(20)

FINANCIAL ACCOUNTING DATABASES: ISSUES FOR MIS DESIGN

Miklos A. Vasarhelyi* and David C. H. Yang**

* Associate Professor and Director of the Accounting Research
Center, Graduate School of Business, Columbia University.
New York, NY 10027.

**Assistant Professor of College of Business Administration
University of Hawaii at Manoa. Honolulu, HI 96822.

Abstract

The increased usage of financial accounting databases requires
the examination of its role and potential within corporate
management information systems. Decision support systems now
often also use externally generated information. This substan-
tial increase of the usage of acquirable data structures brings
us to the discussion of key MIS questions related to: (1)
increased exposure due to acquisition of foreign data, (2)
integration of foreign data structures into corporate MISs, (3)
costs and needs, (4) potentials and liabilities, (5) database
updates and reliability, and (6) relationship with databases
suppliers.

The integration of these financial accounting databases into
corporate information systems brings in a new design dimension
that must be considered and analyzed. This paper discusses
these issues in light of features and comparisons performed
between two financial accounting databases---The Value Line and
COMPUSTAT databases.

1. INTRODUCTION

The increased usage of financial account-
ing databases (FADs) requires the examin-
ation of its role and potential within
corporate management information systems.
Traditional emphasis in management inform-
ation system (MIS) studies is on informa-
tion and decision support using corporate
generated and supplied information. Deci-
sion support systems (DSS) now often also
use externally generated information
(Gerrity[1], 1971).

The integration of these databases (DBs)
into corporate information systems brings
in a new design dimension that must be
considered and analyzed. This paper dis-
cusses these issues in light of features
and comparisons performed between two
financial accounting databases - The Value
Line and COMPUSTAT databases[2].

Among current usages of acquirable data
structures (ADSs) we find portfolio manage-
ment and brokerage services (security price
and accounting data), airline reservation

---

[1]Gerrity's decision support system linked portfolio information
with market information (provided by outside sources) to support
investment trust officers. The paper's discussions focus are
on the DSS features and little is said about the stock price and
information database.

[2]The Value Line Data Base is distributed by Value Line Data Ser-
vices of Arnold Bernhard & Co., Inc., New York, New York. The
COMPUSTAT database is distributed by Standard & Poor's Compustat
Services, Inc.

by travel agents (access to online reser-
vation systems and schedules), analytical
review by auditors (financial accounting
databases), acquisition and financial ana-
lysis by large corporations (financial
accounting databases), among many.

This substantial increase of the usage of
acquirable data structures brings us to
the discussion of key MIS questions related
to: (1) increased exposure due to acqui-
sition of foreign data, (2) integration of
foreign data structures into corporate
MISs, (3) costs and needs, (4) potentials
and liabilities, (5) database updates and
reliability, and (6) relationship with
databases suppliers.

Particular emphasis will be given to finan-
cial accounting databases notwithstanding
the increasing popularity of information
sources[3] such as The Source and others.

II. FINANCIAL ACCOUNTING DATABASES

Financial Accounting Databases (FADs) are
defined as acquirable data structures that
contain information on accounting and
security price data for a large number of
corporations. Among the most popular ones
(see Cuadra et al., 1981, for a more com-
plete list of online financial accounting
databases) we find:

a. VALUE LINE DATABASE.

b. COMPUSTAT.

c. CRSP[4].

d. NAARS[5].

They are typically supplied either as an
online information service, oriented
towards the terminal based query, or as a
periodically updated tape for archival
type of usage. Figure 1 displays a sum-
mary view of the corporate MIS incorpor-
ating FADs as an integral part of their
structures.

(FIGURE 1 here)

Figure 1 indicates the increased exposure
of a corporate MIS that comes from the
usage of acquired data. The purchase of
"mailing lists" and other types of archi-
val data of laborious collection has been
part of corporate life for many years.
The incorporation of these systems into
day to day operations and strategic deci-
sion support substantially increased
exposure risks. Other types of issues
however are more visible and require
discussion.

(A) Technical Issues

FADs come into being due to the economic
circumstances of data gathering. Most

---

[3]Among these we find The Source distributed by Source Telecompu-
ting Corporation and Dow Jones News/Retrieval Service & Stock
Quote Reporter maintained by Dow Jones & Company.

[4]The CRSP tape is maintained by The Center for Research in
Security Prices, University of Chicago.

