Storing too much information in the insurance industry?

Yoram Eden,* Boaz Ronen,† and Israel Spiegler†
* The College of Management, Academic Studies, Tel Aviv, 61480, Israel
† Faculty of Management, The Leon Recanati Graduate School of Business Administration, Tel Aviv University, Tel Aviv 69978 Israel

Abstract: The paper challenges the premise that all data collected in the insurance industry must be stored in the database. This paper applies a cost/benefit model that allows for testing and assessing of data items currently stored in the database, as well as providing a 'to store or not to store' decision policy regarding prospective data. An empirical case study of a life insurance agency is presented through which the ideas are illustrated.

Key words: cost benefit analysis, databases, information, information in process (IIP), life insurance, response time.

Reference to this article should be made as follows: Eden, Y., Ronen, B. and Spiegler, I. (1996) 'Storing too much information in the insurance industry?', *International Journal of Computer Applications in Technology*, Vol. 9, Nos 2/3, pp. 144–150.

1 INTRODUCTION

The idea of achieving strategic advantage through information systems is gaining popularity in the insurance industry. Most insurers are placing an increasing emphasis on the management of their data systems to create or sustain a competitive advantage. The ideal goal is to create an infinite pool of data, from which a manager can get any information he or she wishes [5, 18].

This paper challenges the premise that all data collected in the insurance industry is essential, and hence must be stored in its database. We suggest that existing data must be regularly examined for their contribution and value to decision makers. Out-of-date and valueless facts should be discarded, and new data, about to enter a firm's database, must be quantitatively assessed before storage. Data, and the information produced from them, are thus treated as inventory in a production system [14, 15].

As argued by Ackoff [1], the critical deficiency under which most managers work is not the lack of relevant information, but an over abundance of irrelevant information. Unless the information overload to which managers are subjected is reduced, any additional information made available by an information system cannot be expected to be used effectively.

Apparently, the storage of more data is excusable and can even be justified by two common facts:

(a) There has been an on-going reduction in the price of software, hardware and processing data in the last 20 years [12].
(b) Recent development in information technology seems to overcome the information explosion.

In our opinion, these two facts are of an erroneous type. The cost reduction is usually presented as a *cost per unit*, rather than the total cost [7]. Attempts to tackle the information explosion, and turn data into a timely, storable, manageable and easily accessible commodity, have taken two main avenues—*technology* and *methods* [11]. Technological solutions have involved condensing and compacting the storage media, that is, decreasing physical size and increasing the capacity of the media. Methods have been developed for simple and economic representation of data, e.g. techniques for data compression [6, 8].

Moreover, it sometimes seems that insurers tend to add a lot of data, assuming that the marginal cost is negligible. This may look true considering the existing strategy. However, the analysis presented here may change the fixed cost of a system.

However, no storage technology or methods can provide a long-term solution to the incessant accumulation of data by individuals and organizations. Whatever its size or capacity, the 'container' fills up quickly at a rate that seems to match and surpass any technological developments.

The basic premise in seeking technological solutions to information explosion problems is that *all data are necessary and therefore must be stored*. This premise is challenged by our study.

If the manager's information problem continues to be over abundance of irrelevant information, than the two most important functions of any information system become *filtration* and *consideration* [1]. Our study applies a methodology based on a model developed by Kalfus, Ronen and Spiegler [11] that gives an appropriate answer to these questions. Appendix A shows the decision model and the methodology.

Section 2 discusses the idea of information in process (IIP), its problems and possible solutions. Section 3 describes our proposal for a solution to the IIP problem, and describes the methodology. Section 4 presents the empirical case study that illustrates the model. Section 5 presents a summary and conclusions.

2 INFORMATION IN PROCESS (IIP)

Ronen and Spiegler [15] coined the term information in process (IIP) as analogous to the work in process concept used in operations and inventory management [16, 17]. The following are some of the ills and 'evils' of IIP:

1 Information in process lengthens the *cycle time* needed for decision making.
2 The *quality* of information produced from a voluminous database is likely to be lower than one produced from a more compact database. High levels of work in process, as well as IIP cause more complex systems, lack of quick feedback on mistakes and longer processing time that causes quality problems [15].
3 The *age* of data is often ignored in making the 'to store or not to store' decision. Older facts take up storage space and impede access to needed information.
4 *Management* and *control* of the system is harder when much data and complex storage media need to be handled.
5 *Maintenance* of the information system becomes a problem as data for storage grows exponentially.
6 Data entry, done automatically and indiscriminately, tends to lead to the accumulation of many irrelevant facts, which are time consuming and costly to store.

