

专论与综述

环境管理信息系统的研究现状和发展趋势

袁 进 春

(中国环境科学研究院)

随着计算机科学的发展,信息系统日益成为独立的研究部门。由于出发点及着眼点的不同,人们用几个近义词或同义词来称呼研究对象,如数据库(data base)、数据库(也译数据银行 data bank)、数据网络(data network)、数据系统(data system)、信息系统(information system)、信息网络(information network)、专家系统(expert system)等等,其中以数据库(data base)、信息系统和专家系统几个词的使用最广泛。计算机科学家们习惯于在较狭义的范围内使用数据库一词。专家系统的特点是强调系统本身的功能和智能化。考虑到我们对环境问题研究的目前水平尚达不到如医生看病那种境界,我们建议把我们正在着手研究的领域称为环境管理信息系统。

一、信息系统及其发展

所谓系统就是相互联系的单元构成的有机整体,这些单元称为系统的子系统,这种联系通常表现为物质流或信息流,系统与外界的联系称为系统的输入与输出。所谓信息系统就是以信息为输入与输出的系统。

1. 早期的信息系统——人工管理

所谓管理,实质上就是了解情况、作出决策。了解情况就是收集信息,建立信息系统。

在信息系统的人工管理阶段,这种管理信息系统多是建在决策者的脑海里。只有较大的信息系统才建成看得见摸得着的信息系统——档案。其信息的加工者是人,在有些情况下人还充当信息的存贮介质。人有生老病死,这种信息系统的缺点是显而易见的。

2. 计算机涉足信息系统——文件系统

计算机的出现使信息加工速度成千上万倍地加快,这是落后的靠人工管理的信息系统无法比拟的,于是导致了文件系统的产生。文件系统是专门的数据管理软件,是操作系统的一个子模块,允许对数据进行查询、修改、删除等操作。由于程序和数据间由特殊的软件联系,所以数据与程序之间具有独立性。

文件系统尽管克服了人工管理的种种缺陷,多少缓和了信息加工的高速度与信息系统管理的低效率之间的矛盾,但文件本身是对应于一个或多个应用程序的,从逻辑结构上看它与人工管理的信息系统没有多大差别。尽管程序不必与数据文件打交道而由专门软件联系,但这种信息系统是一个不具弹性的无结构的信息集合,有冗余度大、空间浪费、不易扩充、修改费时等缺点。

3. 信息系统的新形式——数据库

数据库是可共享数据以一定方式组织的

集合,数据库具有较高的数据与程序独立性,数据最小冗余度和完整性。数据库是通过一个叫做数据库管理系统的专门软件实现上述功能的。数据的结构化是数据库区别于文件系统的主要特征之一。正是由于结构化,数据库与文件系统相比大大地缩小了冗余度。依据数据结构可将数据库分为层状、网状和关系三种类型。

与传统的“数据库”如数据手册、文献索引等相比,现代数据库的优点更是显而易见的。首先,数据手册的编辑耗费巨大的人力、物力,而且数据每天都在更新,所以可能使手册中的某些数据在手册行将完稿时,已过时或被证明是错误的。而学术专著、文献索引的出版周期过长也往往使其学术价值减低。其次,数据手册、文献索引的功能是提供用户查询,以便获得有用的数据或文献,而这被经验证明是人所不能胜任的,即或勉强为之,也是十分艰辛费时的工作。最后,每个数据手册的编辑者都必须在如下两个极端之间取折衷:要么为篇幅计,精收各种数据,然而这会被证明是不便使用的;要么为方便用户计,广收各种用户可能用到的数据,这不仅增加了工作量,而且最终也会被证明是太复杂、不便使用的。

作为一种新型的数据手册、新型文摘的现代数据库,克服了传统“数据库”的缺点。首先,数据库可以随时修改、更新,所以当发现过时或错误的数据时,可以即刻修正,特别是在大型网络中,只要中心的数据库管理员做一次修改,网络中各个终端前的用户就在顷刻间人手一册“新版的数据手册”。对于文摘则可一边输入一边使用。其次,使人望而生畏的大量数据、文摘的比较、分类,如在计算机上做真是易如反掌。最后,计算机可以精收各种原始数据,而将可导出数据由程序产生,从而巧妙地解决了数据手册编辑者所面临的难题。另外,一般检索数据的目的是为了参加某种运算,而许多场合下运算又是由

