioblastoma patients as revealed by iTRAQ-based quantitative proteomic analysis.

Gautam P, Nair SC, Gupta MK, Sharma R, Polisetty RV, Uppin MS, Sundaram C, Puligopu AK, Ankathi P, Purohit AK, Chandak GR, Harsha HC, Sirdeshmukh R.

Centre for Cellular and Molecular Biology (CSIR), Hyderabad, India.

LinkOut - more resources

Full Text Sources
Public Library of Science
Europe PubMed Central
PubMed Central
PubMed Central Canada

Medical
FTL Gene - Genetics Home Reference
Brain Cancer - MedlinePlus Health Information

Molecular Biology Databases
CALCIUM, ELEMENTAL - HSDB

Figure 1: Full text links in PubMed

Data Model

LinkOut is itself an Entrez database that holds the linking information to external resources. The separation of the Entrez database records (e.g., PubMed citations) 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. If links to external resources change, such as in the case of a Web site redesign, it will not affect the Entrez database records. Consequently, linking information can easily be updated as frequently as necessary.

The logical data unit of LinkOut is the relationship between a link and its target in the Entrez system. It is summarized in Figure 2.

Each Link Target consists of a database name and a record unique identifier (UID) that the link applies to.

Each Link consists of:

- **Provider** that identifies the supplier of the link;
- **URL rule** that will be used to build the URL to link to the resource at the provider site;
- **UrlName** is a text string supplied by the link provider to describe the resource;
- **SubjectType** and **Attribute** are NCBI keywords to describe the resource.

LinkOut data units are retrieved by the frontend program of the target Entrez databases when the frontend is building the display of records. LinkOut filters are also built based on these data units to facilitate retrieval of LinkOut links in the target databases.

Dataflow

LinkOut dataflow is summarized in the following diagram:

