December 1997

Aggregated Data And Decision Processes:Limitations And Bottlenecks For Effrctive Support

Fred Lohman
Delft University of Technology

Henk Sol
Delft University of Technology

Aggregated Data And Decision Processes: Limitations And Bottlenecks For Effective Support

Fred A.B. Lohman and Henk G. Sol
School of Systems Engineering, Policy Analysis and Management
Delft University of Technology
PO Box 5015 - 2600 GA Delft
The Netherlands
Phone: +31 15 2787043, +31 15 2782318
Fax: +31 15 2783429
E-mail: F.A.B.Lohman@sepa.tudelft.nl, H.G.Sol@sepa.tudelft.nl

Executive Summary
As a result of the complexity in which modern organizations operate, the support of decision processes by means of aggregated data has been accelerated. Recently, a number of information systems and techniques have been launched which are aimed at improving the generation and analysis of the available data. For example, the so-called *data warehouses*, in which the operational data of the organization is put together for further analysis, are seen as one of the fastest growing areas. However, their contribution to decision processes has been questioned. We assume, on the basis of a case study and several indications found in literature, that the implementation and use of these systems and techniques do not automatically guarantee that decision processes will improve. In this paper, we present six bottlenecks which may influence the effectiveness of aggregated data:

1. Data integrity: data is unavailable, incorrectly structured or of poor quality
2. Physical realization of data aggregation cannot be accomplished
3. The information presented cannot be interpreted
4. Decision processes and data aggregation processes are not coordinated
5. Virtual organizations lead to dispersed data storage
6. Aggregated data have negative side effects

Based on these bottlenecks, we argue that at least four areas are relevant to the success of the aggregated data. Effectiveness is determined not only by the quality of the *transformation* of data into information, but also by the usability within the *decision process* and the (negative) side effects to the *organization* as a whole. Furthermore, one must give attention to the *data* itself and its structure. With these four areas in mind, we think that the described bottlenecks may be used as a sort of *checklist* to investigate to what extent the existing operational procedures deal with these limitations. More or less the same, these bottlenecks may be used as *agenda items* of a project plan in case a new management information system has to be built. In addition we developed an initial classification model which can be used to determine the impact of the bottlenecks in relation to the characteristics of the organization.

Abstract
As a result of the complexity in which modern organizations operate, the support of decision processes by means of aggregated data has been accelerated. Recently, a number of information systems and techniques have been launched which are aimed at improving the generation and analysis of the available data. However, their contribution to decision processes has been questioned.

We present six bottlenecks which may influence the effectiveness of aggregated data. These bottlenecks are then used to construct inductively a framework in which the impact can be specified. Further research will be focused on an elaboration of these bottlenecks and suggestions for possible solutions.

1. Introduction

Today, both private and public organizations are forced to respond quickly to the dynamic environment in which they operate. The intensification of competition, the emergence of new markets, the rise of network organizations, as well as the increasing demands of their customers, are among the challenges they have to deal with. As a result, more attention must be given to both the

quality and the efficiency of their decision processes. In this light, Huber (1984) argues that these processes will have to be designed and formalized as the central activity of an organization. In addition, one of the key issues to achieve an adequate decision process is the need for accurate and timely information.

Recently, the development and utilization of information systems designed to fill this need for information has been accelerated. The so-called *data warehouses*, in which the operational data of the organization is put together for further analysis, are seen as one of the fastest growing areas (McFadden 1996). Furthermore, techniques like *data mining* and *data visualization,* which are employed mainly to find patterns in the raw data, are among the new data analysis tools.

We assume, on the basis of a case study and several indications found in literature, that the implementation and use of these systems and techniques do not automatically guarantee that decision processes will improve. Many systems and techniques are introduced under the influence of technology push, without questioning or specifying the added value of the information generated or the impact on the organization. As a result, investments may not be successful, and the business need is still not fulfilled.

This paper gives an overview of six bottlenecks which may be responsible for this lack of improvement (section 2). Although various kinds of information may be used to support decision processes, we focus on information generated by the aggregation of data. In addition, an initial framework is presented which may make it possible to specify the impact of these bottlenecks (section 3). Finally, an overview is given of further research activities (section 4).

