

全球沉积岩数据库 (SEDBA) 及其在中国的发展^①

南君亚 叶健骝 杨卫东

(中国科学院地球化学研究所)

一、前言

全球沉积岩数据库(简称 SEDBA)是在日本奈良大学 Niichi Nishiwaki-Nakajime 教授领导下,於 1988 年开始执行的 IGCP (International Geological Correlation Programme) 第 269 项研究项目。

建立 SEDBA 的目的是储存、检索和使用沉积岩数据,实现沉积岩数据的国际共享。

在 IGCP 269 项目执行期间(1988—1992),将完成 SEDBA 原型数据库的设计和建立,包括数据表格式的设计、数据表使用说明书的编写和出版、数据代码及词表的制定和编辑、数据结构的确定、应用数据库技术在计算机上实现对 SEDBA 数据库文件的管理等。

迄今为止,已有 22 个国家及地区的 60 余位学者参与了 SEDBA 的研究和设计工作,SEDBA 原型数据库的建设工作已基本完成。

众所周知,沉积岩是地壳的重要组成部分,它占地壳出露面积的 75%左右,沉积岩不仅记载了地球演化的历史,而且世界上半数以上的金属矿产、80%的非金属矿产、几乎 100%的能源和地下水资源与其密切相关。因此,SEDBA 的建立,不仅有助于沉积岩数据的国际交流,还可以从中开发出许多重大研究课题,对开展全球地质对比研究、探索岩石圈演化和矿产资源及能源形成规律、研究古气候古生态的变化以及灾害事件对地质发展的影响等,均具有重要意义。此外,地球物理学家、石油地质学家、水文和工程地质学家可从中获得沉积岩物理和工程特性参数;化工、冶金、陶瓷及建筑部门的工程技术人员可从中查询到作为原材料的沉积岩其组分及性能参数。因此,SEDBA 将具有广阔的应用前景。

二、SEDBA 的管理机构

SEDBA 是一个国际性的数据库系统,由一个 SEDBA 中心 (SEDBA Central Office) 和 11 个 SEDBA 地区分中心 (SEDBA Regional Office) 组成完整的 SEDBA 管理机构。

① 国家自然科学基金资助项目。

SEDDBA 中心将设在日本, 负责汇集 11 个地区分中心提供的数据, 按期编辑和制作 SEDDBA 数据文件, 分发给各地分中心提供使用, 同时交一份给世界数据中心。随着新数据的补充而不断更新数据文件版本。

按以下地区, 即: 东亚、中国、南亚、中东、西欧、东欧、独联体、北美、南美、非洲、澳洲, 建立 11 个地区分中心。各分中心负责收集本地区数据, 编辑和制作 SEDDBA 数据文件, 提交给 SEDDBA 中心, 并向本地区用户提供数据的查询、检索服务。

为避免引起混乱, 各地区中心只能通过 SEDDBA 中心获得其它地区的数据, 不能直接与其它地区中心交换数据。

SEDDBA 的组织机构如图 1 所示。

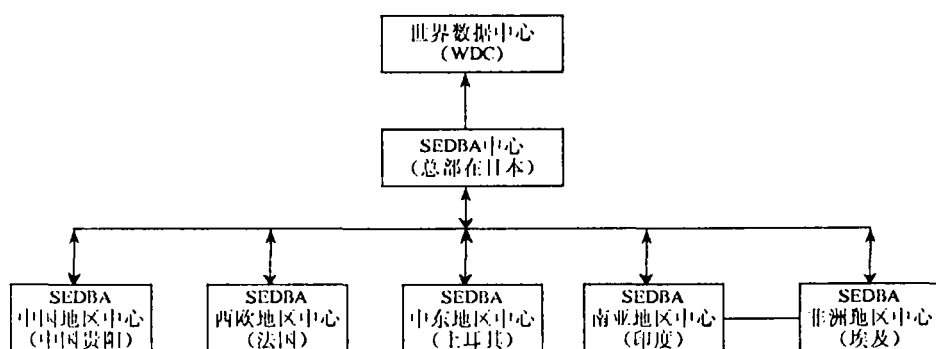


图1 SEDBA的组织机构

三、SEDDBA 的内容及数据的组织方式

SEDDBA 存贮除第四纪松散沉积物 (如海滩或湖滩砂, 深海淤泥风成砂等) 外各地质时代的陆源碎屑岩、火山碎屑岩 (即火山成因的碎屑经水或空气搬运、沉积的岩石)、碳酸盐岩 (岩浆成因的除外)、蒸发岩、硅质岩、磷酸盐岩、含金属的 (如铁质、锰质、铝质) 沉积岩、煤及褐煤等沉积岩的产地、产状、地质特征及分析测试数据。