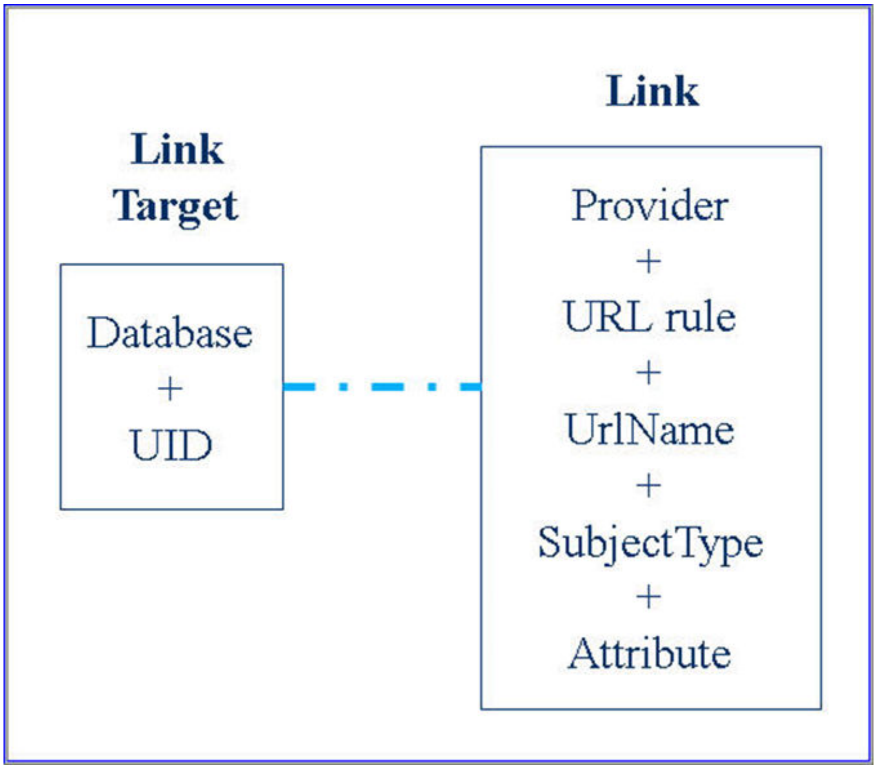


Figure 2: LinkOut Data Model

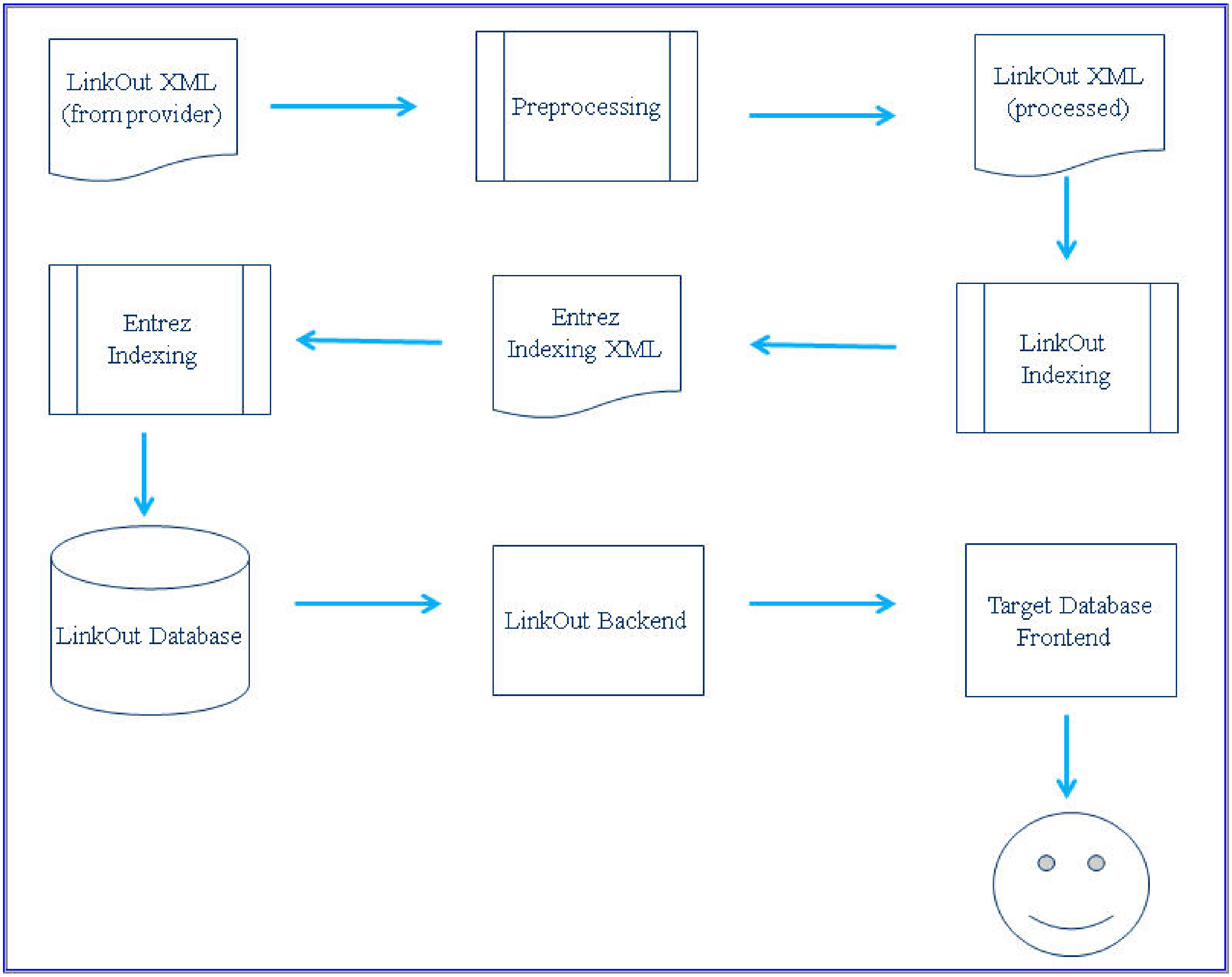


Figure 3: LinkOut Dataflow

LinkOut XML Files from Providers

LinkOut information is submitted by link providers in XML. LinkOut XML is defined by the LinkOut Document Type Definition (DTD).

The LinkOut DTD specifies all the elements needed to build the logical data unit for the LinkOut database. The DTD also specifies the link provider information needed for future communication with the providers regarding their links.

Two root elements are specified in the LinkOut DTD: the <Provider> element, which specifies information about a link provider; and the <LinkSet> element, which describes information about the link. Each root element is submitted to NCBI in a separate file. An identity file contains the <Provider> element, and a resource file contains the <LinkSet> element.

The identity file, *providerinfo.xml*, describes the identity of a provider, including an ID <ProviderId>, an abbreviated name <NameAbbr> assigned by NCBI, the provider's name, and other general information about the provider. There is only one providerinfo.xml file for each provider (see the details of the identity file).

The resource file, which contains the linking information, specifies a set of Entrez database records by UIDs or a valid query to the database, a specific rule to build the URL to an external resource, and the description of the resource using the SubjectType, Attribute, and UrlName fields. There is no standard for naming resource files, except that they must use the *.xml* extension. There may be any number of resources files associated with a ProviderId (see the details of the resource file).

