

RESEARCH CONTRIBUTIONS

Some Realities of Data Administration

BEVERLY K. KAHN Boston University

Beverly Kahn's current research focuses on the practice and effectiveness of data administration. She is a frequent contributor at conferences, member of advisorv panels, and consultant on information resource management, database design, and requirements analysis. Her methodologies for requirements analysis and database design are used by several major corporations and government agencies. Author's Present Address: Beverly K. Kahn, Graduate School of Management, Boston University, Boston, Massachusetts 02215.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission. © 1983 ACM 0001-0782/83/1000-0794 75¢

INTRODUCTION

Data is recognized as a valuable organizational resource, one that should be managed like other corporate resources, such as manpower, money and materials [12]. According to John Diebold [1]:

It is clear that organizations which will excel in the 1980s will be those that recognize information as a major resource and structure it as efficiently as they do other assets.

Providing accurate, relevant, and timely information entails a significant cost for the organization and presents a substantial challenge. Enterprises spend at least ten cents of every single revenue dollar on the management of information [2]. In the 1970s, a certain degree of specialization (or division of labor) in managing the data resource emerged from the data processing function [12, 13]. This specialization was referred to in two ways: data administration and database administration. Today, however, these names are used to describe two separate functions.

There is a distinction between these two functions. Data administration (DA) is the establishment and enforcement of policies and procedures for managing the company's data as a corporate resource. It involves the collection, storage, and dissemination of data as a globally administered and standardized resource. Database administration (DBA) is a technical function which performs database design and development, provides education on database technology, provides support to users in operational data management-related activities, and may provide technical support for data administration. Data administration and database administration are not the same though some authors use the terms interchangeably, and others refer to the entire technical-to-administrative continuum as "data administration."

Little information exists on the practice of data administration which is defined in this study as the administrative and policy side of this continuum. This data administration is thought to be uncommon. Previous studies [5–9, 11] actually investigated database administration even though they called it data administration when confronted by virtually no data administration functions as defined herein. ABSTRACT: Little information exists on the practice of data administration, i.e., the management of data as an enterprise resource. A survey of data administrators at predominantly large enterprises was conducted to determine the extent of the existence of data administration and to assess its effectiveness as implemented. New insight into the realities of data administration, including its realworld objectives and perceived effectiveness, is presented.

Presented here is a survey/research investigation assessing the effectiveness of data administration as currently implemented, establishing the existence of data administration, and increasing the base of knowledge about its practice. This

What form of the DA function exists in enterprises?

research set out to answer the following questions:

- What enterprise characteristics, if any, influence the
- form of this function? What are the properties of existing DA functions?
- What is the relationship between the form of the DA function and the enterprise's data management environment?
- How does the form of the DA function correlate with the scope and effectiveness of its activities?

Methodology and Overview

During the latter half of 1982, a survey was administered by mail. The questionnaire employed multiple-choice questions (available upon request from the author), and collected information in six areas: enterprise characteristics, MIS function characteristics, DA function characteristics, DA and DBA activities, and data management environment. The survey respondents represent 56 enterprises and do not constitute a scientific sample. Most of the respondents serve as the data administrator or perform such a function for their enterprise. They are interested in effective information management and, therefore, were assumed to present a positive view on data administration.

Most of the survey respondents were employed by Fortune 1000 companies [4]. The responses were classified according to the size of the enterprise (Table I):

- Large: Fortune 300 with over \$1 billion in revenues and over 50,000 employees (47% of the respondents).
- Medium: Fortune 300 to 1000 with revenues of \$150 million to \$1 billion and 10,000-50,000 employees (36% of the respondents).
- Small: Revenues less than \$150 million and less than 10,000 employees (17% of the respondents).