在 SEDDBA 中, 以岩石的“样品” (Sample) 作为信息储存的基本单元。所谓“样品”, 通常是指用于研究的手标本或若干相同岩石手标本的组合。也包括一些组合样, 例如从钻孔岩芯上按一定间隔截取的样品组合。每个样品除数据提供者给予的原始样号外, SEDDBA 系统将赋予它一个标识符 (SEDDBA-ID), 作为能唯一标识样品的实体键。

数据提供者可将一系列地质上相关的样品组合成一个样品“集” (SET)。例如, 来自同一地质剖面、钻孔、采场或同一地质时代岩层的一系列样品, 均可定义为一个“集”。数据提供者需说明“集”的定义及名称。SEDDBA 系统将赋予样品集统一的标识符 (SET-ID), 作为样品集的实体键。

SEDDBA 中数据按三个层次分别安排在一组数据表中 (如图 2 所示)。

第一层由以下几个表组成, 分别记录样品集 (SET) 和样品 (SAMPLE) 的一般信息, 文献及图件资料信息。

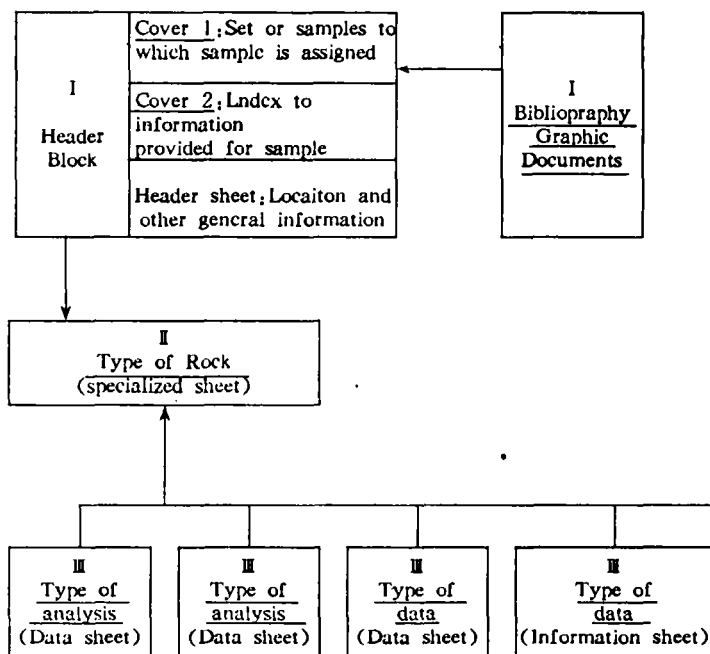


图2 SEDBA数据的层次结构

1. **COVER1** 包括样品集的标识符、定义和名称, 数据提供者和收集者的姓名、工作单位和地址, 包含的样品、文献和图件的数量等。

一份 COVER1 表对应着一个样品集。

2. COVER2 记录与样品有关的基本信息，包括样品的原样号，岩石类别，所属样品集的名称，分析数据表的数量以及一份用以登记分析项目、分析物名称和分析数据表索引号的表格。

COVER2 主要用于对样品提供的信息进行索引。一份 COVER2 对应着一个样品。

3. **HEADER SHEET** 存放对各类沉积岩都共同的数据内容。如采样地点、样品的保存情况、年龄、颜色、所属地层单元、地区的名称、岩石名称、采样点周围岩石的地质特征、样品的代表性、样品所在层位岩石的经济价值、沉积作用和沉积环境、大地构造位置、古地理环境以及后生变化等。一份 **HEADER SHEET** 表对应一个样品。