Auxiliary Tools

The following tools facilitate LinkOut file submissions:

- Library LinkOut Files Submission Utility This utility was 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, without needing to construct the LinkOut files by hand.
- LinkOut File Validation This utility is used by providers to parse their LinkOut files to ensure the accuracy of the files before submission. Besides validating the file syntax against the LinkOut DTD, this tool ensures that only allowable SubjectType and Attribute terms have been provided.

Preprocessing

LinkOut files from providers are passed through a preprocessing stage that converts them to standardized files for the LinkOut indexer.

The preprocessing stage includes three sub-processes:

1. Loading—Files submitted by providers in their FTP accounts are copied over to a staging area for the LinkOut indexer. During this sub-process, the XML files are adjusted to make sure the information in the file is valid. For example, ProviderId is a valid ID assigned by NCBI and only terms from the control lists of SubjectType and Attribute are used. The adjusted files are also validated against the LinkOut DTD to ensure XML files passed to the LinkOut indexer are with valid syntax. The loading sub-process compares the XML files in a provider's FTP account and the files in the staging area and decides which files in the FTP account should be processed. Only new and changed files in a provider's FTP account will be processed.
2. XML files generation—Holdings information entered into the Library Submission Utility are transformed into LinkOut XML files and added to the loading sub-process discussed above. Only holdings from libraries that have changed holdings information are processed.
3. Icon processing—Icons specified in the <IconUrl> element are downloaded. The icons downloaded are adjusted to make sure their size is within the allowed dimensions specified by NCBI. The processed icons are transferred to the target databases to form a part of the record display.

LinkOut Indexing

The LinkOut indexing process takes the standardized LinkOut XML files in the staging area and converts them into the standard XML suitable for Entrez indexing.

This process converts all queries specified in LinkOut file <Query> elements to UIDs of the target database. Since a <Query> in a LinkOut file can be translated to a different set of UIDs as the target database changes over time, it is necessary to process all LinkOut files in each LinkOut indexing. For example, *<Query>plos one [journal]<Query>* will be translated to a different set of PubMed IDs as citations for PLOS One are added to PubMed between LinkOut indexing.

LinkOut indexing is optimized to execute duplicate queries across all LinkOut files only once. In doing so, the number of queries needed to be processed has been reduced significantly and the speed of LinkOut indexing has been improved by tenfold.

Entrez Indexing

The Entrez indexing process builds the LinkOut database using the LinkOut files in the standard Entrez indexing format. The unit of the LinkOut database corresponds to the logical data unit described in the Data Model section. The LinkOut database can be queried directly with Entrez search commands internally at NCBI. It is also the data source for the LinkOut backend program. During the Entrez indexing process, LinkOut filters are generated for each target database. A LinkOut filter is a list of UIDs of a target database with a name. Filters are sent to each target database to be included during the indexing process for the database. As a result, the filters can be searchable in the target database. See the Access section for the details on LinkOut filters.

LinkOut Backend and Frontend of the Target Database

LinkOut backend handles the interaction between the LinkOut database and the frontend program of the NCBI databases. It has standardized protocols and commands that allow access to linking data in the LinkOut database more effectively and efficiently.

Each Entrez database may present LinkOut information to users differently. Generally, a frontend program accesses the LinkOut backend to find out:

1. If a specific record has any LinkOut links;
2. If there are LinkOut links, the information for each link.

The frontend program uses the information returned by the LinkOut backend and the XSLT rules set by the target database to generate the display. The frontend program is also responsible for transforming the LinkOut entities returned by the LinkOut backend to the value for a specific record to form a valid URL.

For example, in NCBI sequence databases, *&lo.pacc;* translates into the Primary Access Number of the corresponding record. In the Nucleotide database, record GI: 467096719, the following linking rule:

http://genome.ucsc.edu/cgi-bin/hgTracks?org=human&position=&lo.pacc;

will be translated into the following URL:

http://genome.ucsc.edu/cgi-bin/hgTracks?org=human&position=NR_102347

Access

LinkOut resources associated with a record in a LinkOut enabled database can be accessed in a variety of ways. The Using LinkOut chapter of LinkOut Help contains up-to-date information on how to access LinkOut resources.

LinkOut resources are typically presented to users by SubjectType in the LinkOut display portlet of the target database and within the My NCBI system, making it easier to browse. Attributes are used to describe the nature of a LinkOut resource, e.g., whether a resource requires a subscription for access. A short text string may be used in the UrlName element to provide an additional description for 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 record in the target database.

