

Historic and Scientific Documentation of a One Hundred Year Old Rock Collection, Now Supported by a Computer Catalogue Database

L.M. BARRON

Geological Survey of NSW, Mineral Resources Development Laboratory,
Cnr Weeroona Road and Joseph Street, PO Box 76, Lidcombe, NSW 2141, Australia

ABSTRACT. The Geological Survey of New South Wales maintains a Petrology Collection of some 50,000 rocks and thin sections which date from the 1880s. The Collection actively supports all current projects within the NSW Department of Mineral and Energy Resources, and is used by other state and federal departments, local councils, exploration companies, consultants and academics, all concerned with regional mapping, mineral deposits and mineralisation styles, building and construction materials, and with historic samples. Old bores, abandoned mine workings and areas under dams are all well represented in the Collection.

The Petrology Collection is now reinforced by a 40 megabyte computerised storage database, presented here as a working example of a first stage computerisation scheme driven by the collapse of the existing catalogue system, but constrained by prolonged and continuing severe budgetary restrictions. Completed in October 1989, this new database is still very primitive; it is only available at a single site in the Specialist Services and Applied Research Section (SSARS), and is an inhouse, incompatible storage system occupying 450 floppy discs. It is not very suited for some important database functions such as retrieval, backup or export/import. However, through its use the serious problems noted in a 1984 survey of the Catalogue have been totally rectified. For the first time ever, the catalogue is now totally typed and edited and accompanied by a preliminary index, while copies of the new catalogue volumes and technical reports and indices are kept at twenty Departmental venues including regional offices, and at the Australian Museum. A considerable amount of upgraded information derived from published and unpublished reports needs to be merged into the database and this will be targeted over the next two years.

From 1972 onwards, a major proportion (70%) of the thin sections have been examined only by the geologist who collected the rock with none of the resultant information added to the catalogue. This appalling situation must be rectified by installing the database on a centrally available computer with much better software, so that the geologists can work directly with the database at their desk and microscope.

Ultimately the Petrology Catalogue database should be reconstituted as a sub-entity within a much larger relational database covering general samples, accompanied by the introduction of a standardised indexing scheme that is compatible with all the Department's applications. Many

obvious commercial and scientific benefits can be derived from such a nexus but the present system, with its inherent inefficiencies, subverts the viability of these benefits.

BARRON, L.M., 1992. Historic and scientific documentation of a one hundred year old rock collection, now supported by a computer catalogue database. Records of the Australian Museum Supplement 15: 129–135.

The Petrology Collection of 50,000 rocks is maintained at the NSW Department of Minerals and Energy Core Library at Londonderry. The geological materials and ancillary information are traditionally supplied by geologists during the course of their mapping projects. Ideally each entry in the catalogue accompanies a thin section, a rock cutoff exactly matching the thin section, and a 10 to 20 cm piece of the sample. The last 2000 samples and all thin sections are kept at the Specialist Services Applied Research Section (SSARS) at the Geological Survey of NSW, Lidcombe. Historically, the Collection was supported by a handwritten catalogue and by an incompletely typed catalogue, but lacked any form of index.

The Petrology Collection splits naturally into two parts. The older part (from 1894 to 1954) of the Collection covers thin sections 1 to 16,000 (i.e., T1-T16000), where the only information usually available is the locality name and the rock name. The newer part (1954-) of the Collection covers from thin section 16,000 onwards (T16000-) where additionally map references, grid references, formational names and map symbols are available.

The Collection is very historic. Eminent Australian geologists figure prominently in the early lists of

collectors, while many of the samples were collected from mines and mineral prospects that are no longer operating and/or no longer available for sampling.

A survey made in 1984 by a volunteer showed that the Petrology Catalogue for the newer part of the Collection was in drastic need of typing and editing. The established procedures in use were rapidly and steadily failing, and a computerised database was obviously needed to control the current additions. The creation of such a database, if extended back to the beginning of the Collection, would: i) resolve the catalogue problems with current and recent acquisitions, ii) permit indexing of the whole Collection, iii) make the whole Collection much more functional, and iv) conserve a useful assemblage of information for merging or linking with other databases.

What hardware and software could be used for this database? The dominant share of the Department's computer funding since 1986 has been devoted to a very large graphics geological database. Virtually all of the remaining funds were committed to outfitting the Administrative Services sections and to boosting entities traditionally viewed as commercially 'viable', such as geophysics.