设置 HEADER SHEET 的目的,是把各岩类的共同数据内容存放在一个数据文件中,便于查询宏观的信息。

4. **BIBLIOGRAPHY SHEET** 用于建立 SEDBA 专用的文献数据库, 存放与样品有关的文献信息。其中包括关键词、作者、标题、出版情况 (公开出版或未出版)。如属杂志或期刊, 提供其名称、年、卷、期、页等; 如属专题报告或专著, 提供其名称以及编辑、出版的有关信息。

一份 BIBLIOGRAPHY SHEET 对应于一篇文献。

5. COLLECTIVE SHEET FOR GRAPHIC DOCUMENTS 用于登记与样品集有关的图件其类型(如航空照片、素描图、地质图等)、对象(如露头、采场等)、作者、图件的色调(黑白或彩色)、图件的表现形式(图纸、胶片、磁带等)、大小等信息。至於各图件的图形数据因受计算机内存空间的限制,目前暂不考虑进入 SEDBA 系统。

第二层由八种专用数据表 (SPECIALIZED DATA SHEET) 组成, 分别存放陆源碎屑岩、火山碎屑岩、碳酸盐岩、蒸发岩、硅质岩、磷酸盐岩、含金属的沉积岩、煤和褐煤等沉积岩的一般地质信息, 是对 HEADER SHEET 表的补充。由于上述各类特征各异, 专用数据表的数据项不完全相同, 但是主要包括以下三个内容:

1. 样品局部地质环境的特征 主要的区域性岩石类型和岩石的组合特征。

对煤及褐煤来说, 这部分将记载煤岩系的特征。

2. 样品所在地质单元的特征 主要记录样品所在地质单元的岩石名称、单元厚度、横向变化、均一性、沉积构造、上覆和下伏地层的岩性特征、古流向等信息。

对于含煤、磷、铁、锰、铝等资源的沉积岩, 这部分主要用于记录含矿层的地质特征

3. 样品的地质特征 主要记录由样品反映的岩石地质特征、岩石的结构构造、基质的成分及含量、成岩作用等。岩石的组分分析数据放在专门的数据表中。

火山碎屑岩是一种特殊的碎屑岩, 除物质来源不同外, 其它特征与陆源碎屑岩相同。因此, 火山碎屑岩专用数据表仅记录火山碎屑岩的分类及组分的成因类型, 其它地质特征按陆源碎屑岩专用数据表的格式记录。

第三层由 14 种数据组成, 分别记录样品的全岩或其组成部分 (如碎屑、胶结物、单矿物等) 的化学分析数据 (包括主要元素、微量元素、稀土元素)、碳质物质的特别数据、碳氢化合物存象 (Occurrence) 信息、元素的电子探针分析数据、同位素分析数据、化石及孢粉鉴定数据、沉积物粒度测量数据、沉积物中各种液体的分析数据及液体包裹体测温数据、模式分析和组分分析数据、重矿物鉴定数据、X 射线衍射分析数据、沉积岩物理及工程特性数据、图件信息、采样过程的说明等。

碳质物质数据表是用于记录煤岩样品的一系列特殊参数, 如煤显微组分、镜煤反射率等。但其它沉积岩中若有碳质夹层, 其有关参数也可记录在此表内。

碳氢化合物存象数据表用于记录沉积岩中石油、天然气等有机质的赋存状态、成熟度、埋藏史等信息。因此, 这既是一种有机质分析测试数据表, 也是一种有关油气储积层的专用数据表, 对记录油气方面的信息有特殊用途。

图件信息数据表的内容与第一层的图件信息表类似。主要用于登记与样品有关的图件资料。

第三层是按分析测试内容组织数据的, 因而不同岩类的同种分析数据均可放在同一种数据表中。在分析数据表中, 以分析样 (Specimen) 为基本的记录单元。一个样品 (Sample) 可能分成若干个分析样 (如全岩、胶结物、碎屑、长石等)。因此在一种数据表中可以出现一个以上某个样品的记录, 如果该样品未作这方面的分析, 该数据表不会有该样品的记录。

在数据表中, 用样品的 SEDBA-ID 加上数据标记号 DR[#] (Data Reference Number) 来标识一个分析样品。DR[#] 由前后两部分组成, 前半部为数据表名称的缩写 (如 CA 表示化学分析表); 后半部为顺序号。DR[#] 作为索引项, 用来对一个样品的分析数据进行检索, 从而将数据有机地组织起来。