机器完成的。显然,数据存贮、检索与加工相结合是更合理的,特别是在数据库支持下进行数据间关系的研究更是得天独厚,如物质理化性质与毒性关系的研究等。

二、环境管理信息系统 (EMIS) 的研究现状

过去的十多年中学术界在 EMIS 方面做了大量工作,仅 ESA-IRS 系统中 Pollution Abstract 文档到 1985 年 11 月 28 日为止,就收集有关的研究文献 805 篇,其中由计算机支持或与计算机有关的共 192 篇。国外在 EMIS 方面有代表性的工作是:美国环保局的一系列工作、欧洲共同体的 ECDIN 和联合国环境规划署全球环境监测系统的 GRID;国内除做了一些引进工作外,中国科学院环境化学研究所完成了中文环境文献库、中国环境监测总站完成了监测数据库——EIMS。

1. 美国环保局的 CIS 及其它

美国环保局在 EMIS 方面开展了大量的工作,CIS(Chemical Information System)^[1]是环保局与 NIH (National Institute of Health) 联合资助的系统,该系统设计的最低目标是辅助化学分析、鉴定,最高目标是研究化学物质结构与物理、化学性质的关系。该系统由如下四个逻辑子系统组成:(1)数据库;(2)数据分析软件;(3)结构和名称检索系统;(4)功能划分与参见。

美国环保局在 CIS 的开发中采取了灵活变通的方法。有的数据库是自建的;有的数据库是从别的团体租借的;有的数据库是与别的团体联合而由后者在 CIS 中运行的,比如它与粉末衍射联合委员会合作,由该委员会在 CIS 中运行它们自己的数据库,如此等等。

CIS 的开发通常是这样的:将准备好的表格输入 NIH 的 IBM-168 中,而后转入 PDP-10 分时系统内,最后编制程序产生可检索数据文件,当这些工作完成后,就产生了一

个“初版”的 CIS 数据库。“初版”的数据库首先在一定范围内免费提供使用以便检查错误,依据数据库的规模和复杂性,这种验收阶段可以长达 18 个月。经过验收后的数据库便可并入 CIS,转移到商用 PDP-10 上向科技界和政府机构提供服务。从此政府不再向它提供经济上的资助。

除 CIS 外,美国环保局还开发了如下几个系统:

(1) EADS (Environmental Assessment Data Systems) 系统,该系统的设计目标是辅助环境评价、污染源表征及控制技术的发展,它分为 GEDS(Gaseous Effluents Data System)^[2] 和 LEDS (Liquid Effluents Data System)^[3] 等系统;

(2) EFDB (Environmental Fate Data Base)^[4] 系统,该系统主要研究化学物进入环境后的归宿,即迁移和转化;

(3) HEEDA(Health and Environmental Effects Data Analysis System)^[5] 系统,这个系统的设计目标是提供结构与毒性关系研究的基础,并最终从化学结构预测毒性。

此外,还有 AIRS (Aerometric Information Reporting System)^[6]、PHYTOTOX^[7] 等等。

2. 欧洲共同体与 ECDIN^[8-10]

作为欧洲共同体环境政策制定的基础,共同体开展了一个环境研究计划,该计划主要由四个部分组成,其中之一便是环境信息的管理。因此,1973 年共同体便开始了 EMIS 方面的实验研究工作,实验阶段有如下八个目标:

- (1) 确定数据的有效性;
- (2) 依据数据有效性,简化数据结构;
- (3) 评价数据收集费用;
- (4) 建立输入格式;
- (5) 确定数据源及替代数据源;
- (6) 验证计算机系统;
- (7) 测试用户反应;

(8) 为 ECDIN (Environmental Chemicals Data and Information Network) 设计计算机软件系统。

在 ECDIN 的开发中,遵循了以下六项基本原则:

(1) ECDIN 收集一切以相当大量生产的化学物质的有关信息,而无论其毒性及有害性;

(2) ECDIN 所谈的数据仅限于“硬”数据;

(3) ECDIN 主要面向化学数据;

(4) 系统的基础语言是英语;

(5) ECDIN 将从一个分时系统上向用户服务;

(6) ECDIN 将联成从其中意大利的伊斯普拉 (Ispra) 到其各成员国中至少一个地方的网络。

ECDIN 的数据分十类:

(1) 化学物质特征数据(化学名、俗名、编号等);

(2) 结构编码信息;

(3) 物理化学性质;

(4) 化学分析方法及有关数据;

(5) 产量及贸易;

(6) 运输、包装、装卸、存贮和危险性;

(7) 使用和排放;

(8) 在环境中的迁移和转化;

(9) 对环境的影响;

(10) 法规数据。

ECDIN 最初是由数据库管理系统 SIMAS 支持的,随着 ECDIN 的发展和 ECDIN 功能的完善, SIMAS 逐渐无法满足 ECDIN 的要求,为此 ECDIN 的研制者们只得选择别的数据库管理系统。市场上通行的数据管理系统有好几十个,要做出一个恰当的选择绝非易事,为此他们例举了一些指标做为选择的依据和标准。这些指标分为两类:一类是要求系统必备的特性;另一类是希望系统具有的功能。用这两类考核指标对

初选合格的几个数据库管理系统打分,最后他们决定选用 ADABAS 代替 SIMAS.

3. 联合国环境规划署的 GRID^[11]

GRID (Global Resource Information Database) 是全球资源信息数据库一词的英文缩写,它的设计目标是:为那些重大决策的制定者们提供信息. 这里所谓“重大决策”是那些有可能影响全球的决策. GRID 是由 GEMS (Global Environmental Monitoring System) 系统中的一个工作小组设计的, GEMS 的任务之一是协调全球环境的监测,人们很快就认识到建立一个收集全球数据的信息系统是十分必要的,其结果就产生了 GRID, GEMS 试图把 GRID 建成联合国的环境数据管理系统.

GRID 的目标是:首先,为更广泛地为用户提供使用 GEMS 监测网络中数据库的机会;其次,将建成一个分时系统,它最终能够接收或发送数据到世界各地;最后,描绘一套全球环境的图象,从而帮助决策者有效地管理环境资源.

在 GRID 的发展阶段有三项任务:(1)收集现有的数据集;(2)分析现有数据并指出重要的环境区;(3)为发展中国家及发达国家培训使用 GRID 技术的人才.

GRID 的数据采集分三种方法:航天、航空和实地. 这三种方法的主要差别不在于立足点而在于分辨率. 人们站在地面上可观测自然环境的细小变化,然而离地球 1000 公里以外的地球资源卫星就只能分辨那些大于 30 米的物体,从飞机上观测地球则介于上述二者之间. 这三种方法互相补充,便可逐渐描绘出一个地面的图象.

GRID 有许多特点,这些特点不仅使它在构造上区别于别的信息系统,而且某些特点是它之优越于其它信息系统的原因所在,其中最主要的便是它的数据的区域性. GRID 收集每个数据都必须先弄清它在地面上的位置. 这不仅便于进行一些区域性的分析,而

且实际上这样才完好地保存了数据,因为许多数据一旦失去了位置参数也就失去了意义. 下面列出了 GRID 的主要功能,很显然没有位置数据其中的许多功能是无法设想的.

表 1 GRID 的基本功能

功 能	任 务	示 例
评价功能	数据提供	提供肯尼亚北部大象数目的信息
	资源管理	全球土壤改良数据
	现状报告	定期环境污染现状评价
	动态监测	森林覆盖率变化报告
分析功能	研究支持	沙漠化原因分析
	预测	蝗虫群或天气预报
	改良管理	用 GRID 数据分析在干旱区进行放牧的最佳时间
	法律制定	试验可供选择的环境法
	紧急区的确定	确定急需发展支持的地区
	项目评价	在干旱区引入灌溉项目的环境影响评价

4. 国内的有关研究工作

国内在 EMIS 方面的研究起步较晚. 中国科学院环境化学研究所于 1983 年建成了“中文环境文献库”,这是一个在 ACOS 500 上运行的信息系统. 另外该所还引进了《Environmental Abstracts》、《Pollution Abstracts》以及《Energy Information Abstracts》等的磁带. 中国科学院兰州分院引进了《Water Resource》磁带. 化工部情报所引进了 CA 磁带. 中国科学技术情报所建成了国际联机站,与 ESA 和 DIALOG 系统联通,从而可使国内用户享受国外的信息资源.

