IGBADAT: A World Data Base for Igneous Petrology

by Felix Chayes

IGCP Project 163, which ended in 1985, designed and began the construction of a retrospective world data base for petrology that contains readily sortable and retrievable petrographic, mineralogical, chemical and geological data concerning igneous rocks of all types and ages. This article describes the data base, IGBADAT, now being expanded and disseminated by the IUGS Subcommittee on Data Bases for Petrology (a part of its Commission for Systematics in Petrology) and IGCP Project 239, each of which has a mandate considerably broader than that of Project 163. An account of how these new organizations plan to maintain, expand, and manage the data base is also given.

The Public Base; Present Size and Prospects

Late in April of 1984, Project 163 (Fig. 1) submitted to World Data Center A in Boulder, Colorado, its first product - a data file and a file of source reference citations - IGBADAT and IGBREF - each in the format prescribed for it in Project Circular IGB832*. On 7 May, 1985, the first of what we hope will be annual or perhaps biennial "editions," IGBADAT2 and IGBREF2, was submitted to Center A. This was 30% larger than the 1984 submission and contained descriptions of 10,115 specimens, each with quantitative chemical information, cited 949 source references, and was 3.4 megabytes long. If this rate of expansion could be maintained, the base would contain 37,000 specimen descriptions by 1990.

We are now attempting to decentralize the data transfer procedure, and if this proves successful, we will be able to develop a considerably more aggressive campaign for "capturing" data. In this case the current expansion rate could be exceeded for some years. For example, if each of 6 sub-centers spent half the effort on data capture and transfer work that the central office expended in 1984, the base would contain over 55,000 specimen descriptions by 1990. Such a growth rate is not sustainable, however, without considerable decentralization of the data transfer operation. Indeed, without this, even the current level of activity cannot be maintained.

The Data Transfer Step and its Decentralization

Two jargon terms common to data base management systems characterize the two so far rather distinct steps of the IGBADAT data gathering process. For the (non-electronic) movement of data from publication to project coding sheet (Fig. 2) we use the term "data capture," and for the movement of data from coding sheet to first machine readable form we use the term "data transfer."

Figure 3 shows the proposed flow of information into IGBADAT from contributors submitting digitized copy (CD), or hard copy on project coding forms (CH). Hardcopy is routed through the office of a national or regional group if

possible, but otherwise directly to the Subcommittee office. Coded copy is referred by either office to a coordinator (not necessarily the same one). A contributor transferring data via microcomputer submits output discs, either directly or via a national group, to a coordinator able to convert them.

The coordinator transfers the data or converts the data discs to blocked, fixed length record files in IGBA format, and forwards the files to the Subcommittee office. When both the contributor and the central office are satisfied that no further editing is required, the file is marked for inclusion in the next update scheduled by the Subcommittee, at which time it enters the publicly available base.

Figure 1: On the beach. Some survivors of the terminal meeting of IGCP Project 163 in Lisbon, 1984. Front (l to r): J. Frizado, C. Stewart, D. Stewart (U.S.A.), M.T. Miranda, R.A.D. Gomes (Portugal), J. Hiron (Brazil), M. Godinho (Portugal). Back (l to r): J. Durham, A. Duncan (U.K.), F. Chayes (U.S.A.)

In principle, data transfer may be done either on stand-alone microcomputers or on terminals linked to large mainframes or sizeable minicomputers. The use of microcomputers for this purpose, a most promising development, will be an important focus for the early activity of Project 239. In current work along these lines much of the distinction between data capture and data transfer vanishes. Programs now undergoing most active development are carefully designed to draw the petrologist who does the capture work into the transfer process. In successfully persuading petrologists they should contribute to the base, the group would thus immediately recruit a very large corps of transfer operators.

Success in this endeavor would create new, interesting data handling and management problems for the central office, mostly of a kind whose solution is not likely to involve large labour or equipment overheads. More importantly, such success would have the effect of thoroughly decentralizing

*Copies of Project Circulars are available, in limited quantities, from the author.

A. HEADING INFORMATION

ROCK NAME (TEXT HERE, NO. IN BLOCK B)

GEOLOGIC UNIT

LOCATION LATITUDE
LONGITUDE

REFERENCE NO.