In practice, data administration can be implemented with varying degrees of objectives and commitment. Three levels

TABLE I. Survey Sample by Industry				
Industry	Number of Companies (%)	Number with IRM ¹ (%)		
Banking and Insurance	11 (19.6%)	5 (45%)		
Chemical, Petroleum, and Coal Products	3 (5.4%)	2 (67%)		
Conglomerates	2 (3.6%)	1 (50%)		
Educational	4 (7.1%)	1 (25%)		
Electrical and Electronic	7 (12.5%)	5 (71%)		
Government Agencies and Health Care	4 (7.1%)	2 (50%)		
Manufacturing and Processing	14 (25.0%)	4 (29%)		
Public Utilities and Transportation	3 (5.4%)	2 (67%)		
Service (Data Processing, etc.)	3 (5.4%)	2 (67%)		
Other	5 (10.7%)	2 (50%)		
Totals	56	26		

¹ Two enterprises did not report the form of the data administration function

of practice, each constituting a form of data administration function, were identified:

- The practice of information resource management which consists of both data administration and database administration (48% of the sample including one enterprise with only a data administration function).
- The practice of database administration only (24%).
- No formal data administration or database administration practices (28%).

The survey revealed that an increasing number of organizations was instituting an information resource management (IRM) function. (For clarity, the first level (both DA and DBA) will hereafter be called information resource management.) In contrast to the 28% reported by Goldstein [8], 48% of the total sample had an information resource management function. Of those without a formal function, some DA and DBA activities were performed as part of other enterprise functions. Database administration was still more common than DA: 72% of the enterprises in the study have a DBA function. In addition, the results indicate the existence of a DBA function was increasing at a faster rate than the DA function.

To determine the significance of the form of DA, two different sets of hypotheses were tested to determine the differences between enterprises:

- What characteristics distinguish enterprises that have IRM from those that do not? This can be stated formally as: H_0^1 —There is no difference in Characteristic X between those enterprises that have IRM and those without IRM.
- What characteristics distinguish enterprises that have 2. IRM from those with only DBA? This can be stated formally as: H₀²—There is no difference in Characteristic X between those enterprises that have IRM and those with only DBA.

A third hypothesis was evaluated to determine which enterprise characteristics are independent of the form of the data administration function, i.e., which properties are common to all three DA forms.

FINDINGS

Data Administration Moving Toward Policy Concerns

In past studies, information resource management was found to be a technical function dealing with the operational support of computerized databases [5-9]. A new focus was noted in this survey. Determining, documenting, and maintaining data management standards are now major activities of IRM and are not necessarily activities of organizations without IRM. The survey results indicated that when standards exist, they were enforced.

Four classes of data management standards were investigated in this research. They pertain to: (1) data dictionaries; (2) data definitions; (3) database/file design; and (4) general data management activities. In testing for a relationship between the use of data management standards and the existence of IRM (an example of H_0^1), the hypothesis for no such relationship was rejected at p < 0.01 for three of the four classes of standards. The results of testing this relationship were the same when comparing enterprises with IRM to those with only DBA (Ho2). When IRM exists, all four classes of standards were more likely to be used than in the other two forms of the DA function, i.e., DBA only and neither DA nor DBA.

RESEARCH CONTRIBUTIONS

However, this test is not significant in the case of database/ file design standards. This result can most likely be attributed to the fact that effective use of a DBMS depends on the use of database design standards.

There was a difference in the perceived importance of DArelated activities between those enterprises with an IRM function and those without one. A comparison of enterprises with IRM to those without IRM and to those with only DBA yielded similar results (Table II). The data shows that enterprises with IRM consider the following six DA-related activities significantly more important:

- developing and enforcing policies governing data collection
- developing tactical and strategic plans for data use
- developing definitional requirements for dictionary items
- identifying opportunities for data sharing
- identifying potential database applications
- controlling the integrity and security of the enterprise data.

The ratings of these activities indicate attention to policy concerns. All of these activities were considered more important by those with an IRM function. There was no significant difference between enterprises with each form of the DA function in terms of the importance of the non-DA-oriented activities (Table II). The importance of the DBA- and MISoriented activities appears to be uniform across all forms of the DA function.

When Do Enterprises Have A Data Administration **Function?**

In previous studies [5, 7], the general determinants of an information resource management function were the size of the enterprise and its profitability. As a result of general opinion, the importance of the information to the enterprise, as well as its industry classifications are now added to the list of determinants. In this survey, these and other factors were correlated with the existence of an information resource management function.