综上所述, SEDBA 数据是按书的结构进行组织的 (如图 3 所示), 一本“书”代表一个样品集。COVER1 类似“书”的封面, 包含样品集的名称和标识符; COVER2 是“书”中某一“章” (即一个样品) 的“目录”; HEADER SHEET 相当于该“章”的“概况”; SPECIALIZED

DATA SHEET 则为该“章”中地质特征部分; 若干数据表好象各“节”, 分别论述样品的各方面特征。

BIBLIOGRAPHY SHEET 和 COLECTIVE SHEET FOR GRAPHIC DOCUMENTS 为该“书”的文献部分。

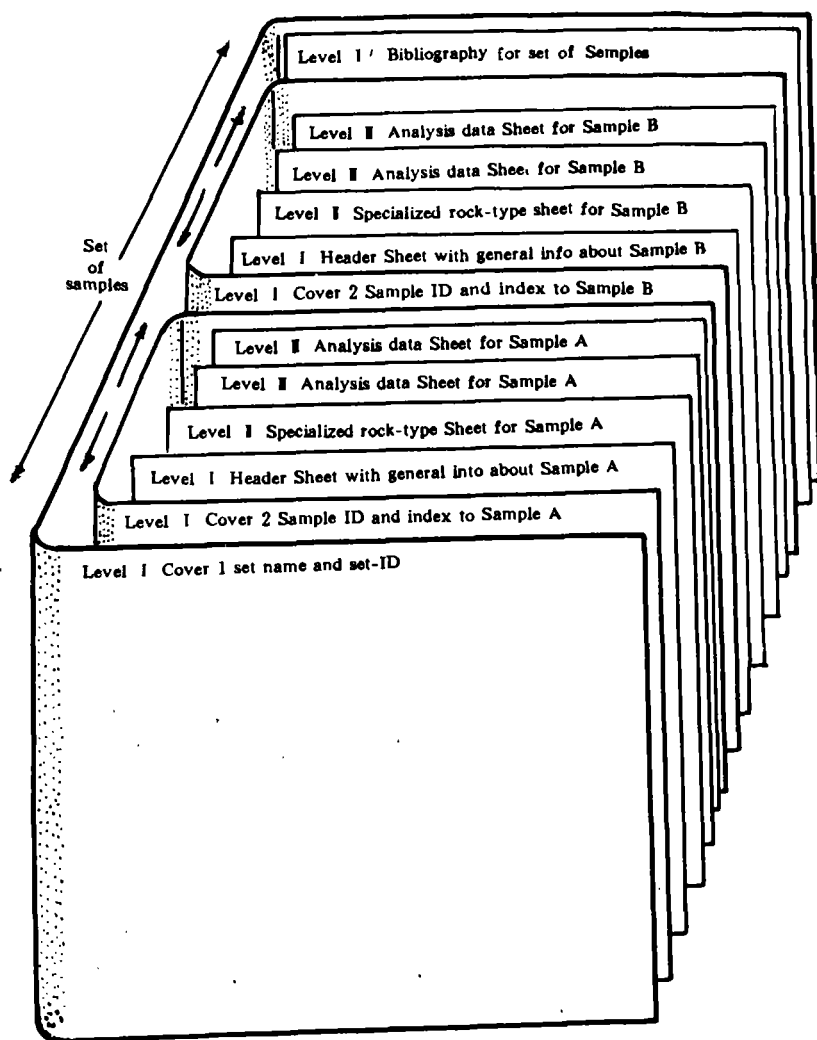


图3 SEDDB 数据组织方式

四、SEDDB 的数据标准化

在 SEDDB 中, 主要由数值型的定量描述信息和字符型的定性描述信息组成。定量描述信息, 如元素分析值, 粒度测定值等, 通常是以比值或用某种度量单位表示的测定值。在 SEDDB 数据表或数据表说明书中, 对各种定量性数据项使用的单位均做了说明。如果原始数据的单位与 SEDDB 规定的不同, 可在注释项中予以说明。

