

## Entrez Programming Utilities (E-utilities)

The E-utilities use a fixed URL syntax that translates a standard set of input parameters into the values necessary for various NCBI software components to search for and retrieve the requested data. You can use any computer language that can send a URL to the E-utilities server and interpret the XML response; examples of such languages are Perl, Python, Java, and C++. Combining E-utilities components to form customized data pipelines within these applications is a powerful approach to data manipulation. See the E-utilities Quick Start guide and related tutorials:

[https://www.ncbi.nlm.nih.gov/books/NBK25500/#chapter1.Searching\\_a\\_Database](https://www.ncbi.nlm.nih.gov/books/NBK25500/#chapter1.Searching_a_Database)

<https://www.ncbi.nlm.nih.gov/books/NBK25499/>

All E-utility calls share the same base URL:

```
https://eutils.ncbi.nlm.nih.gov/entrez/eutils/
```

EGQuery searches all *Entrez* databases simultaneously.

```
egquery.fcgi?term=<query>
```

Use ESearch for text searches in specify database

```
esearch.fcgi?db=<database>&term=<query>
```

ESummary return document summaries (DocSums) for a list of input IDs

```
esummary.fcgi?db=<database>&id=<UID#1>,<UID#2>,...
```

EFetch returns formatted data records for a list of input IDs

```
efetch.fcgi?db=<database>&id=<UID#1>,<UID#2>,...
```

Example #1: Get information at the NCBI databases related about the word “cancer”

<https://eutils.ncbi.nlm.nih.gov/entrez/eutils/egquery.fcgi?term=cancer>

**Note:** Only the number of record found for each database will be obtained as result. The default output is the XML format.

**Tips:** You can add &retmode=html to return output in HTML format

Example #2: Get the PubMed primary identifiers (PMIDs) for articles about breast cancer

<https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed&term=breast+cancer>

**Note:** By default, ESearch only includes the first 20 IDs retrieved in the XML output. You can add `&retmax=<integer>` to increase the number of IDs to be included in the output up to 100,000 records.

Example #3: Get the PubMed primary identifiers (PMIDs) for articles about breast cancer published in Science in 2008

[https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed&term=breast+cancer+AND+science\[journal\]+AND+2008\[pdat\]](https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed&term=breast+cancer+AND+science[journal]+AND+2008[pdat])

**Note:** Avoid placing spaces in the URLs, particularly in queries. If a space is required, use a plus sign (+) instead of a space.

Example #4: Download the list of protein ID for human proteins related with breast cancer.

[https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=protein&term=breast+cancer+AND+human\[organism\]](https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=protein&term=breast+cancer+AND+human[organism])

**Note:** The E-utilities return the GI as the primary identifier (ID) for the retrieved records.

**Caution:** The GI number has been used for many years by NCBI to track sequence histories in sequence databases. However, NCBI is in the process of changing the way they handle GI numbers for sequence records. That means, accession.version identifiers, rather than GI numbers, will be the primary identifiers for sequence records. Presentation of GI sequence identifiers in the GenBank flatfile format was discontinued as of March 2017.

**Tips:** See *Entrez Unique Identifiers (UIDs)* for selected databases at:

[https://www.ncbi.nlm.nih.gov/books/NBK25497/table/chapter2.T\\_entrez\\_unique\\_identifiers\\_ui/?report=objectonly](https://www.ncbi.nlm.nih.gov/books/NBK25497/table/chapter2.T_entrez_unique_identifiers_ui/?report=objectonly)

Example #5. Display entry summaries for organisms with Taxonomy IDs 9913 and 30521

<https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esummary.fcgi?db=taxonomy&id=9913,30521>

Example #6. Get nucleotide sequence with GI 21614549 in FASTA format in a plain text file:

<https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=nucleotide&id=21614549&rettype=fasta&retmode=text>

**Tips:** Values of `&retmode` and `&rettype` for EFetch will define the data format of the records returned. See the list of valid values for each dataset at:

[https://www.ncbi.nlm.nih.gov/books/NBK25499/table/chapter4.T\\_valid\\_values\\_of\\_retmode\\_and/?report=objectonly](https://www.ncbi.nlm.nih.gov/books/NBK25499/table/chapter4.T_valid_values_of_retmode_and/?report=objectonly)