TABLE II. Organizations Rating DA-Oriented Activities as Important¹

Activity	iRM (%)	DBA (%)	No Formal (%)	Totai (%)
Develop/enforce naming conventions	50	45	43	47
Control data integrity and security ²	85	67	57	73
Develop and enforce policies governing data collection ^{2,3}	31	9	7	20
Identify data sharing opportunities ^{2,3}	43	9	21	29
Develop tactical/strategic plans for data use ^{2,3}	31	0	21	22
Develop definitional requirements for dictionary items ²	60	18	21	40
Identify potential DB applications ²	40	27	21	32
Plan the evolution of the corporate DB	35	18	29	29
Maintain the corporate DB	32	36	29	32

¹ Importance was evaluated on a 5-point scale with "1" denoting very important and "5" not important. This is the percentage of organizations ranking the activity as a 1 or 2. ² Humobics H 1 was rejected at 0.01 are a 0.07

² Hypothesis H₀¹ was rejected at 0.01 .³ Hypothesis H₀² was rejected at <math>0.01 .

TABLE III. Relationship Between Type of DA Function and Enterprise Size¹

Form of the DA Function ²				
Size	DA and DBA	DBA only	No Formal	Total
Large	19 (76%)	2 (8%)	4 (16%)	25 (52%)
Medium	5 (26%)	9 (48%)	5 (26%)	19 (40%)
Small	0 (0%)	1 (25%)	3 (75%)	4 (8%)
Total ³	26 (49%)	13 (25%)	14 (26%)	

 $^{1}p < 0.01$. ² Count (percentage across row). ³ Considers all enterprises that denoted the form of the DA function regardless of whether they responded to the company size question.

Past research indicated that the IRM function was scarce and existed only in large enterprises [7, 11]. In this study, 76% of the large enterprises have an IRM function. However, IRM also exists in smaller enterprises. In medium enterprises, the DBA function, without a DA, dominates; 47% have DBA, while only 26% also have a DA function. In small enterprises, the lack of any formal DA or DBA function dominates. However, this lack also exists in a sizable number of large (16%) and medium (26%) enterprises. This relationship between size of enterprise and type of DA function is summarized in Table III. Enterprise size was found to be a major factor (p < 0.01) in determining the form of the data administration function.

Goldstein [8] and Uhrbach [15] claimed that the form of the data administration function is dependent on the enterprise's industry. Uhrbach [15] proposed that industries whose primary service is information, e.g., banking, finance, or insurance, would more likely support IRM. Survey respondents in these enterprises reported that information was extremely critical to their enterprise's success. In theory, IRM is expected to enhance the provision of high-quality information. However, respondents from the financial industries were less likely than the average of the sample to have an IRM function (only 40%). Additionally, over 70% of companies in the electronics industry were found to have an IRM function. This finding refutes the common belief that this industry is not progressive in its internal MIS operations.

Analysis of survey results suggests that the form of the DA function is independent from the enterprise's industry affiliation. This conclusion was based on analysis (Table I) over all industry categories and on four aggregated industry groupsbanking and insurance (20.4% of the respondents), manufacturing (31.5%), electronics (16.7%), and other (31.5%).

The chief information officer (CIO) provides the leadership for all MIS-related activities including information resource management. A preliminary hypothesis of this research was that characteristics of the CIO should determine the form of the DA function. Analysis indicated that the form of the DA function was dependent on the reporting level of the CIO with respect to the president of the enterprise (0.01 . The IRM function was more likely to existas the CIO was placed higher in the enterprise's reporting structure. To whom the CIO reports was not a factor in determining the form of the enterprise's DA function.

Characteristics of the Data Administration Function

The data administration function was accomplished by a small and young organizational unit: 80% of the functions were less than four years old; 40% were less than two years old. These functions are slightly younger than those studied

by McCririck [11] and Gillenson [6] who studied database administration functions. Over half of the DA organizations had four or fewer employees and 19% had only one. All DA functions reported within the MIS/DP organization. Surprisingly, the resources devoted to their data administrationrelated efforts were considered sufficient by 84% of the respondents that had a formal DA or DBA function. This correlates with the fact that 73% of the enterprises felt that their data administration efforts were partially successful as implemented.