The tendency to accumulate data indiscriminately stems from the notion that storage cost is low, and the 'fear' of losing some important data that may be valuable 'some day'. It is based on the fallacious idea of a decreasing unit cost per megabyte of data, when the real key issues are total organizational cost and the fact that IIP increases at a *faster* rate than unit cost (e.g., if the cost per unit were decreased by 50%, and the volume were increased by two orders of magnitude—100—the total cost would be increased by a factor of 50. For further information the reader may refer to [5]).

The less we understand a phenomenon, the more variables we require to explain it. Hence, the manager (or the system designer) who does not understand the phenomenon he or she controls wishes to store 'everything', and thus increases the quantity of irrelevant information [1].

2.1 Possible solutions

Two approaches have been taken to solving the problem of over-abundance of data:

1 Technologies which increase and stretch the capacity of storage media. These include the development of new chips, CD-ROM, and optical storage methods.
2 Methods for the efficient and economic storage of data to maximize data storage per unit of area. These include coding methods, representation, and compression.

The real constraint and the scarce resource of the system is management time. The more irrelevant information we store in the system, the more management time will be consumed to manage it and to explore the effective pieces [15]. The more quality problems occur, the more management time will be spent to recover the damages.

3 SOLUTION

The proposed solution to the IIP problems stated above consists of two parts: a *methodology* for structuring decisions on accepting data into a database, and a *model* to define the costs and benefits of the different variables relating to the data resource.

We use the Pareto principle, which claims that items vary in their relative importance. The general categories of the Pareto (often called the ABC model) are:

Group	*Size*	*Importance*
A	20% of population	80% of importance
B	30% of population	10% of importance
C	50% of population	10% of importance

This 80 : 20 principle of the ABC model is applicable to the handling of IIP, the implication being that management should focus its attention on these more important data items which constitute a mere 20% of the system's data population [9, 10].

The relative importance of data items can be prioritized according to parameters such as the cost of the

Table 1 *Data item profile.*

Parameter	*Meaning and use*	*Need and values*
Length/Area	Storage space the item takes	Cost calculations. (number * length)
Meaning of item	Understanding the data item	– in processes – relevance
Is item used (x_1)	Is item used as input to processing programs	0 = No, 1 = Yes % use of system
Relevance to population (x_2)	Is item relevant to other occurrences	0 = No, 1 = Yes % use in system
Updating relevant population (x_3)	Are there data in the defined range	0 = No, 1 = Yes
Item raw or result of operation (x_4)	Is the item source data or produced by routine	0-Yes, 1 = No
Item supplied from other source (x_5)		0 = Yes, 1 = No For alternatives
Critical for work process *j* (q)	Is item critical to success of process	0 = No, 1 = Yes

Note: x_1 and x_2 are continuous variables; the others are treated as 0's and 1's only.

item, how critical the item is to production, the cost when the item is lacking, difficulty in obtaining an item, and the like.

3.1 Methodology

The methodology consists of a series of steps which help in structuring decisions and policies for handling data and IIP. The steps are:

1 Establish Organizational and Information System Goals.
2 Identify Boundaries and Data Population.
3 Divide the Population into A, B, C Categories.
4 Define Data Profile.
5 List Decision and Work Processes.
6 Estimate Parameters for Data Items.
7 Define Importance of Groups.
8 Classify Data Items into Groups.
9 Cost/Benefit Analysis.
The cost/benefit analysis is performed for each of the five alternatives for handling a data item (see also Table 2).
The alternatives are:
(a) Retain the item.
(b) Retain the item, but change the data structure. This alternative deals with data items which are relevant only to part of the overall data population. Isolating the relevant part may result in a substantial storage saving. For example, a file of 100 000 records, where each record has 100 bytes, may have a primary key field of 15 bytes and another field of 10 bytes which is relevant to only 5% of the population. Pulling out this item from the population means the system will require
$(15 + 10) \times (0.05 * 100\,000) = 125\,000$ bytes,
instead of
$100\,000 \times 10 = 1\,000\,000$ bytes,
resulting in storage saving of 87.5%. Thus, a change of structure should be considered. Surely, including in a file a field relevant to only 5% of the population is a design error that was probably missed at file design time.
(c) Remove the item and produce it upon demand.
(d) Remove the item from the system and supply it from another source upon demand.
(e) Remove the item from the system and do not supply when demanded.
10 Operational Decision.
11 Overall Cost/Benefit Calculation.
12 Follow Up.