STATUS INDICATORS: 1B, 1C, 1D, 2A, 2B, 2C, 2D, 2E, 2F, 2G, 2H, 2I,

3A, 3B, 3C, 3D, 3E, 3F, 3G, 3H, 3I, 3J, 3K, 3L, 3M, 3N, 4A, 4B, 4C, 4D, 4E, 4J, 4K

B. ESSENTIAL OXIDES

SiO2	TiO2	Al2O3	Fe2O3	FeO	MnO	MgO	CaO	Na2O	K2O	P2O5	CO2	H2O+	H2O-	TOTAL	RKNUM

C. TRACE ELEMENTS

NAME															
AMOUNT															
REFERENCE															

NAME															
AMOUNT															
REFERENCE															

NAME															
AMOUNT															
REFERENCE															

D. AGE

STRATIGRAPHIC

REFERENCE NO.

ISOTOPIC OR PHYSICAL

YEARS															
METHOD															
MATERIAL															
REF. NO.															

E. PETROGRAPHIC DESCRIPTORS

ERUPTIVE TYPE OR
MODE OF OCCURRENCE

AA NO MODE OR TYPE
INFORMATION

AB AA
AC AGGLOMERATE
AD ASH
AE ASH FLOW
AF BATHOLITH
AG BLOCK LAVA
AH BOMB
AI BOSS
AJ BRECCIA
AK CONE SHEET
AL DIATREME
AM DIKE (DYKE)
AN DOME
AO EXTRUSIVE
AP FLOW
AQ FLOW BRECCIA
AR HYALOCLASTITE
AS HYPABYSSAL
AT IGIMBRITE
AU INTRUSIVE
AV INTRUS(IVE), (ION)
BRECCIA
AW LACCOLITH
AX LAPILLI
AY LAVA
AZ LAVA LAKE
BA LAYERED INTRUSION
BB LOPOLITH
BC NECK
BD NUÉE ARDENTE
BE NODULE
BF PHACOEHOE
BG PHACOLITH
BH PILLOW LAVA
BI PIPE

BP ROPY LAVA
BQ SCORIA
BR SEGREGATION
BS SILL
BT SPATTER
BU STOCK
BV SUBAERIAL
BX SUBMARINE
BY TEPHRA
BZ TUFF
CA TUFF BRECCIA
CB VEIN
CC VOLCANIC
CD VOLCANICLASTIC
CE WELDED TUFF
CF XENOLITH
CG XENOCRYST
DP OTHER(MODE, TYPE)

TEXTURE, STRUCTURE

DG NO FABRIC INFO.
EE ALLOTRIOMORPHIC
DR AMYGDALE
DS APHANITIC
DT APHYRIC
DU APLITIC
DV BANDED
DW COARSE
DX CRYPTOCRYSTLN
DY CUMULATE
DZ DEVITRIFIED
EA DISCRYSTALLINE
ER DOLERITIC
EC EUCRYSTALLINE
GI EUTAXITIC
ED EQUIGRANULAR
EF FELSITIC
EG FINE

EN GRANULITIC
EO GRAPHIC
EP HOLOCRYSTALLINE
EQ HOLOHYALINE
ER HYALINE
ES HYALOPILITIC
GE HYPIDIOMORPHIC
ET HYPOCRYSTALLINE
QH IGIMBRITIC
EU INTERSERIAL
EV LAMPORPHYRIC
EW LINEATED
EX MASSIVE
EY MEDIUM
EZ MICROCRYSTALLINE
FA MICROGRAPHIC
FB MICROLITIC
FC MICROPEGMATITIC
FD MICROPOIKILITIC
FE MICROPORPHYRITIC
FF MICROSPHERULITIC
FG MYRMKITIC
FH OPHITIC
FI ORBICULAR
GF PANIDIOMORPHIC
FJ PEGMATITIC
FK PERLITIC
FL PHANERITIC
FM PILOTAXITIC
FN PLATY
FO POIKILITIC
FP PORPHYRITIC
FG PUMICEOUS
FR RECRYSTALLIZED
FS SCHISTOSE
FT SCORIOCEOUS
FU SERIATE
FV SPHERULITIC
FW SPINIFEX
FX SUBOPHITIC
FY TRACHYTIC

HY OTHER (TEXTURE-
STRUCTURE)

STATE OF PRESERVATION

IA NO ALT. EXT. INFO.
IB FRESH
IC ALTERED
ID SLIGHTLY ALTERED
IE MODERATELY ALTERED
IF EXTENSIVELY ALTERED

IJ OTHER ALT. EXT.

TYPE OF ALTERATION

IK NO ALT. TYPE INFO.