This has resulted in severe and prolonged budgetary

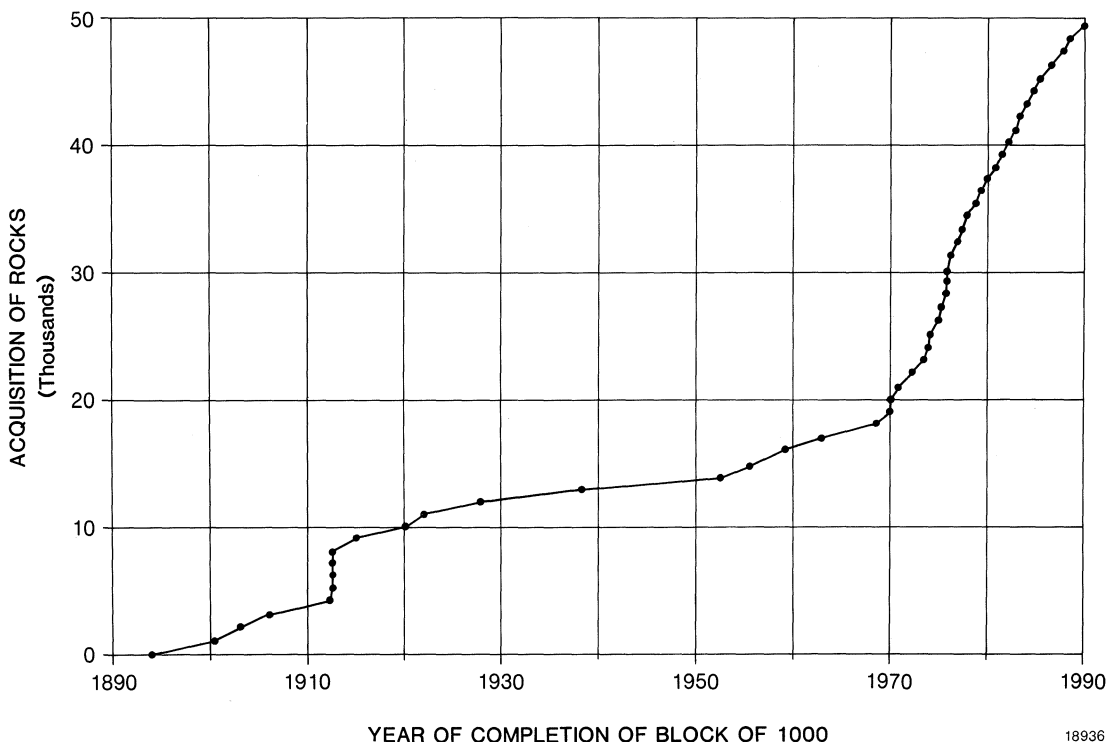


Fig.1. Acquisition history of Petrology Collection. Note that the data is assembled for the completion date at the end of a batch of 1000 rocks, i.e., the 10,000 block (10000-10999) plots at 1920.0 because it was completed at the beginning of 1920.

restrictions for all other computer applications. Any new database system for the Petrology Collection thus had to be designed as an absolutely minimal cost system, using only existing small computers, with no funds available for purchase of more suitable software or hardware. In SSARS, the only computer available was an HP9920 computer with two 260Kbyte disc drives, operating under HP BASIC III. These funding restrictions, still in force in 1990, mean that systems such as those presented in Light *et al.*, (1986) are out of reach, but some of their database functions have been written into the system described herein.

Growth of the Petrology Collection, 1984 Review of Catalogue and Design and Structure of the Computer Database

Figure 1 shows the history of sample acquisition in the Collection from about 1894. There are three average growth rates that apply over the period of the Collection, namely the first phase growth rate of 160 rocks per year lasting from 1890 to 1953, an intermediate one of 290 per year from 1954 to 1970, and the latest phase of 1620 per year since then. The gap at 1912 is an administrative

gap, in the range T4000-8000, as a result of merging a rock collection and the related but non-parallel collection of thin sections, whereby one large gap was left in the thin section numbers T4300-T7600 so that there was no overlap with similar but already assigned rock numbers. The change in slope in 1953/1954 indicates the start of more comprehensive mapping including the 1:250000 scale mapping. The next change in slope is at 1970/1971 and this occurs as a result of a new policy to catalogue ALL rocks that were thin sectioned. Prior to this, only the rocks examined by a petrologist were catalogued into the collection.

