Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 1995 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

December 1995

A Generic Approach to Information Systems Architecture to Support Quantitative Methods in Socio-Human Research

Henry Linger
Monash University

Arkady Zaslavsky
Monash University

Frada Burstein
Monash University

Chris Grigsby
Monash University

David Allen
Monash University

A GENERIC APPROACH TO INFORMATION SYSTEMS ARCHITECTURE TO SUPPORT QUANTITATIVE METHODS IN SOCIO-HUMAN RESEARCH

Henry Linger[+], Arkady Zaslavsky[°], Frada Burstein[+], Chris Grigsby[+], David Allen[+]
[+]Department of Information Systems, Monash University
[°]Department of Computer Technology, Monash University
Email: {H.Linger, A.Zaslavsky, F.Burstein}@Monash.edu.au

Socio-human research involves conducting extensive and expensive surveys and subsequent statistical processing of accumulated data. Constructing the comprehensive and reliable questionnaires is a non-trivial task which is performed by a social science researcher with little support from database and information systems technology. This paper proposes and discusses a generic approach to information systems architecture to support quantitative research in socio-human systems based on a set of conceptual models representing various stages of information systems development. A number of advanced information systems concepts have been applied to increase flexibility and efficiency of the approach. An instantiation of an approach is illustrated by an example from epidemiological studies.

1 Introduction

Any system dealing with human activity, behavior or organisation is by definition open, complex, uncertain and incomplete in terms of the information available about the system. This is further compounded if the ontology adopted is nominalism (Burrell and Morgan, 1979). This position treats reality as subjective, interpretive and is influenced by a wide variety of phenomena, including language, culture, symbols and names, which are, in general, not well understood (Avison and Fitzgerald, 1988).

Quantitative research (QR) is a technique which is used to support social sciences. The main tool of QR is a data collection instrument administered to subjects of the study. The aim of such studies is to collect data, store and manipulate it to report on the phenomena. Such a cycle is consistent with the definition of information systems presented by Buckingham et. al (1987), where an information system is viewed as a system which assembles, stores, processes and delivers information relevant to an organisation (or to society), in a way that is accessible and useful to those who wish to use it. An information system is a human activity (social) system which may or may not involve the use of computer systems.

Information technology has had a significant impact on quantitative research with the advent of personal computers (PC). The role of advanced technologies (eg, expert and knowledge based systems) in social research was also acknowledged (Benfer, Brent, Furbee, 1991). Such expert systems address areas of theory building, defining research problem, designing the research, collection of qualitative data and data analysis, as well as teaching, applications and dissemination of research findings (Benfer, Brent, Furbee, 1991).

Each phase of the qualitative research is dependent on the previous phases especially in terms of data. What is a glaring missing is a meta-level integration of the technology of each phase and a generic approach to building information systems able to support QR. Such integration and generalisation would utilise additional technology to provide an active and intelligent support for research in social sciences.

This paper proposes an architecture for such an integrated information system development highlighting the three existing areas but reconceptualising them within a model of the process of QR. The architecture will identify where additional technology can be effectively used to intelligently support the different activities that are part of QR. A conceptual model of a survey-driven database is then proposed and discussed. It is believed that this model is general enough to support QR in general. We illustrate the instantiation of the architecture to a large, longitudinal epidemiological cohort study relating to blood borne diseases.

2 An Architectural View on Information System for Quantitative Research

In the context of QR document preparation generally applies to the design and production of the data collection instrument. The researcher has the ability to directly design and prepare the instrument for publication.

Information technology, however, offers a number of other facilities that can be potentially utilised by QR. Pattern matching is an important area. A major concern of researchers is to discover underlying trends in the data sets and to investigate these trends in details (Fowler, 1988). Pattern matching can provide insight into trends in the data. Information technology can employ techniques to identify trends ranging from artificial neural networks through to algorithms, logical systems and knowledge based approaches. Identification of trends and general dependencies in data can create a basis for formulating rules and constraints as event-condition-action (ECA) constructs (Chakravarthy et al, 1993; Navathe et al, 1992) and thus making the database active.

We propose an architecture of information systems for quantitative research which is depicted in figure 1. The main idea of the proposed approach is based on the assumption that most of survey-based studies in socio-human research have much in common and can be viewed from the generic perspective. This assumption allows to create a framework for an information system architecture supporting social sciences research which may incorporate heterogeneous technological components, eg., databases, knowledge bases, tools for statistical analysis of survey data, etc.

