

DBcat: a catalog of 500 biological databases

Claude Discala, Xavier Benigni[1,2], Emmanuel Barillot[1,2,*] and Guy Vaysseix[1,2]

CNS, 2 rue Gaston Crémieux, BP 191, 91006 Évry cedex, France, [1]Infobiogen, 7 rue Guy Môquet, BP 8, 94801 Villejuif cedex, France and [2]Généthon, 1 rue de l'Internationale, BP 60, 91002 Évry cedex, France

Received October 4, 1999; Revised and Accepted October 8, 1999

ABSTRACT

The DBcat (http://www.infobiogen.fr/services/dbcat) is a comprehensive catalog of biological databases, maintained and curated at Infobiogen. It contains 500 databases classified by application domains. The DBcat is a structured flat-file library, that can be searched by means of an SRS server or a dedicated Web interface. The files are available for download from Infobiogen anonymous ftp server.

INTRODUCTION

Identifying sources of information is crucial in today's fast moving world of Biology: not only does it allow researchers to find information rapidly for their experiments, but it also allows biologists a place (and sometimes the only one) where their results can be published and for science politicians who fund (or not) databases, a means of impacting and analyzing the directions of research (1). In Biology, as in other sciences, information is disseminated in various heterogeneous databases and this situation is unlikely to change with the announced deluge of data from the genomic and proteomic projects (2).

We developed the DBcat, a comprehensive catalog of 500 biological databases, to help all kind of biological database users to identify the information they are seeking.

ORGANIZATION

DBcat has a simple data model implemented under the form of database entries in a flat-file. All entries have the same structure that consists of a list of fields and their associated values: the main fields are the database name (NAME), its description (DESCRIPTION), its domain (DNA, RNA, Protein, Literature...), authors and contacts (AUTHOR, CONTACT) and the WWW address (URL-FTP) where the database can be browsed. As an example, Figure 1 shows the entry corresponding to the DBcat database itself.

DATA ACQUISITION

The DBcat was started at Généthon in 1994 as part of a technological survey that also gave birth to a catalog of software for molecular biology and genetics: the BioCatalog (http://www.ebi.ac.uk/biocat/) (3). The DBcat is now produced at Infobiogen. New databases are searched in the Web, either by means of general purpose Web search engines or biology-oriented Web sites. Journals, such as the *Nucleic Acids Research* Database Issue, are also consulted. The producers of the database are asked, via Email, to complete a form and to check their entries in DBcat. They can also use a Web form for spontaneous submissions (http://www.infobiogen.fr/services/dbcat/file/dbcat_form.html). If the author has validated the entry corresponding to its database, it is marked as CHECKED.

```
AC              DBC00001
NAME            DBcat, Catalog of biological DataBases
DOMAIN          Literature
DESCRIPTION     The purpose of this catalog is to list all the existing
DESCRIPTION     biological databases, including relevant information
DESCRIPTION     about these databases
CHECKED         YES
AUTHORS         Claude Discala, Marion Ninnin
RA              Discala C., Ninnin M., Achard F., Barillot E.,
RA              Vaysseix G.
RT              DBcat: a catalog of biological databases
RL              Nucleic Acids Research, Vol. 27, 10-11, 1999.
ORIGINAL-SITE   INFOBIOGEN
ADDRESS         7, rue Guy Moquet
ADDRESS         94801 Villejuif cedex
ADDRESS         FRANCE
CONTACT         discala@infobiogen.fr
CONTACT         mninnin@infobiogen.fr
SUBMIT          http://www.infobiogen.fr/servic es/ dbca t/f ile /dbc at_ for m.ht ml
URL-FTP         ftp://ftp.infobiogen.fr/pub/db/ dbc at/
URL-WWW         http://www.infobiogen.fr/servic es/ dbca t/
URL-QUERY       http://www.infobiogen.fr/srs
RELEASE         Daily
UPDATES         Daily
COMMENTS        Corrections and additions will be greatly appreciated
COMMENTS        as well as spontaneous submission
OTHER-SITE      -
ADDRESS         -
ADDRESS         -
URL-FTP         -
URL-WWW         -
URL-QUERY       -
UPDATES         -
COMMENTS        -
//
```

Figure 1. The DBcat entry of the DBcat database.

ACCESS

The DBcat contains 500 database entries, available in one flat-file. To reflect the areas of interest of the users, the database entries are also grouped into eight application domains: DNA, RNA, Protein, Genomics, Mapping, Protein structure, Literature, Miscellaneous. The number of databases listed in each domain is given in Table 1. Note that databases may belong to several domains, and only the first one is taken into account in the statistics.

The DBcat provides the users with a variety of modes of access:

- Download the flat-files: ftp://ftp.infobiogen.fr/pub/db/dbcat
- Web interface homepage with a simple query by name interface: http://www.infobiogen.fr/services/dbcat/
- SRS server: http://www.infobiogen.fr/srs/

*To whom correspondence should be addressed at: Infobiogen, 7 rue Guy Môquet, BP 8, 94801 Villejuif cedex, France. Tel: +33 1 49 58 36 83; Fax: +33 1 49 58 36 89; Email: manu@infobiogen.fr

Table 1. Statistics of the DBcat entries per application domains

Domain	No. of records
DNA	82
RNA	29
Protein	93
Genomic	58
Mapping	30
Protein structure	18
Literature	39
Miscellaneous	151
Total	500

CONCLUSIONS

Recently, a survey conducted by Ellis and Kalumbi (1) based on the DBcat concluded that two-thirds of biological databases were ‘facing uncertain funding for a very near future’ and many interesting and free biological databases were ‘on the verge of financial collapse’. This survey shows the utility of the DBcat in identifying the existing sources of information.

We invite database managers to submit any new database to DBcat. Updates and correction are also encouraged.

ACKNOWLEDGEMENTS

The authors wish to thank Patricia Rodriguez-Tomé and the GREG for supporting the first version of DBcat. We also wish to thank the authors/curators of the 500 databases listed in this catalog for providing their work and expertise to the community.

REFERENCES

1. Ellis,L. and Kalumbi,D. (1998) *Nature Biotechnol.*, **16**, 1323–1324.
2. Reichhardt,T. (1999) *Nature*, **399**, 517–520.
3. Rodriguez-Tomé,P. (1998) *Bioinformatics*, **14**, 469–470.