The survey results indicated that the DA function had achieved a high position in the MIS organization (relative to the CIO): 64% of the DA functions reported to the CIO. This shows a positive and upward movement from the low position observed by McCririck [11].

A generic organizational structure of the information resource management function was identified. When both DA and DBA existed, they were most likely to occur as two independent functions. Database administration was subordinate to or part of DA in only 25% of the enterprises. The DA and DBA functions were usually at the same organizational level. The most common organization structure was for both the data administration function (64%) and the database administration function (42%) to report to the chief information officer.

Impact of the Data Management Environment

The data management environment of the enterprise (that is, the use of database management systems (DBMS) and data dictionary systems (DDS), and the degree of data integration) was shown to be related to the form of data administration function [6, 11]. In this study, it was determined that the existence of data management software is very common and widespread regardless of the form of the DA function or the size of the enterprise. At least one DBMS existed in 86% of the enterprises, and a computerized DDS existed or was being acquired by 78% of the respondents.

In previous research [5, 11], it was not determined whether the use of DBMS and DDS caused the establishment of DA or DBA or whether the establishment of these functions caused the use of the DBMS and DDS. The order of occurrence of the following data management events was analyzed across the survey population:

- Installation and use of a DBMS
- Installation and use of a computerized DDS
- Establishment of a DA function
- Establishment of a DBA function.

In a majority of the cases the events occurred in this order:

- 1. DBMS is installed and used.
- 2. DBA function is established.
- 3. DDS is installed and used.

This ordering follows common sense and is the same ordering that was conjectured by Nolan [14]. The DBMS was in place the longest and has an established reputation. Often, the DBMS did not solve the enterprise's data management problems, and, in fact, created new ones. Then, DBA was established to increase the effectiveness of the DBMS and to solve some of the related organizational problems. The database administrator would determine that a DDS was required to solve these problems and to increase the effectiveness of the DBMS. The DBMS and DDS were installed simultaneously in 13% of the enterprises. Enterprises that did not have a DBMS and DDS until recently are usually procuring this software simultaneously. This appears to be the current trend. Establishment of the DA function did not follow any specific pattern. Of the organizations with DA, 22% established it first and the remainder established it after the installation of the DBMS (39% after the installation of both DDS and DBMS). In enterprises with both DA and DBA, DBA was established before DA in 59% of the cases.

Success Versus Effectiveness of Data Administration

Respondents indicated the degree of success of their data administration efforts as implemented by their enterprise: successful (13%), partially successful (60%), and unsuccessful (27%). Respondents were not specifically asked to indicate their metrics for success. The degree of success correlated highly with the form of the data administration function: 96% of those with IRM considered themselves to be partially-tototally successful. Two-thirds of those with DBA only considered themselves to be partially to totally successful, while only 36% of the enterprises with no formal function considered themselves to be partially to totally successful. As expected, the highest degree of failure was found in enterprises without a formal function (63%).

Unfortunately, the high degree of perceived success of information resource management apparently did not result in the desired (positive) impact. This impact was assessed by comparing those activities that were identified as being improved as a result of IRM with those that were identified as being performed effectively. IRM-related activities were not substantially improved. Only the following activities were improved in a majority of the IRM organizations: data documentation (79% of the enterprises with IRM), consistent information (68%), database design (67%), and data sharing (55%). These improvements were analyzed with respect to the effectiveness of the activities required to support them.

An improvement in data sharing correlated with the enterprise's effectiveness (0.05) in promoting data sharing but the effectiveness was assessed to be marginal at best.An improvement in information consistency did not correlatewith effectively accomplishing those activities required toachieve this objective—maintaining data integrity, maintaining data accuracy, and defining integrity constraints. Enterprises which indicated improvement in their information consistency were no more effective in these activities than thosewhose information was not improved as a result of IRM.

Of the seventeen IRM-related activities studied, only two were being accomplished effectively by a majority of the respondents: procedures for data restoration (75%); and maintaining data integrity (53%). These results are summarized in Table IV. The degree of effectiveness in accomplishing various IRM-related activities was analyzed with respect to the form of DA function. In comparing those enterprises with an IRM function to those without IRM (H₀¹), enterprises with IRM were found to be more effective in restoring data, maintaining the data dictionary, and maintaining a conceptual model of the enterprise's data (0.05). In comparing thoseenterprises with IRM to those with only DBA (H₀²), enterprises with IRM were found to be more effective in promotingdata sharing, performing feasibility studies, restoring data, andmaintaining a conceptual model (<math>0.05).