Figure 2 shows the volunteer's survey made of the Petrology Catalogue from T20000-45000. In spite of reasonable degrees of typing in the range T20000-39000 of 70 to 99%, most of these typed pages were totally unedited, and so full of mistypes as to be unreliable to use.

A database called FELIX was created (1986) on the HP9920 computer to control the newer part of the Collection. The main purpose in creating the FELIX database was to replace the Collection's supporting catalogues, namely the incompletely typed and poorly edited half bound catalogue and the rapidly degenerating 'master' catalogue composed of geologists' handwritten submittal sheets. There was a total lack of any index

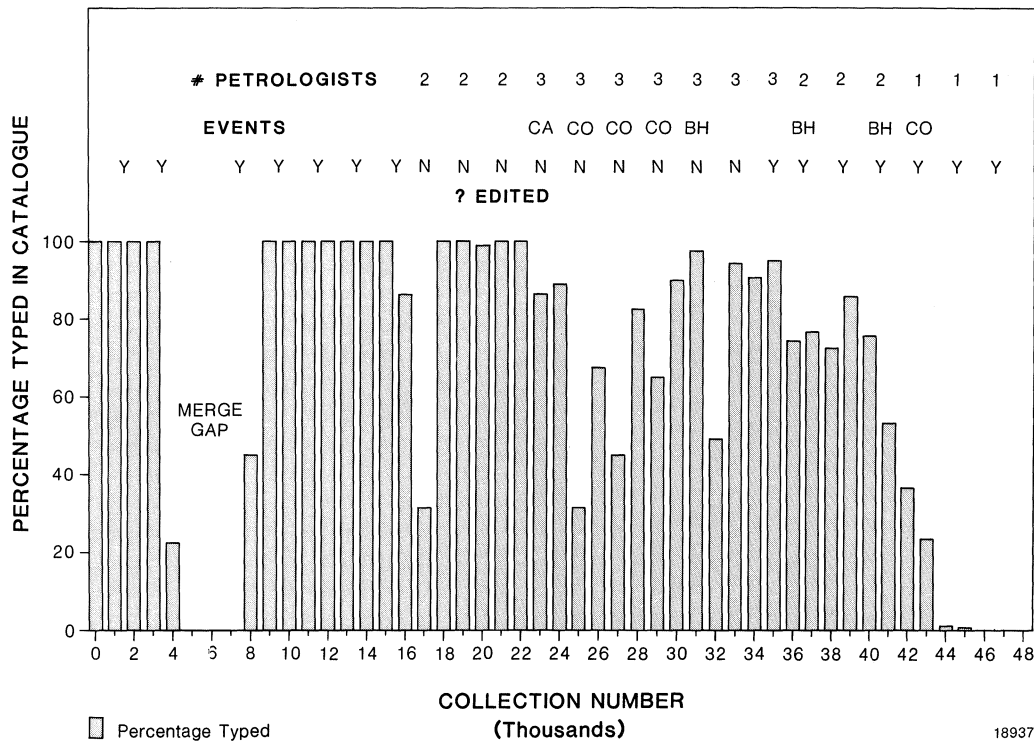


Fig.2. Graphical portrayal of the survey (25/6/1984), made by volunteer Mrs G. Maceka, on the Petrology Catalogue, arranged in blocks of 1000 thin sections. The abbreviations shown on the graph include: Y or N - EDITED-YES or -NO for the typed catalogue; 1, 2, 3 - 1, 2, or 3 petrologists employed; CA - 'Catalogue All' thin sections from 1972; CO - 'Catalogue Old' batches of ancient uncatalogued thin sections; BH - for batches of numbers assigned to Broken Hill office. For the interval from T20000 to T44000, about 70% of the catalogue was typed. The bar diagram shows that the degree of typing dropped to less than 70% in the ranges T25000-29000, T32000 and T36000-44000. Almost none of the typed catalogue pages in the range T20000-35000 were edited, but when they were briefly checked for accuracy it was found that there were so many typographical errors in both text and numbers that it was not safe to use the typed sheets.

to assist in the use of the Collection or the catalogue information, so a second major purpose in setting up the computer database was to index the Collection, initially by map reference. In 1987, in order to improve control over the older part (T1-T16000) of the Collection, a floppy-disc hosted database called PETREG was set up on a privately owned XT computer. The overall Petrology Catalogue computer database, represented by the FELIX and the PETREG data structures, was completed in October 1989.