IL ARGILLITIC
JE CARBONATIC
IM CHLORITIC
IN DEUTERIC
IO FENITIC
IP HYDRATED
IQ HYDROTHERMAL
IR LEACHED
IS METAMORPHIC
IT METASOMATIC
IU OXIDATION
IV PALAGONITIC
JF PREHNITIC
IW PROPYLITIC
IX PYRITIC
IY SAUSSURITIC
IZ SERICITIC
JA SERPENTINIZED
JB SILICIFICATION
JC SOLFATARIC
JD WEATHERED

Figure 2: Project coding sheet (in part).

the data transfer operation, an activity that is labour intensive. If this operation has to be carried on as a head-quarters activity at anything approaching the desired level, it will require stable funding for which, at the moment, there seems not much hope.

Attractive as it is in prospect, however, practical data transfer by stand-alone microcomputer is still in its infancy. Indeed, as far as IGBADAT is concerned, it is still prenatal, for all the information in the data base has reached it via terminals linked to mini-or mainframe computers, and well over 80% of this transfer was carried through at the central office of Project 163. Software developed for this purpose has recently been extensively reorganized with a view to increasing its portability. All of it, thoroughly tested, is now in routine onsite operation. The objective of this work is to generate a program package, IGBAIO, that will compile and execute on any mainframe or sizeable minicomputer, the operating system of which includes a FORTRAN-77 compiler.

Figure 3: Proposed flow of information into IGBADAT: C - contributor; D - digitized copy; H - hard copy.

IGBAIO, unlike the project's new software for IGBA input via microcomputer, makes no overt attempt to draw the contributing petrologist into the data transfer operation. Rather, it sets up that operation in such fashion that it can be carried through by bright, well motivated student assistants who have or soon acquire considerable skill at data entry. A copy of IGBAIO is available to any petrologist or group of petrologists sufficiently interested in using it to send in a tape on which it can be written. Project participants from India, Italy, Kuwait, Spain, U.K. and U.S.A. have already expressed interest in serving as "coordinators" for the transfer of data from coding sheets submitted by individual contributors. Indeed the Subcommission gives top priority to helping defray the costs of IGBADAT data-entry incurred by coordinators.

Although routine contribution to IGBADAT via microcomputer is still in the future, recent developments suggest that it may not be long in coming. Project participants have recently seen demonstrations or heard reports of microcomputer software for conversational transfer of coded data to international IGBA format on Apple-IIe, Dec-Rainbow and IBM-PC computers, to U.K.-IGBA format on CP/M-operated and BBC computers, to Japanese-IGBA format on PC-9800 computers under system SORD, and to Portuguese-IGBA format, via DBMS "Knowledgeman," on a Sirius computer. Translators from national to international IGBA format are

nearing completion (Japan), under development (U.K.), or being discussed (Portugal). Though not yet released, a PDP-11 version of the IGBAIO transfer processor has recently been developed, with initial output in international IGBA format.

File management procedures and some other programming practices used in the current IGBAIO package, which compiles and operates routinely under VAX/VMS, are being modified for on-line transfer to other larger computers, and a version for operation under UNIX has now been independently prepared by two members of the US-IGBA.

Public Access; Current and Prospective

IGBADAT cannot, of course, perform its intended function unless it is readily accessible to the concerned scientific public. In this section, the current situation with regard to public access is outlined, and some indications of how it might be improved are offered.

At present, complete tape copies of the base and bibliography are available from World Data Center A; other Centers may obtain these by exchange and individuals and other organizations by purchase (~ \$100). Enough documentation is provided to enable an interested and reasonably adept programmer unfamiliar with IGBA grammar to prepare *ad hoc* high-level language programs with which to begin exploiting some parts of the base.

IGBADAT, however, requires extensive symbol glossaries and numerous rather arbitrary grammatical conventions, described in detail in Project 163 circulars. Users who find the documentation provided with the WDC-A copy of IGBADAT insufficient for their purposes should request fuller documentation from the Subcommission. An information system equipped with a query language that would make it unnecessary for users to know anything at all about the glossary or symbolic vocabulary of IGBADAT, though certainly conceivable, has not yet been developed. Meanwhile, only casual acquaintance with IGBADAT vocabulary and grammar would be required of a user who had access to a library of simple subroutines, each of which decoded information of a particular kind from the base or extracts of it. Development of just such a user library is the specific charge of Working Group C of IGCP Project 239, chaired by R. Le Maitre; the initial version is about to be released for testing.