定性描述的数据分成两种, 一种是用文字描述实体属性的信息, 如岩石名称, 结构构造等; 另一种是注释性的信息。前一种数据是 SEDDB 中的重要部分, 它们主要用专业术语

表示。由于与沉积岩有关的术语在各国甚至同一国家内都不完全相同,为避免出现“同词异义”或“同义异词”而造成混乱,建立标准词表统一概念是十分必要的。目前 IGCP 269 项目暂定使用由中国工作组编辑的 SEDBA 荐用词表 (LIST OF RECOMMENDED TERMS FOR SEDBA) 作为标准。

SEDBA 荐用词表是参考了中国国家标准局颁发的中华人民共和国国家标准 GB 9649—88《地质矿产术语分类代码》,按沉积岩颜色、沉积岩组分(包括矿物组分和结构组分)、沉积结构、沉积构造、沉积物物理性质、沉积岩类型、沉积环境、沉积作用和成岩作用、古生物与古生态、次生变化等十个方面编辑而成。

该词表共收词 2400 余条。每一词条均包括“中国国家标准码”、“中文名称”、“英文名称”和“SEDBA 代码”四个栏目。词表还附录了国际 FIPS 系统国家(或地区)代码及中国各省、自治区及直辖市代码。经过试用后,将不断修改、补充、使其逐步完善。

SEDBA 的描述性字段均使用英文表示,因此 SEDBA 荐用词表对今后汉化 SEDBA 系统有特殊用途。

SEDBA 中的某些特殊字段使用了代码,这些代码分别列在数据表及说明书中。

此外,SEDBA 的各种数据表都用缩写词表示。例如,CL 表示碎屑岩专用表,CA 表示化学分析数据表,HS 表示 HEADER SHEET 表,BT 表示文献信息表等。一些数据字段用缩写词表示填写的方式。例如(C)表示该字段需使用代码;(M)表示该字段必须填写;(R)表示需使用荐词表中的标准词等。

五、SEDBA 数据转换格式

为使 SEDBA 数据的国际间交流不受计算机型号及数据库管理系统(DBMS)的限制,从而便于在各国推广使用,IGCP 269 项目决定采用压缩的 ASCII 文件格式作为 SEDBA 数据转换的标准格式。在还不具备通讯网络条件的情况下,决定以磁盘或磁带作为交换数据的介质。

目前许多通用的 DBMS 都具有将该系统的数据库文件自动转换成 Delimited ASCII 文件以及将后者转换成数据库文件的功能。因此,只要 SEDBA 系统中各数据库文件的数据结构被确定,数据交换就极容易实现。

六、SEDBA 的运行环境

在 SEDBA 中以样品为基本逻辑单元,即岩石样品作为实体。与样品有关的属性多达千余个,分别存放在不同的数据表中。在 SEDBA 的数据结构上,存在着一对一,一对多和多对多的复杂关系。如一个样品集对应多个样品;一个样品对应多个分析物;一篇文献涉及多个样品;一个样品集中包含多篇文献等。只有使用关系型数据库管理系统,才能实现对 SEDBA 系统的管理。

由于数据转换格式采用压缩的 ASCII 文件格式,对 SEDBA 的运行环境提供了相当自由的选择。SEDBA 既可选用微机,也可使用较大型的计算机或主机(Mainframe);既可应用各种通用的数据库管理系统,也能自行编制管理软件。但必须保证转换出的数据库文件其结

构与 SEDBA 标准文件绝对一致, 否则无法实现数据交流。

目前 IBM-PC 系列计算机在各国普及程度很高, 因此本项目选用这类微机作为建立 SEDBA 原型数据库的计算机选型, 并对该类计算机的基本配置作出如下规定:

- (1) 具有 $\geq 640\text{Kb}$ RAM;
- (2) 必需有至少 5Mb 空间的硬盘;
- (3) 至少有一个软盘驱动器;
- (4) 使用 ≥ 3.0 版的 MSDOS 系统。

七、SEDBA 的数据来源

建立 SEDBA 的最终目的是存贮、管理全世界各地的沉积岩数据, 为所有需要这些数据的用户服务。因此, 凡是沉积岩的数据, 不管是出现在公开出版的书籍、专著、杂志、期刊中, 还是未出版的报告、资料或保存在个人手上的, 都是 SEDBA 收集的对象。任何人都可以向 SEDBA 提供数据, 也能向 SEDBA 索取数据。