The methodology outlined above requires the formal definition of variables needed to assess the various alternative decisions and policies. The full description of the model is explained in Appendix A.

4 CASE STUDY

The methodology and model described above were empirically tested in a case study of a large life insurance agency which has been in business for twenty-five years, selling most types of life insurance policies to firms and individuals. The agency provides a wide range of services to its customers, from underwriting to claims

Table 2 *Data categories.*

Category Importance level	*A High*	*B Medium*	*C Low*
Membership criteria	Item: 1 In use 2 Relevant 3 Updated 4 Raw data 5 Critical to work process	Data items belonging neither to A nor to C	1 Not in use ($x_1 = 0$) AND not updated ($x_3 = 0$) OR 2 Item result of operation ($x_4 = 0$)
Policy for handling the category	Retain in system	*Alternatives*: 1 Retain (A) 2 Retain, and change data structure 3 Remove, produce if needed 4 Remove, supply from external source 5 Remove, do not supply	Remove from system if: cost to retain > cost of removal (for criterion no. 2, produce if no affect to response time).

payment. The agency's information system, which was developed in-house, has been in operation for over fifteen years.

We selected one record from the agency's list of clients, and took it through all the steps of the methodology, including organizational goals, processes, boundaries, data item profile, categories and cost/benefit analysis.

The record as held by the agency contains facts such as name, address, telephone number, and other insurance data. For the purposes of our study, we describe it as follows:

Data	*Detail*
Fields in record	59
Length of record	246 Bytes
Age of record	10 years
Type of data	Raw data mainly
Number of occurrences	1000

(The number of occurrences refers to 1000 records randomly selected from the file.)

The data item in this case is a *field* in the record. Its profile includes the variables detailed in Table 1. These variables were assumed to be qualitative, taking the values of 0 or 1. For x_3, we assume that a field updated at least 70% of the time as being relevant.

Six work processes were identified, and were given equal importance. Thus, the range of values for perceived importance of an item is between 0 and 6, which is the number of different process possibilities. The processes are:

Process Number	*Description*
w_1	insurance design for client
w_2	client consulting
w_3	insurance sale
w_4	accounts receivable
w_5	information services
w_6	software maintenance

The full range of measurements on the data item are not shown.

Table 3 *Classification of items in case study.*

	Category A	*Category B*	*Category C*	*Total*
Number of fields	31	17	11	59
% of total	52.54	28.82	18.64	100
Length in bytes	125	84	37	246
% of total	51	34	15	100

The categories are defined according to the criteria set in Table 2. The classification of the items is shown in Table 3.

Table 3 depicts the number of fields in each category and their length in bytes as a measure of space. This initial analysis shows that category C has 18% of the fields and 15% of the record space. Since members of category C are subject to removal from the file, this space may be saved. On the other hand, Category A has 52% of the fields and 51% of the space, which indicates that approximately 50% of the data is certainly important.

5 COST/BENEFIT ANALYSIS

We show the cost/benefit in terms of three possible data decisions: (1) retain in the system, (2) remove from the system, and (3) remove with structural change.

1 *Retain in the system*: Data to remain in the system are: category A, those parts of category B which have no other source and are needed for automatic processing by an external user, or data for which it is impossible or impractical to implement a structural change. The following results shows that 70% of the record (number of fields or bytes) will remain in the system.

	Remain	*Remove*	*Structural Change*
Fields	41 (69.5%)	14 (23.7%)	4 (6.8%)
Bytes	175 (71%)	64 (26%)	7 (3%)

2 *Remove from the system*: Data to leave the system are category C and part of category B, which amounts to 24% of the fields and 26% of the file space. Removal costs included physical conversion cost, cost of program updates, and cost of user program compilation. These costs were estimated as 5 working days. A working day refers to the time that *one* employee works on the job.