The rapidly increasing number of petrologists who want to do their petrographic calculations on stand-alone microcomputers face two additional problems in gaining access to IGBADAT. The first concerns the volume of data. On average, the "card image" file requires about 1/3 megabytes of disc space for each 1,000 specimen descriptions, and the base already contains over 10,000 specimen descriptions. The microcomputer user operating without hard disc storage, thus, can neither house nor use the base. A fixed cartridge disc of 5 or 10 mb capacity would not much improve the situation, unless it were dedicated to storing and processing IGBADAT. Now, and for some time to come, most petrologists who plan to exploit it via microcomputer will require sorted and filtered extracts from the base.

Attempts to provide capability for generating extracts of suitable size are now underway at World Data Center A, in the central office of Project 239, and elsewhere in the U.S.A. At the project office, an *ad hoc* information system in FORTRAN-77 is under construction, with the initial objective of generating verbatim retrievals, in project card image format, of specimen descriptions that satisfy a user-specified sorting battery.

The second special obstacle faced by the user attempting to scan IGBADAT or extracts from it on a microcomputer is that floppy disc formats differ from machine to machine. In the course of being written to a disc of the right size, a retrieval (also of the right size!) must be converted to a disc format acceptable to a particular machine and operating

system before it can be used. To effect this reformatting in routine fashion, some of the coordinators will be requested to provide two-way conversion facilities. The logistics of the operation are still unclear, but the first of the Subcommittee's two-way coordinators will probably be one that handles both the reading of discs already prepared and the generation of discs to be read under MS-DOS and DOS 3.3. Whether it will prepare input for the latter operation from its own copy of the base or be supplied with it by the central repository has not yet been decided.

Figure 4 shows the proposed flow of service requests (dashed lines) from, and digitized data retrievals (solid lines) to, users with access to large computers (UL) or stand-alone microcomputers (UM). Either class of user may direct requests to the central repository or to the office of a national or regional group. UM users may also direct requests to certain specified coordinators. UL users receive output directly from the offices to which their service requests are addressed, but a UM user always receives output from a coordinator capable of transferring blocked, fixed length record files to discs of the size and format specified in the service request. Given sufficient demand, the central repository and certain national or regional offices will probably develop capability for making the more frequently requested conversions. Similarly, a coordinator equipped with the current version of IGBADAT could also provide service to UL users.

It is apparent that the communication problems facing both the new managers and the potential users of IGBADAT are many and varied. With continuance of the good will, friendly cooperation and hard work contributed so generously by participants in Project 163, however, I believe their solution is simply a matter of time, and, as these things are measured, not very much time. The base is already both sufficiently large and accessible enough so that, with caution and good sense, we can begin to think seriously of using it in some aspects of research and teaching.

Figure 4: Proposed flow of requests (dashed lines) and retrievals (solid lines) to large computers (UL) or stand-alone microcomputers (UM).

What is Stored in IGBADAT2?

Because IGBADAT depends on volunteer contributions for its input, what is stored in the base is what has been offered. The resources that would make possible a more balanced coverage have simply not been available. Nevertheless, certain significant patterns are beginning to emerge from the data now accumulated.

Geographic Coverage: Of the 243 first level "sovereign entities" assigned symbols in the two-level U.S. Government (FIPS) place code adopted by Project 163, 73 are now represented in IGBADAT2. Summed over continents, large subcontinental regions, and oceans, the sample distribution is roughly as shown in Table 1. Because of the nature of our

data capture operation, balance (however this is defined) in the matter of areal distribution can only be anticipated in a very large base. The current distribution seems quite impressive, until one dismembers and examines the rather grandly named geographic groupings.

For example, 1,340 of the 1,558 descriptions of Atlantic Island specimens are from the Canary Islands. Similarly, of the 1,024 descriptions of specimens listed as from the Eastern Mediterranean and Asia Minor, 942 are from Turkey; of the 1,262 from the European mainland, 420 are from Etna and another 200 from elsewhere in Italy; of 457 from South America, 287 are Brazilian, and so forth. Most of the Pacific Island information is from Fiji and the southwest, and the enormous amounts of published data about the Hawaiian archipelago and French Polynesia have not yet found their way into IGBADAT.

Table 1: Geographic coverage of IGBADAT2 (number of samples).

Africa	472
Antarctica	152
Atlantic Islands	1558
Asian Mainland	820
Australia, New Zealand, Papua New Guinea, Indonesia	534
Central America, Caribbean Basin	176
Eastern Mediterranean, Asia Minor	1024
European Mainland	1262
Indian Ocean Islands	41
Japan	69
North America	2376
Ocean Floors (mostly DSDP)	595
Pacific Islands	325
South America	457
United Kingdom	223

It is curious, too, considering the relative intensity of petrological research in the two countries, that IGBADAT2 contains 630 descriptions of specimens from China and only 79 from Japan! Regrettably, IGBADAT2 as yet contains no information at all from the data-rich Slavic language literature of petrology. This kind of spottiness is probably inevitable in any decentralized and essentially volunteer operation of this kind. However, as the base grows and its language and literature coverage are broadened, these imbalances should lessen.