LinkOut Filters

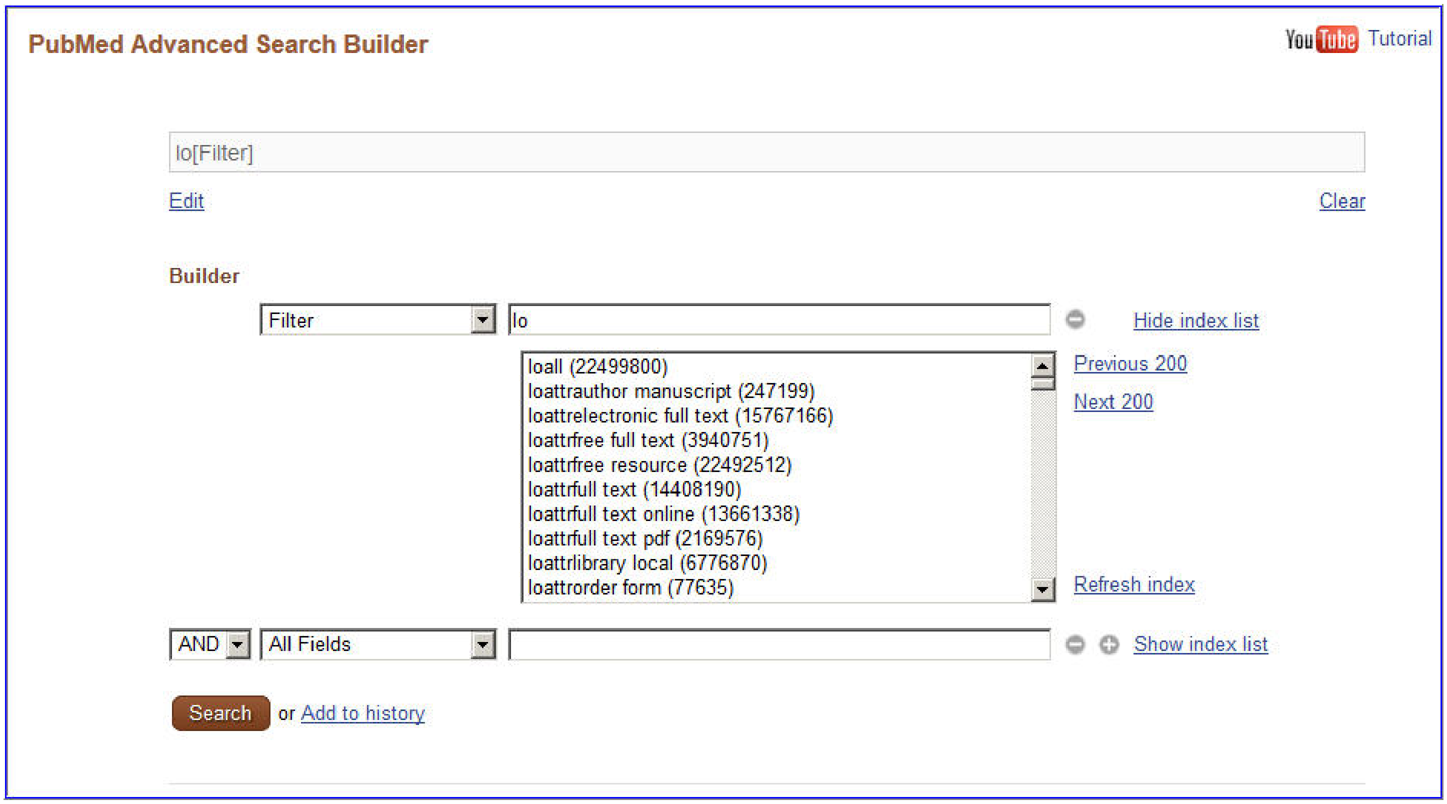
To facilitate search and retrieval of LinkOut resources, there are a number of filters in the LinkOut-enabled Entrez databases. These filters, although not part of the LinkOut database, are based on the results generated in the LinkOut indexing process. A LinkOut filter is a list of UIDs for a target Entrez database. The UIDs identify records with a common property in the target database. This property is reflected in the filter name.

LinkOut filters are all prefixed with ***lo***. Filters are available for all allowable SubjectType and Attribute terms, and the NameAbbr of a provider. To retrieve a set of records by a certain LinkOut property, a filter name can be entered as a search term in a database search box.