On a pragmatic level survey-driven studies can be viewed as consisting of three phases:

- survey design and administration;
- storage of collected data;
- statistical analysis of the collected data.

Such an approach is strictly oriented on one particular study, is usually designed and administrated by a researcher that will be responsible for a statistical analysis of data as well. It does not allow any sort of generalisation or re-use of the collected materials as well as has a limited flexibility from the viewpoint of the questions used in questionnaires. The flaws of such an approach are in direct proportion with the experience of the researcher. There is a minor possibility for an information system which will copy such an approach to be self-consistent and support a better quality data processing for survey-based studies.

To generalise a process of studies in socio-human systems and provide a quality information system support it should be considered on a metalevel. Three areas of generalisation should be taken into consideration: study design, data collection and data analysis. Each of this areas can be described by a generic model that will be valid regardless of the particular researcher or a nature of the research. Furthermore, the knowledge describing the best expertise be elicitated, processed and organised at a meta-level to support a survey-based research with adequate rules and procedures. Large collections of historical data also can be used to discover general dependencies and trends in socio-human systems should (Yoon & Kerschberg, 1993). Accumulation of prior knowledge have always been mentioned as a week point of social science research (Rosenthal, 1984). However there are a number of expert systems constructed to assist in uncertain situations during social research with the human operator as a 'pattern recognition engine'(Benfer, Brent, Furbee, 1991).

Data Collection Model provides consistency between content of the questionnaire and a data model that exists in an information system the data will be stored in. Data Analysis Model associates extracted data with the correspondent statistical techniques. The kind of questions in a questionnaire and their structure have to be relevant to the purposes of the analysis the study is done for. This is responsibility of the Study Design Model to monitor the link between these two parts of study. This approach is general across all survey cased study. This gives an opportunity to introduce one more level into generic framework. It can be called conceptual level. Conceptual level consists of three models: Conceptual Survey Model, Conceptual Data Model and Conceptual Statistical Model. Each of this models represents a component of an information system. The conceptual data model will be considered in more detail below.

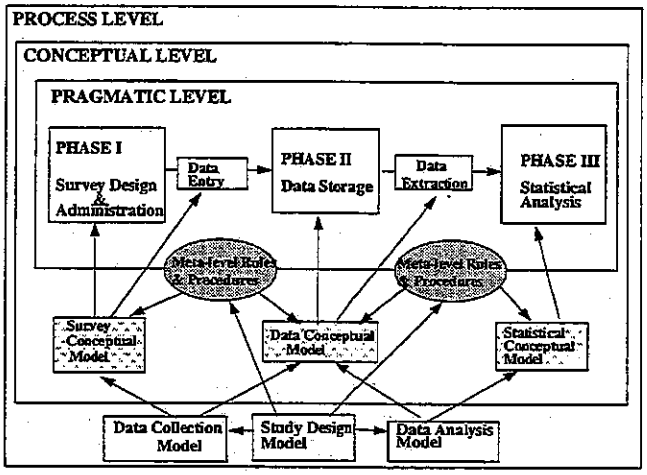

Fig. 1. Generic IS Framework for Socio-Human Systems

The conceptual model of a database to support QR is depicted in figure 2. It is designed to address and provide two dynamic abilities, namely:

- Undetermined current or future purpose. How the system will be used is often not determined or changes are to be introduced after a system has already been constructed.
- Undetermined current or future representation. How the system is represented both conceptually and internally to a computer system often endures over time.

The model is defined as consisting of objects/abilities that represent either physical or conceptual ideas in the problem domain. These objects are all defined identically, separated only by their object name, and provide two fundamental dynamic characteristics mentioned above.

This model is driven by the following assumptions:

- an individual can be involved in many relationships including a self-referential relationship (SRR) corresponding the individual only with no other people;
- these relationships can be based on situations consisting any combination of (including none of) time period information, geographic location information, or merely social and/or time meaning information;
- transactions can occur that indicate the activities of the people/person within the relationship, located by the attributes defined in the given situation.

The following section considers epidemiology as an example of survey-based studies and the conceptual database model will be discussed in the context of these studies.