[5]The National Automated Accounting Research System (NAARS) is
maintained by American Institute of Certified Public Account-
ants, National Automated Accounting Research System.

corporations would probably not be able to afford the costs of gathering publicly available data on their competition with other companies. On the other hand the collection and supply of this data can obviously benefit from economies of scale. Some of these databases are clearly the outcome of internal corporate needs (Value Line databases came into being due to the needs of the Value Line Investment Survey and then became a product on its own) while others were developed from clear research needs (e.g., the CRSP tapes) and are slowly sinking into the meanders of corporate usage. This hybrid origination brings out issues to be discussed:

(1) The data in a FAD is not necessarily the data that a corporation would have chosen to gather. Most often it has substantial amount of useless information and also does not contain some information that may be available but was not coded. In consequence we can observe an increasing usage of multiple merged databases in research and in industry. These merges increase the risk of missing potentially inconsistent data due to definitional and other types of problems.

(2) FADs tend to be fixed record size, sequentially based information with data added year after year presenting high percentages of missing values in the early years. This limits the feasibility of time series forecasting models based on early data. In addition sequential files do not provide the flexibility needed to classify and analyze available data according to alternative concepts (different from the original key). It also makes it difficult the integration of FADs with internal databases.

(3) Online databases on the other hand present a different set of problems. NAARS which contains substantial parts of the text of more than 4,000 financial

statements per year is queried through NAARS/LEXIS terminals which are supplied by Mead Data (the source for NAARS) and limited by key query words and categories. Users must use a NAARS terminal leading to major limitations on the integration of its content into a corporate MIS.

The trend for online databases is increasing availability and facilitation of user interface. This would be accomplished by allowing microcomputers and other systems interface to the base and capture data and/or execute users' generated data. Benjamin (1982) identifies key trends for MIS and DP during the next decade among which:

- decreased terminal costs
- decreased communication costs

This mode of operation creates a series of MIS design problems.

a. Security - if the MIS is fully interfaced (with automatic update) with a foreign source it abdicates its data screening prerogatives.

b. Costs - online databases searches, when in a foreign environment are typically extremely costly and present very steep learning costs leading to very high "usage initiation fees" even if these are not explicit costs.

c. Definitional Differences - a recent study (Vasarhelyi and Yang, 1986) compared the data content of two popular FADs and found over a period of 11 years effecting about 1,400 comparisons each year (see figure 2). The high percentage rate of discrepancies basically reflects the inherent difficulty of classifying data into pre-set categories that are consistent over time and across industries. A brief examination of the explanation of these discrepancies is shown in Figure 3.

(FIGURE 2 here)

Data in Figure 2 illustrates the scarcity of comparable information in the early years of data as well as some degree of increased comparability in the latest years. It is noteworthy to observe the consistency of the discrepancy rates among the two databases. A more detailed view of the discrepancies is shown in Figure 3 for 1981. It can be observed that substantial explanation of the differences are due to between-industry coding differences while the estimated actual error rates are small (in the 1%-3% range).

(FIGURE 3 here)

(E) MIS Design Issues

FADs will be used in formal organizational systems either on a production function or most likely as a source of analytical reference. As a production aid organizations may prepare, for example, investment newsletters based on the data from the database and distribute it as a paid service or an internal investment advice sheet. This usage poses interesting legal problems on the propriety of distributing data purchased from a different source and these will have policies concerning this secondary distribution. Other production type services may routinely prepare early warning reports on potential problem firms among the bank's client's (calculation of ratios), analytic data on the competition, etc.

Analytical reference on the other hand is more ad hoc and potentially difficult to predict. Recent developments in the DSS area provide general guidelines which can help in the design of these applications. In particular design should focus on (Alavi, 1982):

- flexibility to handle varied situations
- ease of usage
- responsiveness/flexibility to users' needs

- users' friendliness

As examples corporations will defend or plan takeovers based on theoretical "consolidation" models which merge two companies and allow for the insertion of "what if" type of actions by the protagonists. Often the best data available on potential acquirees comes from FADs. NAARS data will be used to search for particular accounting disclosures that may be required in the course of a particular fiscal year. CRSP data may be used to test alternative portfolio strategies by investors.

All these uses lead to a series of questions that must be responded by users/potential users and suppliers of FADs:

(1) Expected applications for the FADs.