The design phase of the project showed that three factors were paramount. Firstly, the large number of samples and the relatively large quantity of information per sample (sum of the field sizes is 2000 characters) meant that per sample entry and editing times had to be kept to an absolute minimum. Other text-based problems included historical variation in geological terms, and in naming and spelling of localities. These, and lack of confirmation of much of the geological information, precluded any homogeneous word-indexing on rock types or geology etc., and indirectly excluded any additional components of collection maintenance such as loans or stocktake.

Secondly, the available computer equipment is inappropriate for any use except as a host for a single-site storage database with the only access being through the catalogue number, aided by the simplest of numeric indices based on map references. The database is not an effective retrieval system, only a storage system.

Thirdly, the database is regarded only as a temporary capture vehicle. At some time in the future, the database information must be transferred over to a centrally available computer system with appropriate software and hardware, specially aimed at general use by the geologists.

With these factors in mind, the first priority was to retain the natural structure in the data in spite of the very real difficulties imposed by the hardware/software combination, and secondly the computer programming (in the BASIC language) had to be kept to a reasonable, low level. The software only had to be sufficient to encompass entry, general and field-specific editing, low level context-editing, and mass and selected printing. Intermediate lists of selected thin section numbers would be maintained as data files to aid the selected printing mode, for instance all rocks collected by a particular person, or all rocks on a certain map area.

Even though the above design considerations were ruthlessly enforced, the completed computer database occupies 450 floppy discs, including an on-site and an off-site backup copy. Concatenation of descriptive fields and encryption of reference fields are the main processes used to cram the potential 100 megabytes of information into the 40 megabytes held on the database discs. The first process saves 50 megabytes of storage space, while the second not only saves 10 megabytes but also automatically generates the numbers which are used as the map index values. Each data disc is 98% full and totally devoted to the database, carrying the information for a block of 250 consecutive records portrayed in a random access file, an encryption plus index file and

lastly a dummy file with a name identifying the thin section range for the disc. Because of the large number of data discs involved, about 15% of the coding of the programs is devoted to ensuring that any disc read or disc write is directed to the correct data disc.

The database has a simple hierarchical structure, with fields which parallel the Catalogue layout that has evolved over the last twenty years:

- 1) name of collector, date, project and related general references;
- 2) thin section number which is the primary key field and is assigned at the time of preparation of thin section;
- 3) field number as given by the field geologist at the time of mapping and collecting;
- 4) map name and number of 1:250000 scale map;
- 5) map number of the 1:50000 scale map;
- 6) grid reference for the sample locality;
- 7) locality name and reference information about the sample-depth inside mine or down a drill hole;
- 8) formation name and geological age and map symbol;
- 9) geological context of hand specimen, macroscopic observations plus references to further processing;
- 10) rock name with an special implanted character to mark the ensuing words for bold printing and ultimately for word indexing; and
- 11) brief petrological description.

On the database, fields 1, 4 and 5 are encrypted, while fields 7 to 11 are concatenated. Entry of information onto the computer database can be sequential or random by way of the catalogue number. A limited but effective validation is performed on the map names and references, but the rest of the information is verified by inspection, editing and printing. Unfilled records are tagged so that duplication of catalogue numbers is not possible. That is, once an new entry has been made, the record no longer can be treated as new, and it can only be edited or printed. There is provision for duplicating record information, either in parrot fashion from the last record entered (the default) or in prompted mode from any existing record. Editing is greatly assisted by programmable soft keys, a recall key and by parametric control functions that direct the program to any specified records or fields to perform batch and context editing. The program must prevent storage of more than a total of 1000 characters of information per rock, so some editing and parsing of petrological descriptions is frequently required. About six man-years of total effort have been expended in capturing and proofing the database information, roughly ten minutes per rock. Figure 3 shows the capture history of the computerised database, from the date of inception in mid-1986. The Collection has grown by 5000 rocks since then, but the database catalogue procedures are so much more efficient than the old manual techniques that there has been no difficulty in managing the capture of current acquisitions alongside the dominant job of bringing in the rest of the Collection.

