
NAR's new requirement for data submission to the EMBL data library: information for authors

Patricia Kahn and David Hazledine

EMBL Data Library, European Molecular Biology Laboratory, Postfach 10.2209, D-6900 Heidelberg, FRG

As of 1 January 1988, manuscripts submitted to Nucleic Acids Research (NAR) and containing or discussing sequence data must be accompanied by evidence that the data have been deposited with the EMBL Data Library. The background to this new policy and a general description of how it is being implemented were discussed in a recent NAR article (volume 15, number 18).

The following is a set of instructions describing how researchers can submit their data to the EMBL Data Library and obtain an accession number as quickly as possible.

AN OVERVIEW

The system works as follows. NAR requires that all primary sequence data reported or referred to in a manuscript have a corresponding database accession number. This number, which permanently identifies a sequence (or set of sequences) in the database, constitutes evidence that the data have been submitted to and accepted by EMBL.

This means that authors who intend to submit manuscripts reporting sequences should first send the data to the EMBL Data Library. (The data will, as usual, be shared with the other major nucleotide and protein sequence databases but should be submitted to EMBL.) The Data Library staff will assign an accession number to data which appear to be complete and accurate, and will inform the authors what this number is.

THE FIRST STEP IN GETTING AN ACCESSION NUMBER

Before you do anything else, get a copy of a sequence data submission form. This form solicits all of the information needed to make a database entry; that is, the primary sequence data together with descriptive information such as the source of the sequenced segment (e.g., organism, strain, tissue) and the location of interesting regions within the sequence (e.g., signal peptide, coding regions, regulatory signals). It also contains information about data formats.

The data submission form exists in both a hard copy and a machine-readable version; the latter can be completed using a text editor. These versions are presently available from the following sources:

- (a) Hard copy form: printed at the end of this article and available upon request from EMBL and GenBank.
- (b) Machine-readable form:
 1. With all releases of EMBL (from Release 11 on, in the file called DataSub.txt) and GenBank (Release 48 and later).

Nucleic Acids Research

2. From EMBL by electronic mail (computer network) via our file server. Anyone with access to BITNET/EARN (either directly or via a gateway) can send a request to the EMBL file server, which will automatically return a copy of the data submission form by electronic mail. Instructions for using the EMBL file server are given in Appendix I. You can also request the form by sending a message to: `datalib@embl.earn` (but this requires someone to read your message and mail a copy of the form to you, so please use the file server if you can).
3. From EMBL, on Macintosh (3-1/2") or MS-DOS (5-1/4") floppy diskettes. Complete information on how to contact the EMBL Data Library is given in Appendix II.
4. From GenBank, via electronic mail or on floppy diskette. Addresses and phone numbers for GenBank are given in Appendix II. For information on requesting the form from GenBank via Telenet, contact David Benton (+1-415-962-7360).
5. Through the BIONET National Computer Resource for Molecular Biology (Mountain View, California, U.S.A.). Anyone who is already a subscriber to BIONET can get information about submitting sequences by typing `HELP SUBMIT-SEQ` at the system prompt. People who do not have accounts should contact Kathy Berg (+1-415-962-7337).
6. In Europe (including the U.K.), by electronic mail from the SEQNET Bulletin Board (Cambridge University, U.K.). Send requests to receive bulletins to M.J. Bishop at:
`BIORELAY%UK.AC.CAM.BIO@EARN.RL.AC.UK` (via EARN), or
`BIORELAY@UK.AC.CAM.BIO` (via JANET).
7. In the U.K. those without a computer linked to JANET can use the Cambridge Phoenix Service, which is accessible by PSS. Apply to the External Receptionist, University of Cambridge Computing Service, Computer Laboratory, New Museums Site, Pembroke Street, Cambridge CB2 3QG.

WHAT TO SUBMIT TO THE DATABASE

A data submission should include the following materials (for further details, see the data submission form):

- (a) the sequence itself, in computer-readable form (computer network mail, magnetic tape or MS-DOS or Macintosh floppy diskette). Printouts will be accepted ONLY in the unlikely event that the authors have no access to any sort of computer.
- (b) a completed data submission form for EACH submitted sequence. The form is available from the sources described above.
- (c) a computer network address, a telex number or a telefax number (advisable, to help speed things up, but not required).