(2) Expected frequency of usage.

(3) Needs of timely access, in particular on-line retrieval of data.

(4) Needs of data retroactivity.

(5) Needs of data interface with other data and other systems.

(6) Nature of user and his access needs.

(7) Needs for data update.

(8) Security and data access needs.

Two anecdotes may serve to illustrate the needs and parameters of choice in different situations.

A small company acquired an Apple II computer for its mailing list and newsletter word processing needs. This company basically sells consulting and information services to a special industry with about 40 major companies. A consultant called their attention to the fact that they were collecting large amounts of financial accounting information on these companies that was already available in FADs and incurring substantial costs. Their needs were simple: financial accounting inform-

ation had to be supplied in a consistent format, on a timely basis prior to the newsletter deadlines after the annual reports and quarterly announcements. They had no need for online access or frequent usage. Their solution encompassed using the Apple II as a data transfer terminal out of a time-sharing service that contained a popular database. Negotiations with the database supplier resolved the data privacy issue as most of the newsletter subscribers were also subscribers of the database.

A large western based conglomerate is often involved in the acquisition of other organizations. These are often privately held and must have their data verified against industry standards for the assessment of its quality and the desirability of the acquisition. These acquisition screening meetings are surrounded with secrecy requiring the minimum possible number of people to be involved. This excluded systems and financial analysis people. They purchased the usage of a FAD and using its data created preferential models for typical company performance in different activities and at different sizes. These models serve as "benchmarks" to examine the potential health of an acquiree. In addition while the data is standardized and entered for comparative analysis it is also used to prepare a "pro forma consolidated" statement that captures acquisition terms and outputs a sector based merger "result". The system is conversational and rather rigid in features but captures updated industry standards from an often updated FAD and allows computer neophytes (as the financial vice president) to operate it and try different acquisition strategies. A fully interactive and flexible acquisition analysis system has been proposed and is being considered for development.

III. CONCLUSIONS

FADs are increasingly being incorporated into corporate information systems creating an unusual problem for the information system designers. FADs are typically based in sequential data structures and have built-in time-series inconsistencies. In the future databases with substantial improved query features are going to be made available by vendors. These will bring with them problems of access limitations, lack of tailorization and considerable access exposure.

There is a real need for research into accounting information databases in particular FADs. A survey is needed of what is available, where and what does it contain and for how much it can be obtained. Surveys are also needed to establish the frequency and intensity and purpose of use of FADs within MISs. Field studies are needed to experiment with potential state-of-the-art uses of FADs. Laboratory studies could be performed to examine human interface with data rich systems and way of expert query into these large knowledge bases.

REFERENCES

COMPUSTAT Manual, (Investor Management Sciences, Inc., 1982).

Alavi, Maryam, "An Assessment of the Concept of Decision Support Systems as Viewed by Senior-Level Executives," MIS Quarterly, Vol. 6, #4, December 1982, pp. 1-10.

Benjamin, Robert, "Information Technology in the 1990s: A Long Range Planning Scenario," MIS Quarterly, Vol. 6 #2, June 1982, pp. 11-32.

Cuadra, Ruth N., Abels, David N. & Wanger, Judity, Directory of Online Databases, Cuadra Associates, Inc., 1984.

Emery, James C., Organizational Planning and Control Systems: Theory and Technology, (New York: MacMillan, 1971).

Gerrity, Thomas P., "Design of Man-Machine Decision Systems: An Application

to Portfolio Management," Sloan Management Review, Winter 1971, pp. 59-75.

Keen, Peter G., "Value Analysis: Justifying Decision Support Systems," MIS Quarterly, Vol. 5, #1, March 1981, pp. 1-16.

The Value Line Data Base-II (A User's Manual), (Arnold Bernhard & Co., Inc., 1981).