2. Bottlenecks In The Effective Use Of Aggregated Data

This section describes a set of six bottlenecks that showed up during our research. The description of each bottleneck is supported by the literature. We aim to show that in the case of data aggregation, each of these bottlenecks may limit the effectiveness of the generated information.

Data integrity: data is unavailable, incorrectly structured or of poor quality (1)

It can be argued that a first condition to aggregate data is the availability of accurate data (Dourish et al. 1993; Beheshti 1995). Nevertheless, in practice this condition is apparently difficult to meet (Lohman, Bots and Sol 1996-a; Lohman, Van Engers and Sol 1996-b). This can be very embarrassing, especially if decision makers are not informed about the quality of their data.

> *'Company 'Power' felt that a major strength of the corporate planning model was the accuracy of the data contained in the system. Decision makers relied on the accuracy and integrity of the data. Conversely, the major failing of Rail's system was its lack of data integrity.'* (Molloy and Schwenk 1995, p.299)

In the first place, the data quality is dependent on the usability for those who enter this data. If these users only 'feel obliged' without gaining an additional value they feel less motivated and the data becomes less suitable for aggregation. But if there is a good reason to monitor over this quality, a second complication arises. The data generated is seen more as an instrument of power than as neutral information (Feldman and March 1981; Alsyne, Brynjolfsson and Madnick 1995). As a result, the owners may be unwilling to share this information.

Third, the structure of the data stored also plays an important role. Lohman (1995) found that the project-oriented data structure applied within an engineering company limited the re-use of information. The underlying product structure was neglected, even though this information could be extremely useful for re-use and evaluation.

Physical realization of data aggregation cannot be accomplished (2)

The process of aggregating data as applied in descriptive statistics does not usually present any serious complications. However, if the aim of the aggregation process is to find relations between the data elements results may be disappointing. Some people would argue that it is impossible to get results, especially if these relations are directed towards the underlying business processes of the data stored (Sol 1982; Schrijver 1993; De Jong 1992; Reuyl 1982).

> *'The analysis of the situation in the Multiple Store Company leads to a few significant relationships between exogenous and endogenous variables at article level. By aggregating*

attributes of articles over time, over location or over other characteristics, not many relationships can be accepted for decision-making at a global level. Analysis of time series suggest relationships which cannot found back in a more detailed examination. Thus the currency of information is of little importance.'(Sol 1982, p.191)

'We expect that using frequent trip status updates for other hauliers that also have to contend with major deviations in trip component duration's, will not improve trip executions performance considerably.' (Schrijver 1993, p.188)

Less complicated, but just as crucial as the problem mentioned above, is the impossibility of aggregating qualitative (soft) data. As Leidner and Elam stated:

'EIS have still largely failed to incorporate the need for soft information.' (Leidner and Elam 1995, p.659)

The information presented cannot be interpreted (3)
In addition to the physical realization, we would argue that one of the most important bottlenecks is the impossibility of interpreting information generated by the aggregation process. With the increasing power and potential of information systems, many organizations are taking information which was previously unavailable or distributed only on paper and making it available electronically. Especially data warehouses claim to gain insight in their organization or environment. However, a primary focus of these systems is the ease of information retrieval (Dourish et al. 1993). We believe that without the possibility to interpret this information, the additional value is rather limited. In such cases information is, in terms of decision support, used mainly to trigger for a decision process or to evaluate that process.

But techniques like data mining also suffer from this type of problem. The strength of these techniques is the ability to execute thousands of calculations, based on certain algorithms. The result of these calculations is the extraction of implicit, previously unknown, and potentially useful information from data (Piatetsky-Shapiro and Frawley 1991). However, the detection of an unexpected relation or pattern will have to be translated into a concrete action or strategy in order to be useful.

'It is noted then that even with a data set which is apparently complete, it may be impossible to clearly determine what interpretation may be drawn from any patterns which do emerge.' (Cornish and Elliman 1995, p.351)

Decision processes and data aggregation processes are not coordinated (4)
In addition to the bottlenecks within the aggregation process, problems may also arise between decision process and aggregation process. Most of these problems can be reduced to the lack of coordination between the desired information and information presented. Feldman and March (1981) argue that the separate execution of both decision process and aggregation process may play a role here. This could even clarify such phenomena as technology push (McQuillen 1995; Millet and Mawhinney 1992; Beroggi 1996) or information overload (El Sawy and Pauchant 1988; Huber 1984; Ackoff 1967).