It was surprising that all enterprises, regardless of the form of the DA function, were ineffective in (1) establishing priorities of information requirements; (2) promoting data sharing; and (3) controlling redundancy. Most IRM functions were established to be more effective in accomplishing these activities. It appears that the desired impact has yet to be achieved.

RESEARCH CONTRIBUTIONS

TABLE IV. Effectiveness of Accomplishing	Data Administration Goals
--	---------------------------

Goals'	Level of	F	form of Data Administra	Data Administration ³	
	Effectiveness ²	IRM	DBA Only	No Formal	
FFECTIVE					
Procedures for data retention	1.98	85%	69%	71%	
IARGINALLY EFFECTIVE					
laintaining data integrity	2.54	46%	77%	50%	
nderstanding application's retrieval requirements	2.64	48%	39%	43%	
Naintaining data accuracy	2.65	46%	54%	50%	
roviding consulting to data processing personnel	2.69	44%	54%	29%	
roviding consulting to users	2.71	40%	54%	54%	
Inderstanding and maintaining data security needs	2.74	42%	54%	36%	
roviding training to data processing personnel	2.76	40%	54%	21%	
erforming feasibility studies5	2.91	36%	31%	43%	
efining data integrity requirements	2.98	36%	31%	31%	
roviding training to users	3.04	20%	23%	27%	
etermination of information requirements	3.15	32%	31%	2 9 %	
ARGINALLY INEFFECTIVE					
stablishing priorities of information requirements	3.31	21%	23%	36%	
Aaintaining data dictionary ⁴	3.37	48%	17%	27%	
romoting data sharing ^{4,5}	3.46	19%	0%	31%	
ontrolling redundancy	3.54	23%	0%	0%	
NEFFECTIVE					
Maintaining enterprise model of data4.5	3.98	16%	0%	21%	

¹Goal accomplishment was evaluated on a 5-point scale with "1" denoting highly effective and "5" denoting not effective; the categories were assigned with respect to the mean as follows: effective—less than 2, marginally effective—more than 2 and less than 3.2, marginally ineffective—more than 3.2 and less than 3.9, ineffective more than 3.9.

³ The percent of respondents in each form of the DA function deeming that the goal was accomplished effectively (e.g., 1 or 2). ⁴ Reject Hypothesis H_0^+ at $0.05 . ⁵ Reject Hypothesis <math>H_0^-$ at 0.05 .

There was no consensus in identifying the inhibitors to successful information resource management. A summary of the analysis of potential inhibitors is shown in Table V. Only one of the potential inhibitors was considered significant by a majority of the respondents-management's lack of understanding of the data administration concept (69%). Some of the potential inhibitors that were not considered impediments to the success of IRM were:

TABLE V.	Inhibitors to	o Successful	Data	Administration
----------	---------------	--------------	------	----------------

Inhibitors ¹	Not Considered an Inhibitor ²		
Management lack of understanding of	Poor quality standards		
the DA concept	Lack of user training		
Insufficient DA staff	Poor corporate planning		
Too small a budget	Responsibility without correspondin		
Too heavy a workload	power ³		
Lack of top management support	Inadequate DA tools		
	Insufficient quality DA personnel		
	Poor interface with users		
	Not enough responsiblity		
	Lack of user respect		
	Poor interface with DP personnel		

¹Potential inhibitors were evaluated on a 5-point scale with "1" denoting a major factor and "5" denoting not a factor. An inhibitor has a mean less than 3. The inhibitors are listed in an ascending order based on this mean. ²Those potential inhibitors classified as not considered an inhibitor have mean greater than 3. These are listed in ascending order based on this mean. ³An inhibitor to successful DA in enterprises with an IRM function. H_0^{-1} and H_0^{-2} rejected at p < 0.01.

- poor interface with DP personnel
- not enough responsibility
- . lack of user respect
- poor interface with users
- insufficient quality DA personnel
- responsibility without power
- inadequate DA tools.