Vasarhelyi, Miklos A. & Yang, David C. H., "Financial Accounting Data Bases: Methodological Implications of Using the COMPUSTAT and Value Line Databases," Working Paper, Columbia University, 1985.
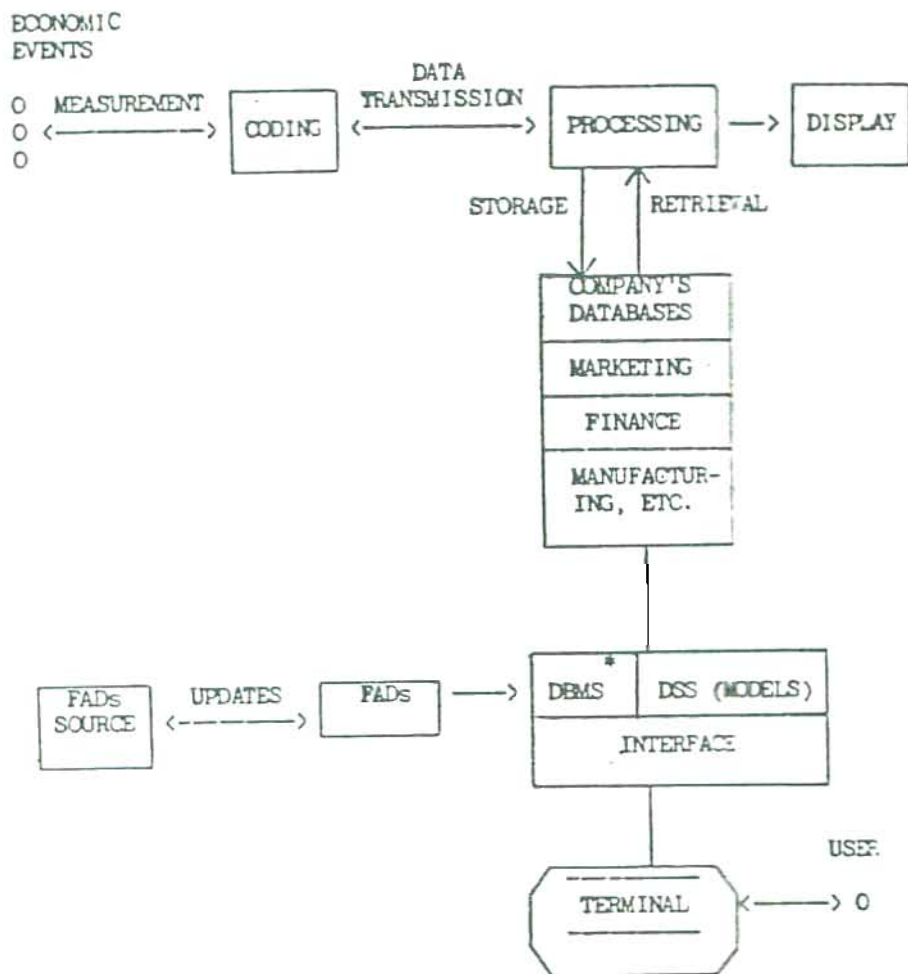
BIOGRAPHY

MIKLOS A. VASARHELYI

Associate Professor and Director of the Center for Accounting Research, Graduate School of Business, Columbia University, (Visiting at the AT&T Bell Laboratories - 1985/86).

B.S. in Economics; B.S. in Electrical Engineering; M.S. in Management, MIT 1969; Ph.D. in Accounting Information Systems, UCLA, 1973. Teaching and research interests: expert systems, computer auditing, financial databases and human information processing.

Received grants from the Touche Ross Foundation, Peat, Marwick and Mitchell Foundation, and the FASB (Database #33/36) and a pool of corporations (Exxon, GE, IBM, Morgan Stanley, and Shell). Was recently commissioned for an Accounting Research Monograph on Financial Accounting Databases by the American Accounting Association.

Author of several books on Accounting (Pensions and Inflation), Computer Related Topics (APL language, Microcomputers, and Advanced Audit) and Research (Literature Directory), and articles in the Journal of Accounting Research, Auditing: A Journal of Theory and Practice, Journal of Accountancy, The CPA Journal, EDP Auditor, and Decision Sciences.

DAVID C. H. YANG

Assistant Professor, School of Accountancy, College of Business Administration, University of Hawaii.

B.S. in Economics, National Taiwan University, 1977; MBA in Finance, University of California, Berkeley, 1979; Ph.D. in Accounting, Columbia University, 1985. Teaching and research interests: accounting information systems, computer auditing and internal controls.