'Since the information-gathering functions are typically separated from the information-using functions of organisations, incentives are modest for gatherers to avoid overloading users.' (Feldman and March 1981, p.415)

Furthermore, the unstructured execution of decision processes may also lead to the inability to coordinate these processes. It is argued by several authors (March 1987; Cyert and March 1992; De Bruin and Ten Heuvelhof 1995) that organizational decision making often has less coherence than decision theory attributes to it. Problems, solutions, and actions are frequently only loosely coupled, or connected by their simultaneity rather than their consequentiality. As a result, specification of the information desired (e.g. time and content) is hard to achieve.

Virtual organizations lead to dispersed data storage (5)
Information-intensive organizations are inclined to operate as virtual organizations. The improvement of information technology to support dispersed workgroups offers the opportunity for people to work at any time and anywhere around the world (Upton and McAfee 1996; Nijhuis 1996). This gives employees more flexibility in organizing their work and may reduce property costs. On the other hand, virtual organizations also result in complications. Handy (1995) argues that these organizations cannot be managed by control, but will have to be based on trust. If this is true, what consequence will have this effect on the control of the dispersed data storage? Aggregating data derived from different departments is already hard to achieve, and virtual organizations will expand this problem to the individual level.

Aggregated data have negative side effects (6)
The bottlenecks mentioned above are problems which, in some way, obstruct the realization of effective aggregated data. This effectiveness depends on the extent to which this information provides additional value for the execution of decision processes. However, the availability of this information may also lead to certain negative side effects. Millet and Mawhinney (1992) maintain that decision makers may focus too much attention on the measurable dimensions of business. The possibility of data warehouses to 'drill-down' through layers of aggregation might result in spending too much of their time in analyzing low-level concerns. But even more important, since lower echelons in the organization know that certain dimensions are monitored, they may shift their agenda and shorten their time horizon.

In addition, Stickel (1995) argues from a macro-economic point of view that the availability of more information will reduce the risk a decision maker may take. As a consequence, a reduction of the market share of the organization may follow.

3. Areas Related To Aggregated Data

The bottlenecks described in section 2 make it clear that focusing exclusively on the aggregation process itself may be too narrow to explore the effectiveness of the aggregated data. Effectiveness is determined not only by the quality of the transformation of data into information, but also by the usability within the decision process and the (negative) side effects to the organization as a whole. Furthermore, one must give attention to the data itself and its structure. These four areas, including the described bottlenecks, are shown in figure 1.