在 SEDBA 中存贮的都是沉积岩的原始数据, 因此, 在收集文献资料中的数据时, 需要请文章作者提供原始数据。

在 SEDBA 发展的初期, 数据主要来源于 IGCP 269 项目的成员。计划在 1992 年底前收集大约 5000 个样品的数据, 用于检验 SEDBA 系统的可行性, 以便进一步改善和提高系统的性能。与此同时, 通过各种形式, 包括在国内外期刊上发表文章, 举办培训班等, 宣传介绍 SEDBA, 使更多的专业人员了解和掌握 SEDBA 数据的收集和使用方法, 不断扩大数据来源和用户队伍。IGCP 269 项目准备于 1992 年年底在印度加尔各答举办一个 SEDBA 培训班 (Training Course), 吸收各国沉积岩专业人员参加, 就是实现这一目标的措施之一。

八、SEDBA 数据的提交和获取方法

从数据提供者收集、提交数据, 到地区中心编辑, 散发数据的过程如图 4 所示。

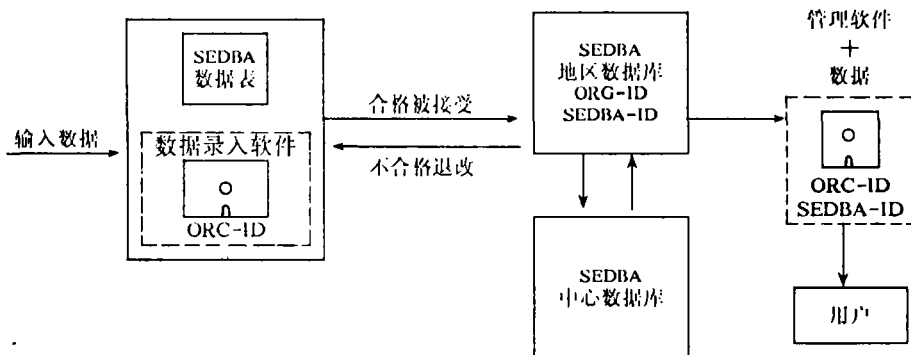


图4 SEDBA工作框图

首先, 数据提供者按照 SEDBA 数据表的内容和格式, 填写数据表, 并用 SEDBA 数据录入软件将数据输入到各数据文件中。每个样品及样品集均需有原始样号 (ORG-ID) 和样品

集的定义及名称。

数据文件提交给 SEDBA 地区分中心后,由该中心的管理人员检查数据的合格性。如果不合格,退给数据提供者进行修改;如果合格,由地区 SEDBA 中心管理人员统一编辑,赋予各样品及样品集 SEDBA-ID 和 SET-ID,制作数据库文件。每当 SEDBA 中心完成对各地区分中心提交的数据的编辑工作后,地区分中心负责向各用户提供带有数据库管理软件的数据盘。

数据录入软件及管理软件是用某种数据库管理系统(DBMS)提供的编程工具制作的。目前中国工作组采用 FOX BASE+2.1,法国工作使用 PARADOX3.0 制作的数据录入软件和管理软件,被 IGCP 269 项目推荐给各地区中心使用。各地区中心可根据自己的条件、需要,选择其它类型的 DBMS 建立相应的软件系统。

由于国际上对计算机软件的版权有专门的法律予以保护,为使数据输入及管理软件的扩散不触犯法律,这两套软件均使用专门的数据库管理系统版本——RUNTIME,重新编译而成。因此,这些软件仅能用于 SEDBA,不能更改 SEDBA 各数据库的结构,也不能用其建立其它数据库。

在 SEDBA 管理系统中,为用户提供了帐号和口令,这是保障数据库的安全和保密的必要措施。

九、SEDBA 在中国的发展情况

中国幅员辽阔,沉积岩及沉积矿产资源、能源、水资源极其丰富,另外在中国从事地球科学研究的部门也很多。因此全球沉积岩数据库在中国有极为广阔的应用前景,并且也有丰富的数据来源。