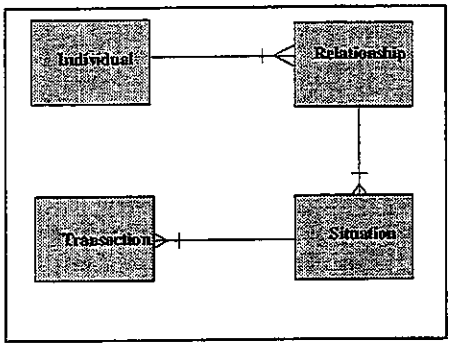

Fig.2. A conceptual data model for survey based systems

3. ADES Application

This section discusses implementation of an approach proposed in previous sections in a particular application domain. A typical example of a socio-human system is epidemiology. Epidemiology is the study of the distribution and determinants of disease and injuries in various populations. It is an example of population medicine, a study of health and disease on a population level rather than on an individual or clinical basis. It is concerned with the frequencies and types of illnesses and injuries inflicting people grouped by many denominators, to gain insight into the factors that influence their distribution. The lack of information systems support for epidemiology is reflected in the current problems. The design and implementation of epidemiological studies is currently plagued by difficulties of varying importance and effect.

The conceptual representation of the epidemiological domain corresponds to the database model in figure 2. The nature of how "individuals" (eg., "subjects" or "significant others",) "relationships", "situations" and "transactions" are interconnected can be simplified for implementation purposes by adopting several assumptions that while simplifying the implemented prototype, are still consistent with the epidemiological model and conceptual representations. These assumptions are listed below:

- study subjects and significant others, being people capable or having the same personal information are generalised and called Individuals.
- the semantic meaning behind the subject - transaction link was broken down due to a requirement that any transaction must have a situation and relationship, even if the relationship is self referential.
- the meaning behind the subject - situation link, for much the same reasons as before, has been removed. A subject in a situation alone with no transaction or relationship data having significance.
- for a complete transaction to be recorded, a path must exist from the subject, through relationship and situation, to transaction.

The Active Database for Epidemiological Studies (ADES) application prototype has been implemented using Microsoft Windows and Access DBMS software. Though the prototype obviously has a number of drawbacks, the lesson learned justifies the approach proposed and prompts a number of potential improvements.

4 Conclusions and Future Work

For the first glance, the nature of information system should be invariable to the application domain. However, the domain of socio-human systems requires a specific attention because not only the information itself is considered as a valuable resource, but the structure and the ways of presenting this information to users are of the paramount importance.

This paper proposed an architecture for the advanced information systems which can support survey-based research in social sciences. This architecture provides a framework to accommodate and utilise a number of heterogeneous information systems technologies based of the general data structures throughout the whole study. An insanitation of the architecture was illustrated with the epidemiology as an application domain. A database system was implemented and briefly discussed with an emphasis on a generic nature of a conceptual model suitable for other application domains like psychology, agriculture, business decision making, etc. The future work will concentrate on implementing other core components of the proposed architecture and the knowledge-based extensions to it.

References

Avison, D. E., and G. Fitzgerald. (1988). *Information Systems Development: Methodologies, Techniques and Tools*,. Oxford: Blackwell Scientific Publications.

Buckingham, R. A, R. A. Hirschheim, F. F. Land, and C. J. Tully (eds). (1987). *Information Systems Education: Recommendations and Implementations*,. Cambridge: CUP.

Burrell, G., and G. Morgan. (1979). *Sociological Paradigms and Organisational Analysis*,. London: Heinemann.

Benfer, R.A.; Brent, E. E.Jr.; Furbee, L. (1991). *Expert Systems*, Series: Quantitative Applications in the Social Sciences, SAGE Publications.

Chakravarthy, S., Krishnaprasad, V., Tamizuddin, Z., Lambay, F. (1993) *"A Federated Multi-media DBMS for Medical Research: Architecture and Functionality"*, Technical Report UF-CIS-TR-93-006, Department of Computer & Information Sciences, University of Florida, January, 1993

Fowler, F.J., Jr. (1988). Survey Research Methods, SAGE Publications, London.

Navathe, S.B., Tanaka, A.K., Chakravarthy, S. (1992) "Active Database Modeling and Design Tools: Issues, Approach, and Architecture", Data Engineering, Vol.15, No.1-4, December.

Rosenthal, R. (1984). *Meta-Analytic Procedures for Social Research*, Newbury Park, CA:Sage.

Yoon, J.P., Kerschberg, L. (1993) A Framework for Knowledge Discovery and Evolution in Databases, IEEE Transactions on Knowledge and Data Engineering, Vol 5, No 6, December, 973-979.