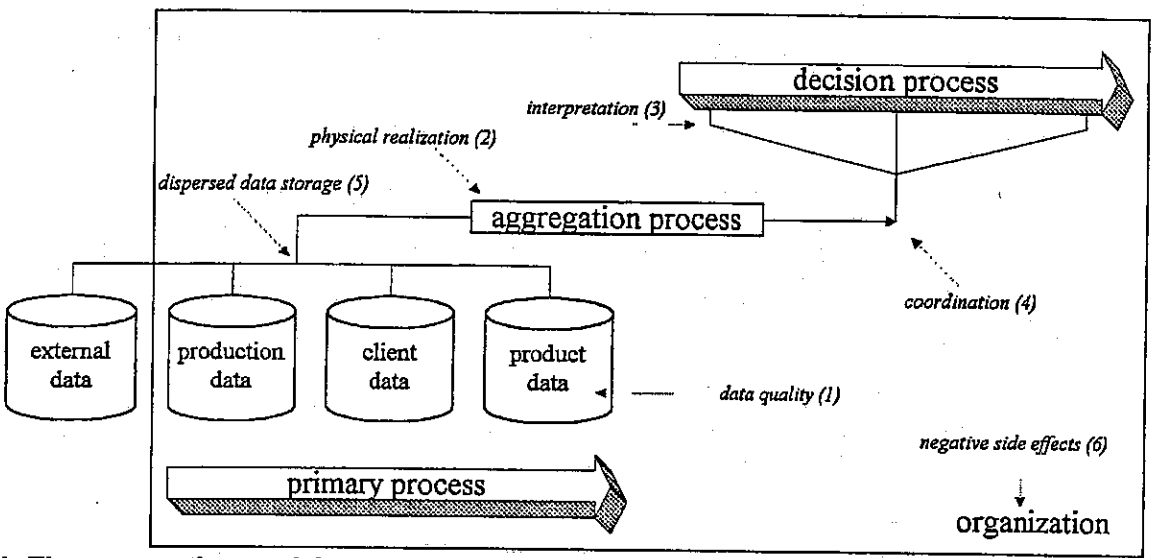

Figure 1: The aggregation model

With these four areas in mind, we think that the described bottlenecks may be used as a sort of *checklist* to investigate to what extent the existing operational procedures deal with these limitations. During our second case study we found that several problems could easily be categorized and specified by using these bottlenecks (Lohman, Van Engers and Sol 1996-b).

Similarly, these bottlenecks may be used as *agenda items* of a project plan in case a new management information system has to be built. It forces one to pay attention not only to the physical

realization of the system, but also to the added value to the organization as a whole. Nevertheless, it may be quite difficult to define the possible *impact* of the bottlenecks beforehand. This impact, however, is directly related to the success of the developed system.

We would argue that the values of some of the variables which describe the four areas shown in figure 1, may determine the impact of the bottlenecks. On the basis of the individual characteristics of these bottlenecks, we developed an initial classification model which presents important variables related to our research area. Although some variables cannot be inferred from the bottlenecks directly, we believe that this descriptive model can be used to specify these bottlenecks in more detail. A further explanation of these variables is given in Lohman, Bots and Sol (1996-a).

AREA	VARIABLES	VALUES (examples)
1. Organization	structure	flat, hierarchical, network
	branch	public, private
	size	number of workers, returns
	history	reorganizations, mergers
	culture	leadership, flexibility
	primary process	engineering, financial
	work force	age, skills
2. Decision process	domain	products, production, marketing, development
	level	micro, meso, macro
	horizon	crisis, short-term, long-term
	type	structured/ routine, unstructured/ non-routine
	information need	triggering, supporting, evaluating
	stakeholder	management, external parties
3.Aggregation process	stage	collecting, transforming, presenting, and interpreting data
	technology	DSS, EIS, data warehouse
	initiators	IT-specialists, decision makers, staff
	method	descriptive statistics, simulation, trend analysis, visualization, data mining
4. Data	structure	hierarchical, relational, object-oriented
	type	quantitative, qualitative
	content	product data, client data, production data
	location	central, dispersed
	quality	number of errors, electronic availability, complete

Table 1: Classification model

4. Directions For Further Research

On the basis of a case study and several indications found in literature, we conclude that at least six types of bottlenecks may limit the effectiveness of aggregated data to support decision processes. With the notification that accurate data is needed to survive in today's dynamic environment, further research with respect to these bottlenecks seems necessary.

This research will focus on two areas. First we want to examine, by the use of case studies, the extent to which the classification model in table 1 can be used to analyze the bottlenecks that obstruct the effective aggregation of data. On the basis of that examination, we aim to construct a methodology for the design of effective aggregation processes.

Second, we will investigate to what extent the third bottleneck described in section 2, the interpretation of aggregated data, can be resolved. At the moment, for example, we are analyzing the possibilities of supporting decision processes by means of simulation tools. By using aggregated and disaggregated data as input for these tools we assume that the output enables decision makers to better understand the consequences of their decisions.

References

Ackoff, R.L. *Management misinformation systems*, Management Science, Vol. 14, No. 4, pp. B-147-56, 1967.

Alsyne, M. van; Brynjolfsson, E.; and Madnick, S. *Why not one big database? Principles for data ownership*, Decision Support Systems, Vol. 15, pp. 267-84, 1995.

Beheshti, H.M. *Downsizing with executive information systems*, Industrial Management + Data Systems, Vol. 95, No. 5, pp. 18-22, 1995.

Beroggi, G.E.G. *Supporting centralized and decentralized organizational decision making with multimedia systems*, Advisory group meeting on the development of decision aiding tools, International Atomic Energy Agency, 1996.