在数值数据库方面,中国环境监测总站和北京市计算中心联合建成的“国家环境信息管理系统”于 1985 年 8 月通过国家鉴定.

作者以“环境管理信息系统的分析和一个微机环境数据库的实现”为题完成了硕士学位论文,这是一个用软件工程方法研制环境管理信息系统的尝试。

三、存在的问题及发展趋势

EMIS 的研究虽然取得上述重大进展,但存在的问题也是十分重大甚至是致命的。

首先是 EMIS 的研究中还没有一个自始至终运用软件工程方法研制成功的先例。软件工程是一般地信息系统研制的方法学,它要求信息系统的研制遵循一定的规范和步骤。比如:首先计算研制的得失、分析可行性,尔后定义系统的目标要求、构划出系统的逻辑模型。否则将导致大量返工和浪费。比如: ECDIN 的研制中由于中途变更系统目标及要求,迫使系统的研制者们必须选用新的数据库管理系统。系统设计中没有理论指导,其结果是各子系统之间协调性极差,美国环保局的工作大都存在这种问题。

其次是系统特性的综合性不够。CIS 和 ECDIN 的特性是数据的化学性,而 GRID 的特性是数据的区域性,即 GRID 从不抽象地谈数据,它的数据都有相应的位置参数,特别是由于 GRID 有很强的图象显示功能,区域性的结果便是直观性。环境模拟、模式方面的大量成果使得信息系统研究中考虑模型化不仅是必要的,而且也是可能的。

由此可见,用软件工程方法兼顾数据的化学性和区域性以及系统的模型化,是 EMIS 研究的发展方向。

致谢:中国科学院研究生院饶纪龙副教授指导了本文的文献检索工作,郭方、李康、吴锦、沈乃新几位先生给作者许多指导,在此表示衷心感谢。

参 考 文 献

- [1] Heller, S. R., Milne, G. W. A., *Environmental Science & Technology*, 13, 798—803(1979).
- [2] Reider, J. P., PB 84—198407.
- [3] Reider, J. P., PB 84—194091.
- [4] Howard, P. H. et al., *J. Chem. Inf. Comput. Sci.*, 22, 38—44(1982).
- [5] Lefrovitz, D. et al., *J. Chem. Inf. Comput. Sci.*, 21, 18—28(1981).
- [6] Mann, C. O., *J. of the Air Pollution Control Association*, 32, 1021—1024(1982).
- [7] Royce, C. L. et al., *J. Chem. Inf. Comput. Sci.*, 24, 7—10(1984).
- [8] Hushon, J. M. et al., *J. Chem. Inf. Comput. Sci.*, 23, 38—43(1983).
- [9] OLE NØRAGER et al., *J. Chem. Inf. Comput. Sci.*, 18, 134—140(1978).
- [10] Ott, H., Geiss F. & Town W. G., Second International Symposium on Aquatic Pollutants, Sept. 26—28, 1977, Amsterdam.
- [11] Fanshawe, J., GRID, UNEP, 1985.

(上接第 94 页)

(4) 污灌后灌区环境舒适程度下降,苍蝇、蚊子增多,污水散发恶臭。如辛院大队污水河绕村而过,空气质量很差。群众要求污水入暗管,以改善居民居住环境。

(5) 污水灌溉对作物病、虫害的直接影响尚不显著,间接影响还是存在的,如污水种稻营养生长旺盛,易发生稻瘟病等。污灌对人群健康的影响不容忽视,如侯家台村反映水田作业生皮炎很难治愈。且各庄反映高血压、心脏病、肿瘤病增多,需设专题调研。

(6) 南排污河灌区三个区域信访评价结果有差别,其原因在于南郊双巨排河系统污灌区污水水质差,其 $BOD/COD = 0.23$,可生化性能差,而且 SS 很高, pH 值不符合农灌标准, TDS 、 Cl^- 、 SO_4^{2-} 浓度均高于咸密系统和纪庄子系统。重金属 Zn、Pb 等的排放浓度也高,土壤中重金属累积多,因此群众对污灌评价最差。纪庄子系统污灌区水质较好,生活污水所占比例较高,群众又有丰富的净化污水、使用污水的经验,污水灌溉、污水养鱼效益明显。据此,在污灌区环境对策中对三个区域应有不同的侧重。