In the immediate future, large infusions of data are expected from India, Japan and the United Kingdom. The Indian contribution is contingent on the development of satisfactory arrangements for transfer of data from Project coding sheets to first machine-readable form - i.e., "data transfer" in exactly the sense used within Project 163. The Japanese and U.K. contributions hinge on completion by their national groups of promised translators for moving already digitized data from national to international IGBA format. Pending translation, the U.K. data in UK-IGBA format can be obtained directly from World Data Center A. Readers interested in the Japanese IGBA data should write to Prof. S. Aramaki (Earthquake Research Institute, University of Tokyo, Bunkyo-Ku, Tokyo 113, Japan).

Incidence of Rock Varieties by Name: Of the 414 rock names in the system glossary, only 240 occur in IGBADAT2; 351 specimens are recorded as unnamed in the source references and 317 as bearing source names not included in the IGBA glossary. The average incidence of IGBA-recognized rock names that actually appear in IGBADAT2 is 39.5, but only 43 are used to denote more than 25 specimens and only 18 suffice to denote nearly 2/3 of the compositionally-named

specimens in the file. The average incidence of the 197 names that occur less than 25 times each in the file is 12, and 34 of them occur just once! Development of a rock nomenclature that is both scientifically sound and psychologically sensible remains a major problem for petrology, and the time is not far off when the IGBA base may prove very useful in addressing it.

Meanwhile, the commonest names in IGBADAT2 are about what one would expect, though their relative frequencies are perhaps not. In order of decreasing frequency, the essentially compositional names used to denote more than 100 specimens whose descriptions appear in IGBADAT2 are shown in Table 2. There are also 102 lherzolites, nearly all described as nodules occurring in volcanic rocks. Of noncompositional names there are 353, the commonest being lava (184), ignimbrite (45) and pumice (43).

Table 2: Relative frequency of rock names in IGBADAT2.

Volcanic or Hypabyssal		Plutonic	
Basalt	2148	Gabbro	464
Tholeiite	332	Granite	452
Dolerite	313	Diorite	204
Trachyte	309	Syenite	182
Phonolite	268	Monzonite	151
Dacite	212		
Rhyolite	190		
Latite	157		
Diabase	152		
Hawaiite	149		
Andesite	133		
Trachybasalt	132		
Spilite	109		

There are two drastic anomalies in this display, the frequencies for andesite and granite. That IGBADAT2 contains so few andesites probably reflects rather extreme sampling bias in literature coverage, for in the literature concerning large areas of the globe for which petrographic data are abundant (such as the main islands of Japan or parts of the Moldanubian arcs) descriptions of andesite far outnumber those of basalt. The very low incidence of granite descriptions in IGBADAT2, however, is probably occasioned by a more fundamental difficulty. Each specimen description in the base must include some quantitative analytical data about bulk composition, and for older material this amounts to requiring a "complete essential oxide analysis." The literature abounds, however, with descriptions of granites that contain no direct chemical information at all, a combination of modal analysis and qualitative mineral assemblage data providing all the information about bulk composition that was considered necessary. Indeed, probably for the same reason, since bulk chemical analyses first began to be used in petrology their incidence has been far higher in descriptions of volcanic rocks than in those of plutonic ones.

The contents of IGBADAT2 suggest that in published petrographic descriptions bulk chemical analyses of volcanic rocks outnumber those of plutonic rocks by considerably more than a factor of 3. Among those including both bulk chemical and quantitative modal analyses, however, the relation is reversed, plutonic rocks outnumbering volcanic rocks by at least a factor of 2.

Age Distribution: Radiochemical age determinations of samples in IGBADAT2 are still very scarce, and for 3,756 (35%) of them no stratigraphic age is recorded in the base. It seems safe to assume that most of our contributors include

all stratigraphic system assignments encountered in the source references; this is certainly suggested by the fact that in over 3,500 descriptions reported, the stratigraphic age assignment is generalized, indicating either that the writer was not, or the reader cannot be sure that it applies to the specimen actually described. Absence of age information in the coded description almost certainly reflects a similar lack of such information in the source. Absence of an age assignment in the source reference may be occasioned by disinterest, complete lack of information, or unwillingness to waste print on information considered either widely known or essentially irrelevant.