Bruijn, J.A. de and Heuvelhof, E.F. ten *Netwerkmanagement. Strategieën, instrumenten en normen*, Lemma Uitgeverij B.V., Utrecht, 1995. (In Dutch)

Cornish, T.A.O. and Elliman, A.D. *What has Mill to say about data mining?*, pp. 347-53, Proceedings of the 11th Conference on Artificial Intelligence for Applications, IEEE Comput. Soc. Press, Los Alamitos, 1995.

Cyert, R.M. and March, J.G. *A Behavioral theory of the firm*, Blackwell Publishers, Cambridge, 1992.

Dourish, P.; Bellotti, V.; Mackay, W.; and Chao-Ying Ma *Information and context: lessons from a study of two shared information systems*, pp. 42-51, Conference on Organizational Computing Systems, ACM, New York, 1993.

El Sawy, O.A. and Pauchant, T.C. *Triggers, templates andt twitches in the tracking of emerging strategic issues*, Strategic Management Journal, Vol. 9, No. Sep-Oct, pp. 455-73, 1988.

Feldman, M.S. and March, J.G. *Information in organisations as signal and symbol*, Administrative Science Quarterly, Vol. 26, pp. 171-86, 1981.

Handy, C. *Trust and the virtual organization*, Harvard Business Review, No. May-June, pp. 40-50, 1995.

Huber, G.P. *The nature and design of post-industrial organizations*, Management Science, Vol. 30, No. 8, pp. 928-51, 1984.

Jong, R. de *An adaptive trip planning support environment*, R. de Jong, Delft, 1992.

Leidner, D.E. and Elam, J.J. *The impact of executive information systems on organizational design, intelligence, and decision making*, Organization Science, Vol. 6, No. 6, pp. 645-64, 1995.

Lohman, F.A.B. *Verbetering van de informatievoorziening in het proposaltraject bij Comprimo*, TU Delft, sectie Industriele Organisatie, Delft, Reportnr. 333II, 1995. (In Dutch)

Lohman, F.A.B.; Bots, P.W.G.; and Sol H.G. *Ondersteuning van het topmanagement met behulp van informatietechnologie*, Technische Bestuurskunde, Delft, 1996-a. (In Dutch)

Lohman, F.A.B.; Engers, T.M. van; and Sol, H.G. *Tussenrapportage informatie-aggregatie Belastingdienst*, Technische Bestuurskunde, Delft, 1996-b. (In Dutch)

March, J.G. *Ambiguity and accounting: the elusive link between information and decision-making*, Accounting, Organizations and Society, Vol. 12, pp. 153-68, 1987.

McFadden, F.R. *Data warehouse for EIS: some issues and impacts*, pp. 120-29, vol II, Proceedings of the twenty-ninth Hawaii International conference on System Sciences, IEE Comput. Soc. Press, Los Alamitos, 1996.

McQuillen, N.L. *Knowledge Queries for Learning and Database Discovery*, pp. 1-28, ICIS Doctoral Consortium, Amsterdam, 1995.

Millet, I. and Mawhinney, C.H. *Executive information systems, a critical perspective*, Information & Management, Vol. 23, pp. 83-92, 1992.

Molloy, S. and Schwenk, C.R. *The effects of information technology on strategic decision making*, Journal of Management Studies, Vol. 32, No. May, pp. 283-311, 1995.

Nijhuis, R.J.H. *Working apart together*, Technische Bestuurskunde, Delft, 1996.

Piatetsky-Shapiro, G. and Frawley, W.J. *Knowledge discovery in databases*, The AAAI Press, California, 1991.

Reuyl, J.C. *On the determination of advertising effectiveness: an empirical study of the German cigarette martket*, Stenfert Kroese, Leiden, 1982.

Schrijver, P.R. *Supporting fleet management by mobile communications*, P.R. Schrijver, Delft, 1993.

Sol, H.G. *Simulation in information systems development*, Krips repro, Meppel, 1982.

Stickel, E. *Werttbewerbsorientierte informationssysteme und produktivitatsparadoxen*, Wirtschaftsinformatik, Vol. 37, No. 6, pp. 548-57, 1995. (In German)

Upton, D.M. and McAfee, A. *The real virtual factory*, Harvard Business Review, No. july-august, pp. 123-33, 1996.