3 *Structural change*: Structural changes were needed in four fields of the record. Thus, for example, a 1 byte field (having the value 0 or 1) appearing in only 2.8% of the population was replaced by a small file containing those records only. Using the sample of 1000 records we get: $1 \times 1000 \times 3\% = 30$ bytes taken, where 30×13 (key length) $= 390$ bytes needed; this represents a saving of 61%. In the case of the other three fields, structural changes led to a saving of 87.5% of space.

Total costs were estimated to be for 12 working days: 5 for removing data from the file plus 7 days for building new files and updating user programs.

Overall space saving for the sample data, after removal and structural changes were calculated to be 28.4% (the new storage space is 176 140 bytes instead of 246 000 bytes).

6 SUMMARY AND CONCLUSIONS

A methodology and a model for evaluating and handling data whose necessity is questionable were presented and suggested for practical use. Steps a manager may take to minimize the amount of information stored were demonstrated. We tackled the problem of identifying such data by using approaches developed in production and inventory management. The methodology and the model were validated by an empirical study of an actual life insurance agency.

ACKNOWLEDGEMENTS

The authors wish to thank the anonymous referees for their enlightening comments.

REFERENCES

1 Ackoff, R.L. (1967) 'Management misinformation systems', *Management Science*, Vol. 14, No. 3, December, pp. 147–156.
2 Ahituv, N. and Neumann, S. (1990) *Principles of Information Systems for Management*, (3rd edn.), Dubuque, Iowa, W.C. Brown.
3 Ahituv, N. (1980) 'A systematic approach toward assessing the value of an information system', *MIS Quarterly*, December, pp. 61–75.
4 Ahituv, N., Munro, M.C. and Wand, Y. (1981) 'The value of information in information analysis', *Information & Management*, pp. 143–149.
5 Aiuppa, T.A. and Wehrs, W.E. (1993) 'Insurer provided rating software based on Bop insurance applications at Wisconsin agencies', *Journal of Insurance Issues*, Vol. 16, No 2, October, pp. 1–19.
6 Bell, T., Witten, L.H. and Cleary J.G. (1989) 'Modelling for text compression', *ACM Computing Surveys*, Vol. 21, 4 December, pp. 557–591.
7 Eden, Y. and Ronen, B. (1993) 'The declining price paradox of new technologies', *OMEGA*, Vol. 12, No. 3, pp. 345–351.
8 Ein-Dor, P. and Carl, R.J. (1985) *Information System Management: Analytic Tools and Techniques*, New York.
9 Floyd, B. and Ronen B. (1989) 'Where best to system invest', *Datamation*, November 15, pp. 111–114.
10 Goldratt, E.M. (1991) *The Haystack Syndrome*, North River Press, Croton-on-Hudson, New York.
11 Kalfus, O., Ronen, B and Spiegler, I. (1994) 'To store or store: an assessment of information as inventory', Working Paper, Tel Aviv University, Faculty of Management.
12 Porter, M. and Millar, V. (1985) 'How information gives you competitive advantage', *Harvard Business Review*, Vol. 63, No. 4, July–August, pp. 149–160.
13 Ronen, B. and Palley, M.A. (1987) 'A topology of financial versus manufacturing management information systems', *Human Systems Management*, Vol. 7, No. 4, pp. 291–298.
14 Ronen, B. and Pass, S. (1992) 'Manufacturing management information systems requires simplification', *Industrial Engineering*, Vol. 24, 2 February, pp. 50–53.
15 Ronen B. and Spiegler, I. (1991) 'Information as inventory: a new conceptual view', *Information & Management*, Vol. 21, pp. 239–247.
16 Schonberger, R.J. (1982) *Japanese Manufacturing Techniques*, Free Press, New York.

17 Suzaki, K. (1987) *The New Manufacturing Challenge*, The Free Press, New York.
18 Venkatraman, N. and Zaheer, A. (1990) 'Electronic integration and strategic advantage: a quasi-experimental study in the insurance industry', *Information System Research*, Vol. 1, No. 4, pp. 377–393.