Whatever the reason for its frequent omission, the incidence of citations of stratigraphic systems in the remaining 6,391 specimen descriptions is as follows: Cenozoic - 4158, Mesozoic - 833, Paleozoic - 392, and Precambrian - 1008.

Incidence of Petrographic and Mineral Assemblage Descriptors: The IGBA petrographic descriptor list is used in almost all descriptions, and contains, on average, 4.3 terms; 3,552 petrographic lists contain more than 4 but only 655 include more than 6. The mineral assemblage list is present in all but 454 of the specimen descriptions and contains, on average, 4.7 terms. More than 4 are used in 4,815 of the lists, and more than 6 used in 2,655.

It would, thus, appear that we are rapidly accumulating and organizing the kind of information that would facilitate authoritative and perceptive characterization of the interrelation between bulk composition, mineral assemblage and petrography, at a scale and with a precision previously far beyond reach. But one must not cheer too soon. The Project coding sheet makes provision for the contributor to record, as with age, his own or the source reference's judgement about whether these classes of descriptor are "generalized," that is that they either may not or do not apply to the material actually analyzed.

No fewer than 4,959 of the petrographic and 4,700 of the mineral assemblage lists are recorded as "generalized." Thus, in about half the descriptions, either it is clear that the petrographic and/or mineral assemblage descriptors do not apply to the analyzed specimen, or it is not clear that they do. The prevalence of this kind of over-generalized description seems one of the most important and disturbing discoveries growing out of the work of the Project. Those in whose work the relation between modal and bulk chemical composition is of major importance also may be somewhat jarred by the realization that not much more than one in every ten specimen descriptions in IGBADAT2 includes a modal analysis.

Incidence of "Essential Oxide" Components: For certain types of petrographic calculation a "complete" major analysis is required. Others cannot be carried out in the absence of values for some particular oxide or set of oxides. The project coding form makes provision for entry of 14 oxide components: SiO₂, TiO₂, Al₂O₃, Fe₂O₃, FeO, MnO, MgO, CaO, K₂O, Na₂O, P₂O₅, CO₂, H₂O⁺, H₂O⁻. A general idea of the level of "completeness" is given in Table 3, which shows that 8,161 of the analyses contain determinations of more than 10 oxides. Furthermore, 775 analyses lack determinations of both iron oxides and another 540 include only one. H₂O⁺ is lacking in 2,211 analyses, and H₂O⁻ in 4,988. No water determinations are given in 1964 analyses, and in another 3,271 the water is not partitioned.

The lack of Fe and H₂O partitions, like the truncation of results for major constituents at the one decimal (tenths) place in 1,808 examples, is often justified by rationalizations now so strongly held as to verge on dogma. Historically, however, all three begin in earnest only with the great wave of instrumentation and automation that transformed analytical chemistry after World War II. Long-established wet and gravimetric procedures were then replaced by much more convenient physical methods that generally allow neither routine determination of water nor partition of water or iron in routine analytical work.

The question of the minimum number of "essential oxide" determinations required to qualify a specimen description for inclusion in IGBA was discussed, and answered differently, several times during the life of Project 163. The current formal requirement is for no fewer than 5 trace and/or essential oxide components. In view of the heat of the 1984 discussion in which this particular decision was reached, it is amusing to recall that even then the file contained a couple of specimen descriptions in which determinations of oxide components were entirely lacking. In IGBADAT2 there are now 132 such descriptions, and in each of another 89 there are fewer than 5 essential oxide component determinations (Table 3).

Table 3: Number of analyses in IGBADAT2 lacking 0, 1, 2, . . . 14 oxide determinations.

0 - 2191	5 - 436	10 - 7
1 - 2770	6 - 267	11 - 31
2 - 1883	7 - 50	12 - 26
3 - 1317	8 - 74	13 - 25
4 - 923	9 - 9	14 - 132

Throughout the life of Project 163, the final decision about whether a specimen description should be included in the base depended on whether it duplicated one already there, and the *ad hoc* test of "duplicity" was a comparison of the amounts of the first four oxide components: silica, titania, alumina, ferric oxide. There are now 257 analyses, or 2.5% of those in the base, on which the test cannot be performed because at least one of the four is lacking! This percentage may be expected to rise in the future, and perhaps a more informative test of "duplicity" should now be devised.

