Using PSILOG, a new acquisition package to update FRANCIS

Maguer, Jacques le

Historical Social Research, Vol. 14 — 1989 — No. 4, 82-85

Using Psilog, A New Acquisition Package to Update Francis

*Jacques Le Maguer**

PSILOG may appear very classical but it offers real sophisticated possibilities if we consider the work step by step:

Creation or Recreation of the Database

The first one is the creation or recreation of the structure of the database including the definition of the controls the user wants to apply, to ensure a safe data input: he has to define the record length for storage optimization purposes; the maximum record length is 16 Kbytes. Then the user has to indicate all the fields he needs (this number being unlimited), the field titles (which are not names but access keys), the fixed or variable length of the field (the maximum being 8 Kbytes); the fields may have a repeating status (authors) and may be divided into sub-fields and sub-sub-fields in a tree structure way. A maximum of 3 hundred fields and sub-fields can be defined, some of them being repeated (affiliations, corporate authors...).

Once the database has been physically defined, the user can implement various field controls. Every time a field content is foreseeable, a control is possible. Those controls may be classified according to three main groups :

1.1. Existence Controls

Has a field a mandatory, an optional or a forbidden status'.' Is the field 2 mandatory IF the field 1 exists? or IF it does not exist? or IF a defined content is present? For example, if it was declared that the language field had a mandatory status, the record will not be validated unless some data are present in those fields.

* Address all communications to Jaques Le Maguer, Institut de Science Humaines et Sociales, 54, Bd. Raspail, F-75270 Paris Cedex 06.

1.2. Syntaxic Controls

They can be of various kinds, checking:

- the length of the field
- the nature of the field content (alphabetical or numerical)
- the validity of an ISSN by recalculation
- the validity of a data by comparing to a look-up table.

Those tables may be a list of character strings, such as a language code list; it can also be numerical: in that case you only have to define the numerical interval including the valid data.

If some tables have not been used for some time, they may be cancelled if they are really not useful.

1.3 Interfield Controls

They link the content of several fields in order to value one in relation to another.

For example: If the source document is in Arab language, no original title must be present but a translated title must be.

If the document type is a journal, the ISSN must be present (or replaced by a pseudo ISSN).

Of course, all those validations suppose the previous introduction of look-up tables which maybe an enumeration table, i.e. a list of character strings (such as country code list) or an interval table: In that case, you only have to define a numerical interval between two limits: the data arc validated if they are included into those limits.

2. The Updating of the Database

2.1. Having created its data structure, the user now has to choose his own data entry form and design the corresponding input screen which is composed of windows placed and designed anywhere according to his choice. The screen may include any information, but in the windows, only data may appear; the lengths and forms are predefined and they may be scrolled vertically or horizontally; the records often need the use of several screens as a window sometimes fills one full screen; in that case, screens appear according to a predefined order; the first screen must contain a window for the key. Some windows are filled through the consultation of a look-up table. A constant data may appear in a particular field of all records; some data may be recalled in several subsequent records (ex.: serial title, publication year...).

2.2. Another important feature is the availability of an expanded character set with:

- upper and lower cases
- accented characters
- greek alphabet
- exponents and indexes

Several work stages can be defined:

- a cataloguing stage
- a separate indexing and abstracting stage with access to a limited number of windows at each stage

Uploading and Downloading of Records

To end with the theoretical presentation of PSILOG, these are some examples of uploading and downloading in connection with a central computer; you can upload records on a mainframe:

- either for mere updating a central database and that is particularly interesting when working in documentary networks, several organizations having decided to gather their records while keeping their primary documents
- or take advantage of the mainframe capabilities for vocabulary chekking operations, or keyword translation, or data generation
- conversely PSILOG allows downloading of records on microcomputer and that may once more be useful when a network member wants to keep its own records by himself for a special use.

So a partner can make use of decentralization of input, and at the same time, creation of private files can be uploaded on host.

Any application program may be modified according to individual needs.

4. Applying PSILOG to FRANCIS

FRANCIS input software having been created in 1970 had become obsolete and we badly needed a new data acquisition package able to manage the input of 80,000 records a year divided into 20 files on human, social and economic sciences; the records' origins are geographically decentralized, the teams working to the constitution of the files are doted with collaborations from anywhere in the world, all work stations have to be connected to a central mainframe in order to create a unique multidisciplinary database covering philosophy as well as economics; what is more,

FRANCIS files use either private or common vocabularies according to the aspects covered by the indexation: for instance, some geographical terms can be used in all domains while a specific vocabulary is necessary for a detailed indexation in some areas.

Having overlooked several software, we concluded that the best way to get an adequate tool was to create it according to our needs and PSILOG was developed by the INIST, the new »French Scientific Information Institute« gathering the former two documentation centers of the CNRS, in cooperation with »Jouve«, a French printing and data processing company.

PSILOG value consists in its ability to be integrated into a whole data processing system; it takes into account data capture and a variety of controls through look-up tables on decentralized micro computers.

The second step of work consists in gathering all data belonging to some scientific fields on intermediary machines.

Then, all informations are to be transferred on a mainframe, an IBM4381 which is to achieve the whole data processing of INIST documentation and produce FRANCIS and PASCAL databases (PASCAL is a database on Science and Technology).

We began to use this software to input the 1988 records from 4 FRANCIS files:

- RESHUS (Human Sciences of Health)
- Energy Economics
- ECODOC (Economics)
- DOGE (Management)
- ISJ (Computer and Law)

It is the first time PSILOG is applied on Social Sciences files and we had to adapt it for this use. The network partners take big advantage of the decentralized upload ability. PSILOG is also used on PASCAL to enter two important parts: PASCAL-GEODE and PASCAL-IALINE and the file World Translation Index.

Several French organizations are implementing PSILOG.