SENDING DATA TO THE EMBL DATA LIBRARY

Data can be sent to the Data Library in one of several ways:

- (a) Electronic file transfer: files can be sent via computer network to `DATASUBS@EMBL.EARN`. This BITNET/EARN address can be reached via various gateways from Arpanet, Usenet, JANET, etc. Ask your local network expert for help or phone us (+49-6221-387-258) or Rich

Roberts (+1-516-367-8388). Alternatively, data submitted through BIONET is automatically sent to us via computer network, so the author does not have to worry about how to do this.

- (b) Telefax to Data Submissions, EMBL Data Library. Our fax number is: +49-6221-387-306.
- (c) Normal post. See address given in Appendix II.

HOW LONG WILL IT TAKE TO GET AN ACCESSION NUMBER?

We will process your data within 7 working days of receipt and send you notification of either what accession number(s) they have been assigned or what additional information is needed.

There are several things you can do to minimize the time it takes to get an accession number:

- (a) Be sure that your submission includes all the necessary materials and that you have answered all relevant questions on the data submission form.
- (b) Check the data to be sure that they do not contain inconsistencies/errors (e.g., a stop codon in the middle of a region you have described as an exon).
- (c) Be sure you have included either a computer network address or a telex or telefax number. If this information is not provided, we will notify you of your accession number by regular post. Telephoning is costly and time-consuming, and we will therefore NOT attempt to contact you by phone.

Although we will process data submissions as quickly as we can, we strongly encourage you to submit your sequence at or before the time you begin writing the manuscript, rather than once you have finished it. This way we can process your data while you are working on the paper, and you won't end up waiting impatiently for an accession number while your completed manuscript sits around ready to be sent.

The data submission form asks whether you would like us to withhold your data from public availability until they are published.

APPENDIX I. EMBL NETWORK FILE SERVER

Computer users with access to BITNET/EARN (directly or via a gateway) can obtain copies of the data submission form, or of database entries, by sending commands to a file server running on the VAXcluster at EMBL. The file server facility is provided free of charge, though you may have to meet some or all of the communication costs, depending on the accounting system of your local computer service. Please note that the file server system described here will be operational from June 1st 1988.

To use the file server, you send commands (as electronic mail) to the BITNET/EARN address NETSERV@EMBL. Each line of the mail message you send should consist of a single file server command, and nothing else. The mail can be sent over BITNET/EARN, or from any other network which has a gateway into BITNET/EARN (e.g. JANET in the UK or ARPANET in the USA). The most important file server command, to get you started, is HELP. If the file server receives this command, it will return a help file to you, explaining in some detail how to use the facility.

In order to send electronic mail to a BITNET/EARN address, you must find out which operating system command you have to use on your own local machine,

Nucleic Acids Research

and how you should format the address NETSERV@EMBL. If you don't already know how to do this, we suggest that you contact your local computer service, or if all else fails, contact the Data Library and we will do our best to help.

For example, suppose you are using a VAX/VMS system which is a BITNET/EARN node running JNET software. To send a HELP command to the file server, you could use the operating system command MAIL as follows:

```
$ MAIL <filename> JNET% "NETSERV@EMBL"
```

where <filename> would be the name of a file containing file server commands. To request help information it would contain just one command, namely HELP. To request a copy of the data submission form, it would contain the following GET command:

```
GET DATALIB:DATASUB.TXT
```

APPENDIX II. HOW TO CONTACT THE NUCLEOTIDE SEQUENCE DATABASES

EMBL Data Library:

- (a) Computer network: datalib@embl.earn (for questions requiring a personal response); datasubs@embl.earn (for data submissions)
- (b) Postal address: Data Submissions, EMBL Data Library, Postfach 10.2209, 6900 Heidelberg, West Germany
- (c) Telephone: +49-6221-387-258
- (d) Telefax: +49-6221-387-306
- (e) Telex: 461613 (embl d)

GenBank:

- (a) Computer network address: genbank@lanl.gov
- (b) Postal address: GenBank Submissions, Mail Stop K710, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
- (c) Telephone: +1-505-665-2177