C7  DATA BASE MANAGEMENT

Introduction

This note will deal with the application of Generalized Data Management Systems (GDMS) to diverse management activities in science and engineering. By a GDMS is meant a set of user facilities for the definition of a Data Base, for loading data into it, for updatiing it and for retrieval from it). At CERN, such systems are applied for the most part in support of basic activities of the laboratory, such as management of different kinds of equipment and projects. GDMS are not applied directly in the analysis of experiments which are best served by special purpose programs for making statistical calculations on data from many independent events, but GDMS are used in other physics activities ranging from the bulk management of data kept on magnetic tape to the retrieval of bibliographical and numerical data from publications. Before discussing some applications, a brief survey will be made of the systems used at CERN.

The most widely used GDMS at CERN is INFOL2) (INFormation Oriented Language) which is a simple system mainly used in a research environment, and written in "portable" FORTRAN. It has a relatively long history, being originally developed by T.W. Olle in 1966, and further developed by the University of Geneva, and later CERN, as a language which can easily be used by research workers, such as geographers or psychologists, to maintain small data collections. INFOL is now used at CERN for about 100 applications mostly small but a few of them very large. It can be used interactively, but the sequential file format and simple hierarchical structure of INFOL data bases limit its power.

The TABLOID system3) was written at CERN, also in FORTRAN, for the CDC 6000 front-end computers as a language for handling applications with more exacting requirements such as were originally encountered in the construction phase of the SPS. TABLOID allows the user to define a main file, and an unlimited number of "dictionaries" whose records are accessed by symbolic key thus allowing a full network structure. As part of the policy of using standard software, the manufacturer-supplied system, EDMS (Evolutionary Data Management System), is now gradually supersed- ing TABLOID on the CDC 6000 computers for the more complex applications. This product allows data bases to be defined in a general manner, according to the latest ideas.

It is not interesting to list all the hundred or so INFOL applications which are mainly inventories of various kinds. Some of the more notable INFOL applications are, in no particular order:

- control of approved experiments: references to the original proposal, the institutes and physicists engaged, the current status, resources used, references to the papers published, etc.
- statistical queries about personnel
- keeping medical records of various kinds
- management of tapes containing experimental data
- maintenance and printing of address lists for various services.

The main systems

The applications
- management of about 100,000 drawings on microfiche
- management of catalogues for the central stores
- control of chemical compounds
- analysis of questionnaires
- space management for groups working on experiments
- retrieval by keyword and author of titles of published papers\(^4\)
  in a) High Energy Physics, b) computer science and c) the topics covered in the INIS data base.

The TABLOID and EDMS applications also include some inventories but of a more complex nature leading for example to the detailed control of racks of electronic equipment. With TABLOID, updating and reporting facilities have been pre-coded according to user specifications, and then operated routinely. This has been the case for the installation of the 100,000 parts of the SPS accelerator and continues to be the case for the technical and financial management of the control cables and power cables of the SPS. In the latter case, new cables are optimally cut from partially used drums which are also managed by computer in the same data base. Finally, TABLOID is also used for compiling interaction cross sections culled from physics publications.

With EDMS, user procedures are also pre-coded, but an interactive query facility offers the user the possibility of easily posing unforeseen questions himself. A major EDMS application is the organization of medical examinations as a function of occupational risk and of previous checks. Another one is the monitoring of the assignment of physicists to experiments.

There are many mini-computers at CERN, and when it came to the problem of managing a collection of some forty Hewlett-Packard configurations with a total of 2,000 pieces of equipment, it was natural to apply the IMAGE GDMS which is available on a Hewlett-Packard 21MX itself. This application which is now in full service has shown that interactive update and simple retrieval are comparatively simple, but elaborate report generation more difficult on a mini-computer.

Another separate development is the use of APL on the IBM 370 for the installation of beam elements for physics experiments, an application involving much calculation, but a relatively small amount of data. APL is used only for retrieval and report generation for which it is very effective.

The problem of easier data entry to the many data bases mentioned above is currently being tackled by the use of intelligent terminals, but the Central Library has for some time been using locally designed terminals on a PDP-11 for updating of their cataloguing data, and for entry to their special data base of preprints and reports, etc. on the CDC 6000 computers\(^5\).

J.-P. Baud, G. Moorhead or S. Santiago.

2) INFOL 2 Reference Manual, in English and in French; Computer Documentation Office
4) Information Retrieval by Computer; Computer Documentation Office
5) C. Piney: "A Generalized Interactive System for the Acquisition of Structured Data". CERN DD/75/12.