Examples of LinkOut filters include:

- ***loprov***—Filter for records with links to a specific LinkOut provider. Example:
  - *"loprovplos" [filter]*
  - Retrieves all records with links to the journal *Public Library of Science* in PubMed.
- ***loattr***—Filter for records with links of a specific LinkOut Attribute. Example:
  - *"loattrfull text online" [filter]*
  - Retrieves all records with at least one full text link in PubMed.
- ***losubjt***—Filter for records with links of a specific LinkOut SubjectType. Example:
  - *"losubjtorganism specific"[filter]*
  - Retrieves all records with links to resources of the SubjectType "organism specific", i.e., resource in a database providing data specific to a particular organism or group of organisms.
- ***loall***—Filter for all records in a database with at least one LinkOut resource. Example:
  - *"loall"[filter]*

The Advanced Search Builder of each database can be used to retrieve LinkOut filters. Select **Filter**, type "lo", and then click the **show index list** link to browse through the filters related to LinkOut.



PubMed Advanced Search Builder

lo[Filter]

Builder

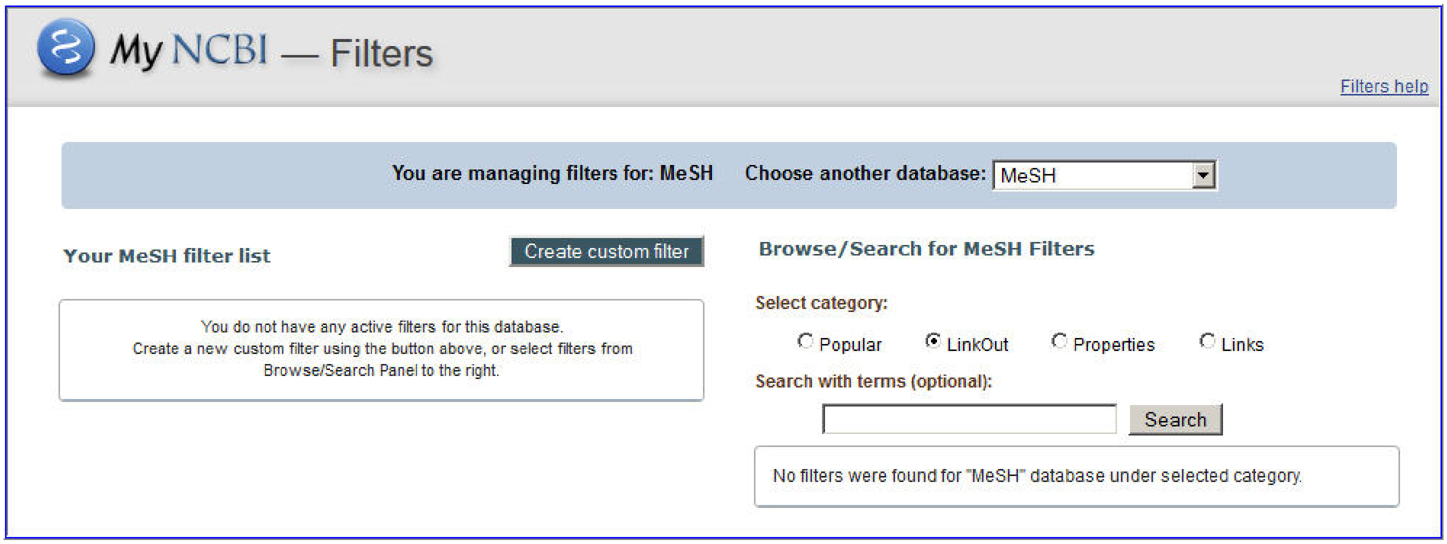
Filter: lo

loall (22499800)
loattrauthor manuscript (247199)
loattrelectronic full text (15767166)
loattrfree full text (3940751)
loattrfree resource (22492512)
loattrfull text (14408190)
loattrfull text online (13661338)
loattrfull text pdf (2169576)
loattrlibrary local (6776870)
loattrorder form (77635)

AND All Fields

Figure 4: LinkOut Filters in Advanced Search Builder

Users can also customize LinkOut filters and icons by setting the Filters preferences in My NCBI.



You are managing filters for: MeSH Choose another database: MeSH

Your MeSH filter list

You do not have any active filters for this database.
Create a new custom filter using the button above, or select filters from Browse/Search Panel to the right.

Select category:

○ Popular ● LinkOut ○ Properties ○ Links

Search with terms (optional):

No filters were found for "MeSH" database under selected category.

Figure 5: My NCBI Filters

Guide to LinkOut Providers

Information on LinkOut participation and file preparation is available from the manual: LinkOut Help. This documentation includes specific chapters for full text providers, libraries, and providers of general resources.

Information about LinkOut is available on the LinkOut home page, including FAQs, a list of existing providers, and access to various informational lists.

Users are welcome to communicate directly with the NCBI LinkOut team. Questions and comments about LinkOut can be sent to the mailing list: linkout@ncbi.nlm.nih.gov