Amount of Trace Element Information: Table 4 shows the frequencies in IGBADAT2 of descriptions containing 0, 1, 2, . . . trace element determinations. The zero class frequency is slightly reduced, and those of classes 1 to 3, especially the first, are slightly inflated by the long-standing project convention for recording "essential" oxide components reported in the source reference as "tr," "trace," or "not determined." These are stored as trace elements of amount zero - that is, less than the minimum reportable. Analyses in which it is necessary to record a single oxide component in this fashion are quite uncommon, and those in which more than one must be so recorded are vanishingly scarce.

Evidently, about half of the descriptions in IGBADAT2 contain an average of about 10 *bona fide* trace element determinations. Both the number of sample descriptions containing trace element analyses and the average number of determinations in them are bound to increase materially as the capture operation gradually shifts from retrospective to current data. The trace element list is often the largest single list in the sample description. As noted earlier, the average length of a specimen description in IGBADAT2 is about 330 bytes, and almost without exception, descriptions longer than this are rich in trace element data.

Source References

Well over 90% of the specimen descriptions in IGBADAT2 are drawn from widely circulated publications available in any major reference library. Some national groups have searched journals that do not enjoy wide international circulation - that is one of the advantages of having national groups - and two have also made some use of "gray" publications, documents available from issuing institutions on request but not otherwise publicly circulated.

From the outset, the data capture operation of Project 163 was both decentralized and voluntary. Except for two of its early "rock type" contributors, the U.S. group assumed its responsibility was restricted to, and included all relevant information in books or journals published in the United States. Some national groups, however, interpreted their mandates in precisely the opposite sense, feeling obligated to capture descriptions of all (but only) indigenous rocks wherever published. The central office interfered from time to time by attempting to arrange coverage of journals whose content and public are dominantly international. Confusion of this sort is probably inescapable in the early stages of any voluntary multinational literature scan of such magnitude.

In the countries in which most geological publishing is done, the simplest and most efficient way to organize the work is by publication, ignoring geographic and petrographic considerations. Where it seems desirable to organize the work geographically or by rock type, however, the tendency will be to scan any publication and extract from it only material satisfying the immediate geographic or petrographic search objectives. The only practical consequence of this lack of agreement about search responsibilities is to increase the risks of both incomplete coverage and duplication of effort. In fact, there were very few instances of unplanned duplication during the life of Project 163, and most were not occasioned by confusion or disagreement of this sort. However, if there had been twice as many active contributors, or they had built a base twice as large as IGBADAT2, the situation might well have been very different.

Table 4: Frequencies in IGBADAT2 of samples with 0, 1, 2, . . . trace element determinations.

0 - 4825	10 - 406	20 - 75
1 - 578	11 - 17	21 - 53
2 - 830	12 - 165	22 - 41
3 - 234	13 - 229	23 - 35
4 - 164	14 - 295	24 - 24
5 - 283	15 - 103	25 - 20
6 - 185	16 - 95	26 - 20
7 - 419	17 - 125	27 - 11
8 - 268	18 - 131	28 - 23
9 - 193	19 - 20	29 - 14
		>29- 22

In planning the coverage of the base, it was decided that specimen descriptions would be drawn only from primary sources. It is easy to imagine difficulties occasioned by this policy - analyses shown to be faulty and later corrected, for instance - but in fact they very rarely arose in practice. It was also decided that 1917, the publication date of the most comprehensive printed tabulation of chemical analyses of igneous rocks, would be used as a cutoff date. Older material was not to be excluded *ipso facto*, but there was to be no systematic attempt to capture it. Policy matters of this sort are of course subject to review by the new Subcommittee, which has already decided to make it possible for contributors to draw material from compendia if they choose.

Project 163 was initially commissioned to design and generate a *retrospective* base for igneous petrology, because it was thought that if attention was directed at once to the flood of modern publications, older material of substantive interest would almost certainly be slighted and might well be lost. The mandate for a primarily retrospective search was never unanimously endorsed by contributing members of the project. With vast amounts of the early literature still unscanned, IGBADAT2 already contains material drawn from as many sources published after as before 1970 (Table 5).

Requirements for the Future

Several requirements are clear. First and foremost, efforts to decentralize the data transfer operation must be accelerated and intensified. Failure of major progress toward this objective in the immediate future will place the life of the base in serious jeopardy.

At considerably lower priority, the coding form should be revised to eliminate mention of all matters that can be determined internally from the data. The transfer processor could easily be modified so that, before sending a completed specimen description to the storage file, it would determine and record, for instance, whether the analysis contains values

Table 5: Incidence of dates of source publications.