Supervised Rock Names Versus Loss of Petrological Information

An old typed Catalogue exists for the T1-T19200 part of the Petrology Collection, and historically it appears that every rock was examined in thin section by a petrologist with the resultant upgraded brief rock name appearing in the typed catalogue. These upgraded rock names have been placed on the database. During the period covering acquisition of this part of the Petrology Collection, the Collection grew at the low average rate of 190 rocks per year. The rate of acquisition since 1970 is much higher at 1630 rocks per year resulting in 30,000 new records in the catalogue and bringing the catalogue up to T50000 as of the end of 1989. Over the same period, petrologists have examined about 9000 rocks, leaving a shortfall of almost 20,000 undescribed rocks, i.e., 70% of the rocks in the latter part of the Petrology Collection. In spite of concerted efforts during 1980 to 1988 to attack the shortfall, it proved intractable and under the present system will continue to grow exponentially as the Geological Survey swings back towards regional mapping with a resultant increase in the acquisition rate to the Petrology Collection.

The submitting geologist: i) makes petrological observations on a majority of the thin sections added to the Petrology Collection, ii) knows the field setting and the purpose for collecting the rock and for having

the thin section made, and iii) makes highly relevant comments, typically including structural details, mineral modes etc., up to 2000 characters of additional information.

This considerable detail is now only available on the back of the geologist's field cards so there is only one copy. This leads to a steady short term wastage of the information in the period after completion of the project, and may even result in a total loss in the long term due to accidental disposal etc. On top of this the comments are usually handwritten and unedited and so are difficult to use elsewhere. Lastly, the field cards are generally not available to the Petrology Catalogue because they are very seldom submitted for capture in the Catalogue. The geologist's revised formational and rock names and brief field and petrological observations are not added to the Catalogue for the majority of the thin sections.

The Catalogue thus contains mostly the unconfirmed rock names assigned by the geologist at the time of the field work, even though much better information is subsequently available. Although the computerised Catalogue database is now operational, this 'loss' of revised information will continue because the database is not available as a general working tool. This gross wastage of information can only be stopped through the use of a central computer facility and appropriate software. The database must be made routinely available to the geologists so that they can record their own