APPENDIX 1 THE METHODOLOGY AND THE DECISION MODEL

The methodology consists of a series of steps which help in structuring decisions and policies for handling data and IIP. The steps are:

1 Establish organizational and information system goals
2 Identify boundaries and data population
3 Divide the population into categories
4 Define data profile
The profile will generally contain parameters of two types:
- objective parameters about the data item,
- subjective parameters which are the result of an evaluation and indicate the relative importance of an item.

Table 1 depicts the parameters that make up the profile of a data item.
The profile is general and is therefore applicable to different levels of data items: field, record, file, and database.
5 List decision and work processes
The list of processes derives from the organizational and IS goals. The relative importance of an item in a work process is given by

$$\sum_{j=1}^{J} q_j w_j$$

$$\sum_{k=1}^{M} W_k = 1 \qquad M \geq J$$

where q is how critical the item is to a work process, w is the relative importance of a work process to all others, M is the all work processes in the organization, and J is the processes relevant to a given category.
Note: The relative weights (w) should be summed up to 1, by definition.
6 Estimate parameters for data items
7 Define importance of groups
Table 2 shows the three groups (ABC) and the criteria for membership in each category. This is a suggestion only which may be amended and adjusted for the particular system.
8 Classify data items into groups
9 Cost/Benefit analysis
The cost/benefit analysis is performed for each of the five alternatives for handling a data item (see also Table 2).
The alternatives are:
(a) Retain the item.
(b) Retain the item but change the data structure.
10 Operational decision
11 Overall cost/benefit calculation
12 Follow up

The methodology outlined above requires the formal definition of variables needed to assess the various alternative decisions and policies.

Variables

1 Item profile variables
$\{x_1, \ldots, x_5 ; q_j\}$ $j = 1, \ldots, J$ where J is the number of processes. This is a set of independent variables which are measured for a data item to determine its importance. These variables are detailed in Table 1.

2 Importance of an item
The importance of an item is defined as

$$h = \sum_{i=1}^{5} x_i + \sum_{j=1}^{J} q_j w_j$$

where h is the overall importance index of an item, and w_j is the perceived importance of a work process j.
The first term defines the *objective* importance as it indicates factual measures such as usage of the item (x_1), relevance (x_2), currency (x_3), and availability (x_4) and (x_5). The higher the index the more important the item.
The second term describes the *perceived* importance of the item; it derives from the perceived importance of the work process and how critical the item is to this process. The evaluation is subjective and done by the analyst together with the decision maker and/or the database administrator (DBA).

3 Categories
Following the ABC model, the importance categories are:
A highest degree of importance
B intermediate importance
C lowest importance
where $q\{A, B, C\}$.
The classification criteria are shown in Table 2.

4 Alternatives
A data item may be handled in one of five alternative ways $0 = \{1, 2, 3, 4, 5\}$. The five alternatives were defined in the previous section.

5 Decisions
The decision relevant to items remaining in the system are: D = {retain, remove, structure-change}

6 Cost Variables

C_s = cost of storage space the item takes up
C_u = cost of regular update, including software maintenance

C_t = cost of item in data retrieval

C_f = cost of removing an item from the system, where $C_f = C_c + C_p$

C_c = cost of converting current structure to another structure which does not contain the item

C_p = cost of converting a program to new data structure

C_y = cost of producing information (including hardware, software, and people)

C_b = cost of supplying an item from an external source

C_a = cost of lacking an item (negative benefit).

The cost of the alternative data policies are:

C_1 = cost of retaining an item in the system, where $C_1 = C_s + C_u + C_t$

C_2 = cost of retaining the item after a change in data structure. The cost of this alternative is cost of conversion + cost of supplying item from new structure—cost of continuing to supply the item from the old structure. $C_2 = C_c + C_p + C'_s + C'_u + C'_t + C'_1 - C_1$

C_3 = cost of producing the data item, defined as $C_3 = C_f + C_y$

C_4 = cost of supplying the removed item from another source $C_4 = C_f + C_b$

C_5 = cost of being unable to supply the item. This is the cost of not having an item upon demand, i.e., the damage caused if a request for the item cannot be met. Thus, $C_5 = C_f + C_a$.

7 *Benefit variables*

There are two aspects to measuring benefits (B): actual cost saving, and improving the value of information.