<1910	-	13	1950-59	-	69
1910-19	-	16	1960-69	-	204
1920-29	-	45	1970-79	-	422
1930-39	-	69	>1979	-	66
1940-49	-	45	Total	-	949

for both oxides of Fe and for H_2O^+ and H_2O^- . It could also detect the number of missing oxide determinations, the number of significant figures in the oxide determinations, the presence of a petrographic mode, and so forth. Marking of the status indicator symbols that denote these and similar "non-judgemental" characteristics of the description is a needless imposition on the contributor's time and patience.

A more systematic central office procedure for assigning, recording and reporting literature coverage will become essential as the file expands, and particularly as the number of contributors involved in data capture increases. It should be understood by all concerned that literature assignments are for a fixed term, subject to renewal if necessary and mutually agreeable. The gradual accumulation of incomplete assignments resulting from the lack of such prior agreement has been an intermittent source of confusion.

If the scheme proposed in 1983 by British participants in IGBA for capturing data immediately prior to its publication proves successful, it should be adopted, with suitable local modification, by other national groups. For the next few years, however, the emphasis on retrospective data capture should be continued throughout the Project, for it is not likely that a scan of similar scope will be attempted again in the near future. Computerization of the whole spectrum of data storage and retrieval activity in earth science is proceeding very rapidly, and it is quite possible that, for most practical purposes, information we fail to capture will be lost for a generation or more.

The major problem remains that data bases soon die if they are not continuously expanded and improved. Who will do the non-computational, non-executive, non-policy work of arranging and managing the periodic updating of IGBADAT? Where is this work to be done, and how will it be financed? Reasonably satisfactory arrangements, under the aegis of the Subcommittee, can probably be made for the next couple of years, but a base that is to serve as a major and continuing resource for the profession will ultimately require both a small, stably-funded office and a fair part of the time of a competent petrologist.

Acknowledgements

The first part of this article is based on decisions reached in, and what the writer took to be the sense of the discussion at the meetings in May, 1985, at Leicester, U.K., of IGCP

Project 239 and the IUGS Subcommittee on Data Bases for Petrology. The following, several of whom also later reviewed preliminary versions of the manuscript, participated in the meetings: A. Al-Mishwt (Kuwait), J. Brandle (Spain), R. Cristofolini (Italy), A. Duncan (U.K.), J. Durham (U.K.), D. Fiesinger (U.S.A.), J. Frizado (U.S.A.), R. Gomes (Portugal), J. Hiron (Brazil), M. LeBas (U.K.), T. LeBas (U.K.), N. Lefolley (U.K.), R. LeMaitre (Australia), N. Nishiwaki (Japan), G. Panto (Hungary), D. Stewart (U.S.A.), P. Sabine (U.K.), K.V. Subbarao (India) and C. Unan (Turkey).

Dr. Felix Chayes has been at the Geophysical Laboratory (2801 Upton St. N.W., Washington, D.C. 20008) "far too long and not nearly long enough." His major research has been in statistical petrography. A former president of the Mineralogical Society of America and 1984 Krumbein medalist of IAMG, he now heads both IGCP Project 239 and the IUGS Subcommittee on Data Bases for Petrology.

Appendix: Organizational Notes and Some Useful Addresses

The Subcommittee and Project 239 welcome active participation in their work by all interested petrologists. The Subcommittee will issue an occasional newsletter and, in collaboration with Project 239, will continue the circular series established by Project 163. Initially, newsletters and circulars will be sent to all persons whose names are on the mailing list inherited from Project 163; others who wish to receive these should notify the Secretary of the Subcommittee.

Orders for or inquiries about tape copies of IGBADAT2 should be directed to: Susan McLean, EDIS-NOAA, Boulder, CO 80302, U.S.A.

General inquiries about IGBA, IGBADAT or the Subcommittee should be directed to: Felix Chayes, Subcommittee Chairman, Geophysical Laboratory, 2801 Upton Street NW, Washington, D.C. 20008, U.S.A.; or Dion Stewart, Subcommittee Secretary, Department of Geology, Bowling Green State University, Bowling Green, OH 43403, U.S.A.

Inquiries about the proposed library of user-oriented subprograms for exploiting IGBADAT should be directed to: Roger Le Maitre, Chairman, Working Group C, Geology Department, University of Melbourne, Parkville 3052, Victoria, Australia.

Inquiries about current and potential applications of the microcomputer in building and using IGBADAT should be directed to: Joseph Frizado, Chairman, Working Group B, Department of Geology, Bowling Green State University, Bowling Green, Ohio 43403, U.S.A.