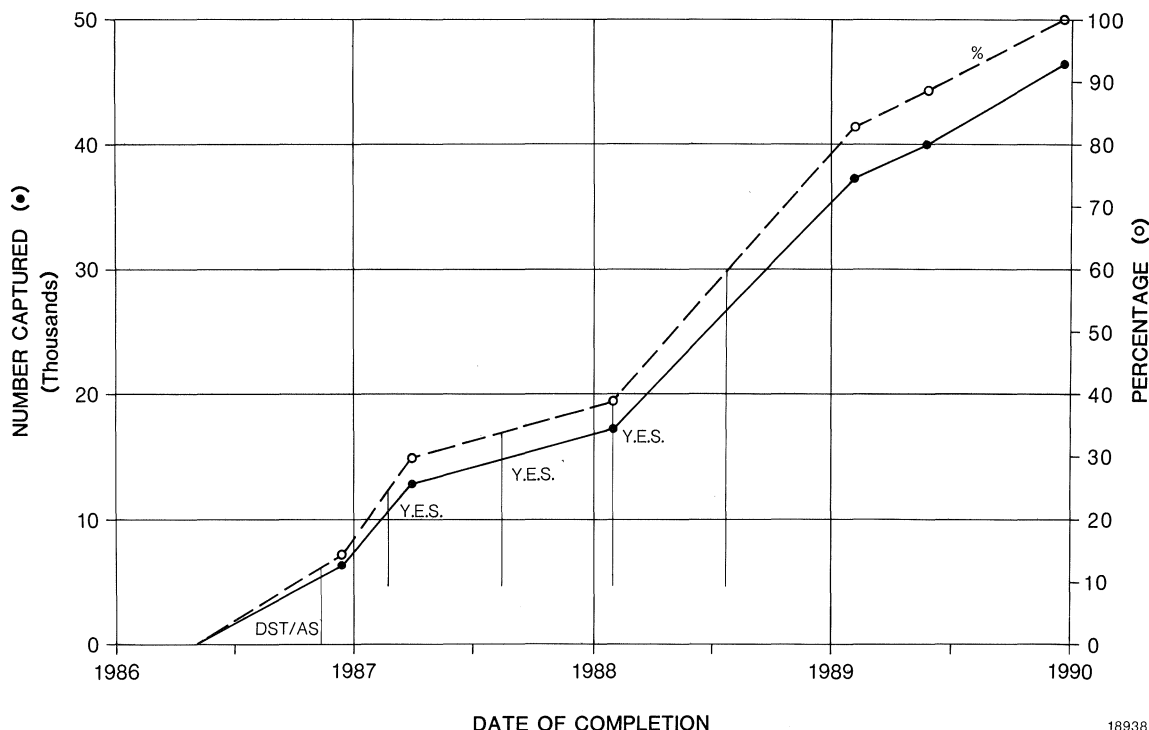


Fig.3. Capture history for the computerised database for the Petrology Collection. The solid line is for the number captured (use left hand scale), while the dashed line is for the percentage of the total pool (use right hand scale). Major assistance was received from the Australian Department of Science and Technology and Academia Sinica at the beginning of the project (DST/AS), while three Youth Employment Scheme (YES) personnel worked solidly in the middle stages of the capture history. The initial stage (DST/AS) involved capture of all records pertaining to the 4000 rocks from the Geological Survey's largest-ever mapping project: Broken Hill (1972-1983).

observations directly from their desk or microscope onto a computer into the Catalogue database. If geologists frequently use the database, then they will find it greatly to their advantage to upgrade the geological age and formational names as well. The assemblage of information would increase in both quality and quantity, allowing the geologist to retrieve rocks collected by previous geologists from the same map area, thus improving the quality of information in the older parts of the Petrology Collection as well.

Future Developments for the Computerised Catalogue Database

Many of the problems associated with placing the Petrology Catalogue on a computerised database are now solved but the database is only available at the single venue in SSARS. This highly restricted availability is maintained in response to the 'access' problem - having

a computerised database with 80 users all updating independently of each other on 'parallel' but not linked copies of the database. This is a problem so major that to date it has prevented any broadening of the pool of active users beyond SSARS, severely hampering the long term widespread purpose of the database. Supplying geologists with information from the database is becoming increasingly time consuming, and, at present, not all requests can be met, so it is obvious that use of the database has to be passed over to the geologists. Another future development is to append highly useful additional information onto the computer database, information that would not normally be placed in an ordinary catalogue, such as rock descriptions, modal variations, references to other work and global indexing.

Most of these future developments cannot be performed with the available computer or presently used software because of insufficient capacity and poor performance. The Department is investigating the purchase of a local area network system plus head-office linkup to upgrade the computer equipment for the SSARS, and one of the