FIGURE 1
**

FADs in MIS

** Modified from Emery's (1969) symbolic view of the corporate MIS.
* Data Base Management System.

FIGURE 2
DISCREPANCIES IN THE COMPARISON OF SEVEN FIELDS OF
THE VALUE LINE AND COMPUSTAT DATABASES
1971 - 1981

ASSET: TOTAL REPORTED ASSETS
SALES: NET SALES
INBET: NET INCOME BEFORE EXTRAORDINARIES
INVNT: INVENTORIES
DEPRE: DEPRECIATION, DEPLETION, AMORTIZATION
LIABL: CURRENT LIABILITIES
PLANT: GROSS PLANT

DATA ITEM DISCREPANCY RATE	ASSET %	SALES %	INBET %	INVNT %	DEPRE %	LIABL %	PLANT %
YEAR (NUMBER OF COMPANIES MATCHED)							
1971 (96)	20.8	24.0	21.8	93.7	93.7	95.8	96.8
(7)*	0.0	14.3	0.0	14.3	14.3	14.3	85.7
1972 (101)	16.8	25.7	19.8	94.1	95.0	96.0	93.1
(7)*	0.0	14.3	28.6	14.3	28.6	42.8	71.4
1973 (1336)	2.6	9.4	10.4	24.2	39.0	12.3	18.4
(1242)*	7.6	13.2	11.5	19.8	40.5	8.1	19.7
1974 (1454)	4.4	10.5	12.2	24.3	40.6	13.4	18.5
(1356)*	11.6	13.2	13.4	21.2	42.7	10.1	19.4
1975 (1480)	5.4	11.3	14.3	25.1	41.0	13.2	19.0
(1380)*	10.8	15.2	14.9	20.6	42.2	18.5	20.0
1976 (1478)	2.9	9.7	10.2	23.7	39.5	11.1	16.2
(1378)*	9.0	14.2	11.8	18.9	41.5	16.9	17.5
1977 (1486)	4.8	11.1	12.0	24.1	41.7	12.4	18.6
(1386)*	9.7	15.2	13.5	19.0	42.8	7.7	20.5
1978 (1475)	4.9	11.2	10.6	23.9	43.3	12.8	20.1
(1376)*	9.5	11.6	15.1	18.5	45.5	8.0	22.0
1979 (1496)	5.2	11.7	12.4	24.5	43.5	13.0	20.4
(1395)*	9.4	13.9	16.2	20.0	46.3	8.6	21.8
1980 (1514)	4.8	11.6	13.6	24.5	42.7	13.1	21.7
(1413)*	8.7	16.0	15.2	20.0	45.5	8.1	22.9
1981 (1431)	3.3	10.0	14.7	23.2	39.5	11.8	19.5
(1380)*	6.6	16.1	14.3	18.1	41.8	6.4	20.9

* Companies having data items with missing values in both databases
were excluded.

FIGURE 3
COMPARISON OF COMPUSTAT AND VALUE LINE DATABASES - 1981
BASED ON 200 COMPANIES
DISCREPANCIES SOURCE

ASSET: TOTAL REPORTED ASSETS
SALES: NET SALES
INBET: NET INCOME BEFORE EXTRAORDINARIES
INVNT: INVENTORIES
DEPRE: DEPRECIATION, DEPLETION, AMORTIZATION
LIABL: CURRENT LIABILITIES
PLANT: GROSS PLANT

DATA ITEM	ASSET	SALES	INBET	INVNT	DEPRE	LIABL	PLANT
NUMBER OF DISCREPANCY	#	#	#	#	#	#	#
DISCREPANCY RATE	%	%	%	%	%	%	%

SAMPLE SIZE: 200 COMPANIES

	ASSET	SALES	INBET	INVNT	DEPRE	LIABL	PLANT
I. EXPLAINED DEFINITIONAL DIFFERENCES:							
A. DIFFERENT CURRENCY USED	5	5	5	5	5	5	5
	2.5	2.5	2.5	2.5	2.5	2.5	2.5
B. DIFFERENT INDUSTRY CLASSIFICATION OR VARIABLE DEFINITIONAL DIFFERENCES	0	20	7	60	23	14	26
	0	10	3.5	30	11.5	7	13
II. UNEXPLAINED DEFINITIONAL DIFFERENCES:							
A. NON-DISCLOSED CODING RULE DIFFERENCES OR CODING ERRORS BY							
COMPUSTAT	0	2	6	9	69	1	12
	0	1	3	4.5	34.5	0.5	6
VALUE LINE	0	1	4	13	9	2	7
	0	0.5	2	6.5	4.5	1	3.5
BOTH	0	0	0	0.	6	0	4
	0	0	0	0	3	0	2