IGCP 269 项目开始执行以来,就引起国内各部门专业人员的浓厚兴趣和关注。参加了 IGCP 269 项目中国组以及关心 SEDBA 建设的人员与日俱增。目前中国组成员已有 30 余位,分别来自地矿部、煤炭部、中科院和高等院校所属的 12 个单位。本项目也得到国家自然科学基金委的经费及各方面的支持。

中国组在 SEDBA 设计和建立的各个过程中,均提交了研究成果和意见建议,为 SEDBA 的发展作出了很大的贡献。例如,在中国第一次工作会议上,设计了 SEDBA 数据表格式。之后,完成了 SEDBA 数据表说明书及荐用语表的编写工作,并用 FOXBASE+2.1 在 IBM-PC 系列微机上建立了 SEDBA 原型数据库等。这些成果为 SEDBA 的发展奠定了基础。

如何推广 SEDBA 的应用,并不断获取新的数据来源,是今后发展 SEDBA 的关键。为此,我们打算采取以下步骤实现这个目标。

1. 宣传和介绍 SEDBA

通过在各种专业性期刊或学术会议上发表文章或论文,详细而系统地介绍 SEDBA 的内容和功能,使全国各部门的地质专业人员了解 SEDBA,是宣传、介绍 SEDBA 的重要方法之一。我们曾先后在一些期刊上发表过介绍 SEDBA 进展的文章。今后准备继续利用全国性的学术会议,扩大宣传效果。

2. 举办 SEDBA 培训班

举办 SEDBA 培训班,系统讲授 SEDBA 的数据收集方法和使用方法,培训专业人员掌

握收集数据的规范和使用 SEDBA 的方法, 是确保数据来源和推广 SEDBA 实际应用的有效措施。

3. 争取各数据管理部门的支持

为使 SEDBA 在中国持续发展, 必须得到国内有关的数据管理部门支持。因为 SEDBA 是一个国际性的数据系统, 在国内如果仅靠少数专业人员用有限的课题研究经费来维持系统, 不仅因经费不足而难以持续下去, 而且无法获得稳定和持久的数据来源。只有纳入国家有关的数据管理系统中, 才能不断发展并在国内发挥最大的经济和社会效益。

4. 争取数据来源

如何鼓励专业人员提供数据, 是一项政策性很强的问题。我们准备通过以下途径为 SEDBA 争取数据来源:

(1) 与某些重大科研项目或生产任务结合, 承担其数据管理和应用的任务, 从而使该课题的数据系统而完整地进入 SEDBA 系统中;

(2) 与包括 IGCP 269 项目成员在内的专业人员建立协作关系, 以他们对 SEDBA 建设的贡献大小, 决定他们共享 SEDBA 的程度。以此来鼓励专业人员积极提供数据;

(3) 与国内有关期刊或出版物建立协作关系, 请编辑部制定规章, 要求文章作者必须按 SEDBA 数据格式提交其原始数据, 否则文章不予刊载;

(4) 对那些需要数据又未提供数据的用户, 采取有偿服务的办法, 可利用数据检索费来维护系统及购买新数据。

以上尚属初步设想。需要与 IGCP 269 项目中国组成员讨论并广泛征求全国有关专业人员的意见, 形成一个切实可行的发展计划予以实现。

十、结束语

全球沉积岩数据库的发展计划得到联合国教科文组织 (UNESCO)、国际地质对比计划 (IGCP) 和国际地科联 (IUGS) 的经费支持。在国内, 得到国家自然科学基金委员会、第三世界科学院和中科院地化所的经费资助。

经过国内外成员三年的共同努力, SEDBA 的原型数据库已基本建成。基中, 中国工作组全体成员以及关心本项目的专业人员为此作出了积极的贡献。

为使 SEDBA 成为我国地质工作者从事科研、生产、教学的有力工具, 需要全国地质人员共同关心和支持 SEDBA 的建设。积极提供数据, 并把它应用到自己的工作中, 结出丰硕的成果。

作者谨向为 IGCP 269 项目提供过支持、作出过贡献的单位和个人表示诚挚的谢意, 并欢迎对文中错误和不当之处提出批评。欢迎广大地质工作者对如何在我国展开 SEDBA 应用的问题提出意见和建议