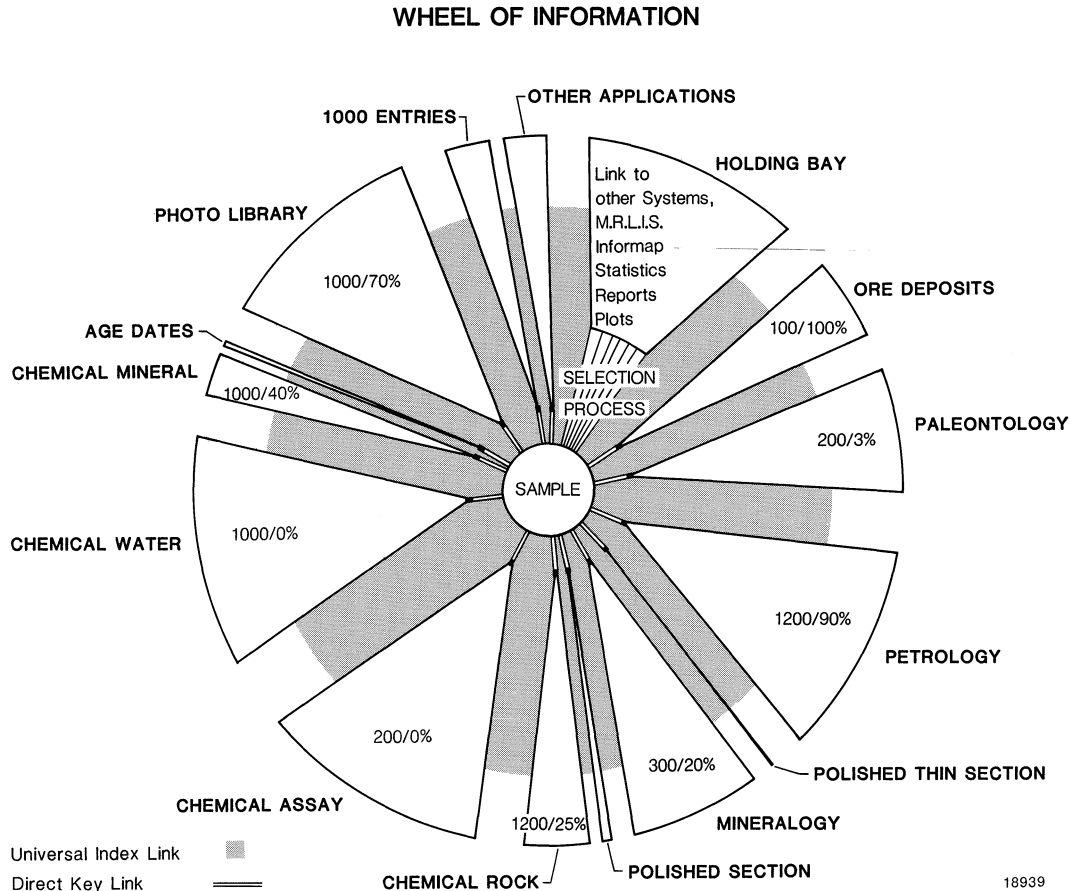


Fig.4. Suggested structure of a centrally installed data complex. The hub of the complex is the database SAMPLE, which carries the standard common information pertaining to a sample, namely collector's name, date, sample number, location, grid reference, formation, rock or mineral name. Additional and more detailed information pertaining to a sample would be placed in the appropriate subdatabase, but with a direct link, via a primary key, back to SAMPLE, i.e., via thin section number for a rock in the Petrology Collection. The size of the pie sections are indicative of the relative size of the existing applications, while the annual rate of addition and the percentage already captured into a computer database are portrayed inside the pie section. Note that the universal index link would provide a means of cross-linking into and retrieving from all related applications.

major benefits from such an upgrade would be a dissolving of most of the hardware and software problems inherent in the widespread use of a Petrology Collection computerised database. The data from the existing PETREG and FELIX computer databases would be transferred off the presently used PC computers, and reloaded as a new unified database, using much more powerful software. Batch upgrading of the information would be used to provide standardised map references for the whole data mass, hopefully along with indexing of locality and formation names.

The more conceptually difficult task of geological and general indexing has yet to be addressed, because this would have to be compatible with indexing schemes currently under development in support of other applications in the Department. It is hoped that a screen-based indexing menu can be installed as a memory-resident utility, so that some initial indexing can be performed during routine entry and editing. This would ensure standardised indexing, a feature which will probably become mandatory for all data structures as they become larger and more interlocked with other data structures, both within the Department and externally to other Departments. With resolution of the problem of lack of grid references for half of the Collection material, it may be possible to link the catalogue database into the Department's Mineral and Land Information System which is a graphical database including information on geology and ore deposits. At this stage, productivity increases of two to five times are expected compared with the just-completed database, which itself is about 20 times more productive than the old manual system of preparing the catalogue.

One major advantage of adopting a standardised indexing scheme for all of the Department's computerised database applications is so that individually the databases can originate with and retain considerable local flexibility and inherent non-standardised or incompatible structures, yet be cross retrieved through the universal index link. See Figure 4 for a possible structure that would include the present Petrology Catalogue application. The Australian Mineral Foundation has published a global geoscience indexing scheme (AMF, 1979), and they should be encouraged to support their contribution by developing a memory-resident module capable of on-screen indexing, using the built-in hierarchical structure to supply or prompt for the inherent up-stream index words.

ACKNOWLEDGMENTS. Publication is with the permission of the Director-General NSW Department of Minerals and Energy. Reviewers comments are appreciated.

References

- AMF (1979) Australian Thesaurus of Earth Sciences and Related Terms, 2nd edition, Australian Mineral Foundation. Light, R.A., D.A. Roberts & J.D. Stewart, 1986 (eds). Museum Documentation Systems: Developments and Applications. Butterworths, London.

Accepted November 19, 1992