Of the list of potential inhibitors to successful IRM, only one was dependent (p < 0.01) on the form of the DA function-responsibility without corresponding power. This relationship implies that those enterprises with an IRM function felt that they were less powerful and influential in implementing IRM-related activities than those organizations with no IRM or with only DBA. This may be attributed to the frustration that accompanies increased awareness and understanding of the scope and importance of IRM. Analysis of the other potential inhibitors showed that none were significant at the 10% level.

CONCLUSION

This study provides some new insights into the practice of data administration. Three forms of DA practice were found to exist in today's enterprises: information resource management (both DA and DBA), database administration only, and no formal function. Information resource management was found to be the most common.

The practice of data administration was found to be widespread but not very effective. Continuing a data administration effort carries a great cost to the enterprise [3]. The impact of these efforts needs to be monitored to determine whether the practice of data administration can be cost-justified.

RESEARCH CONTRIBUTIONS

REFERENCES

- Diebold, J. Information resource management—The new challenge. Infosystems 26, 6 (June 1979), 50–88.
- Edelman, F. The management of information resources—A challenge for American business. MIS Quarterly 5, 1 (March 1981), 17–27.
 Esteblisher the Data Administrative Evention Children and Internetional
- 3. Establishing the Data Administration Function. Guide International Corporation, Chicago, Illinois, 1977.
- 4. The Fortune Directory of the Largest U.S. Industrial Corporations. Fortune 105, 9 (May 3, 1982), 258–286. The Second 500 Largest Industrial Corporations. Fortune 105, 12 (June
- 14, 1982), 178-206.
 Gillenson, M.L. 1981 SRI Data Administration Survey Report. IBM TR 70 20 (2012) 1022 JUNE Survey Report. IBM TR
- 73-22, (June 22, 1981), IBM Systems Research Institute, New York.
 6. Gillenson, M.L. The state of practice of data administration—1981. Comm. ACM 25, 10 (October 1982), 699–706.
- Goldstein, R. C. National Data Administration Survey Summary of Results. June 1979, Faculty of Commerce and Business Administration, The University of British Columbia.
- 8. Goldstein, R.C. and McCririck, I.B. What do data administrators really do? Datamation 26, 8 (August 1980), 131-134.
- Goldstein, R.C. and McCririck I.B. The stage hypothesis and data administration: Some contradictory evidence. Proc. of the Second Intl. Conf. Info. Syst. Society for Information Management, December 1981, 309-324.
- Lyon, J.K. What? why? where? The Database Administrator, John Wiley and Sons, New York, 1976, 1–18.
- McCririck, I.B. A Survey of the Data Administration Function in Large Canadian Organizations, June 1979, Master's Thesis, The Faculty of Graduate Studies, The University of British Columbia.

- 12. Nolan, R.C. Managing the Data Resource Function, West Publishing Co., St. Paul Minnesota 1974.
- Nolan, R.C. Thoughts about the fifth stage. Database 1, 2, 1975, 4–10.
 Nolan, R.C. Data Administration. Managing the Data Resource Func-
- tion, (second ed.), West Publishing Company, St. Paul, Minnesota, 286-296.
 15 Uhrbach H Stors to shared database suggest sufficient Computer.
- 15. Uhrbach, H. Steps to shared database success outlined. Computerworld, May 23, 1982.
- 16. Weldon, J. Organizing for database administration. Working Paper CRIS 6, xvii, 21 (January 1979), Center for Research on Information Systems, Graduate School of Business Administration, New York University, New York.

The changing role of database administration. Working Paper CRIS 7, November 1979, Center for Research on Information Systems, Graduate School of Business Administration, New York University, New York.

17. Wiorkowski, K. and Wiorkowski, J.J. Does a database management system pay off? Datamation 24, 4 (April 1978), 109–114.

CR Categories and Subject Descriptors: H.2.7 [Database Management]: Database Administration; H.4.0 [Infomation Systems Applications]: General; K.6.4 [Management of Computing and Information Systems]: System Management

General Terms: Management

Additional Key Words and Phrases: information resource management, data administration, database administration, data administration function effectiveness

Received 2/83; revised 5/83; accepted 6/83

