Towards a Systematic Evaluation of Personal and Small Group Information and Knowledge Management

Mark GREGORY
Finance and Operations. ESC Rennes School of Business
35065 Rennes Cedex. France
Mark.gregory@esc-rennes.fr

and

Dr. Mario NORBIS
Department of Management. Quinnipiac University
Hamden, Connecticut 06518, USA
Mario.norbis@quinnipiac.edu

5th International Conference on Cybernetics and Information Technologies, Systems and Applications: CITSA 2008

ABSTRACT
This research presents the hypothesis that individuals working in groups should be encouraged and educated to make better use of the available tools, and that the tools themselves should evolve into or be replaced by better ways of representing information and knowledge. As a first step, it is necessary to classify and evaluate the effectiveness of existing tools and techniques. This paper summarizes current trends in the academic and practitioner’s literature in the areas of knowledge representation and communication by individuals and small groups. The paper also suggests that a judicious mix of existing and emerging tools, coordinated by semantic desktop approaches, will permit evolution or revolution in the management of individual and shared information and knowledge.

1. INTRODUCTION
Knowledge and information workers, working as individuals within virtual team structures, acquire information, which they store in a large number of arbitrarily complex ways: some being paper-based, but increasingly computer-based.

Over the last two decades a significant number of computer-based tools (sometimes referred to as Personal Information Managers or PIMs) [Teevan 2006] have been created in order to assist in the storage and management of such information.

So far it is arguable that no tool has achieved ubiquity, whether measured in terms of the extent of its use or the generality of its application. Also there is little consensus about the most appropriate methodology for evaluating information technology or information systems [Beynon-Davis 2004] and so far Information Systems evaluation remains underdeveloped and undermanaged [Love 2004].

The underlying hypothesis of the research-in-progress presented in this paper is that individuals working in groups should be encouraged and educated to make better use of the available tools, and that the tools themselves should evolve into or be replaced by better ways of representing information and knowledge.

This study is further motivated by perceived deficiencies in current data management paradigms and other paradoxes, some of which are highlighted here.

Deficiencies of the current data management paradigms. Current database concepts materialized in software packages, fail to handle properly the complex, non linear relationships among data items or families of them. Examples of this situation arise in diverse areas as new data organization to be utilized at museums information centers [De Vorsey 2006].

The productivity paradox. Paul Strassman [Strassman 1999] has suggested a productivity paradox: there is little or no correlation between the amount which organizations spend on IT and their profitability.

Uncorrelated investment. The authors’ experience suggest that there may be relatively little correlation between the investment made in personal productivity tools and those oriented to small-group productivity.

The usage question and the expressiveness of information. This research aims to examine the reasons why what should be widely-used techniques, are not. It is our belief that computer users voluntarily sacrifice freedom in favor of structure in order to facilitate storage, retrieval, and especially querying and communication; but they still do not achieve the level of precision and communication that they strive for.

Towards a better Group Information Management paradigm. In this context, a new paradigm should provide the user with the flexibility to accommodate a variety of interactions among data items maintained by individuals or groups: by defining structures and mechanisms to better accommodate particular needs while keeping a more rigid framework to guarantee consistency while allowing for expressiveness of the data to be present.
2. APPROACHES TO DATA AND INFORMATION MANAGEMENT

The state of the art of personal and small group data and information management can be described largely as practitioner-based. It is difficult to identify a particularly coherent and extensive academic literature in the area. With this idea in mind, this section will present the most common tools because they constitute the bases on which we can build more complex (or simple) structures for information and knowledge management.

General Spreadsheets. VisiCalc originated by Dan Bricklin, has turned the personal computer from a hobby into a business tool. VisiCalc from Apple was followed by Lotus 1-2-3 (IBM PC 1983) went on to become an application that was so compelling, people would buy a particular computer just to own it. It is suggested that today a typical knowledge worker possess from the 100’s to the 1000’s of spreadsheets. They enforce a degree of structure onto data such as lists, that result in their strengths in the presentation of information. There are however many problems associated with spreadsheets. See Burnett [Burnett 2001] and Peyton Jones [Peyton 2003] for a discussion and suggestions of ways forward; and the discussion of Functional Spreadsheets below.

Functional Spreadsheets. The Functional Spreadsheet is a new idea by co-author Mark Gregory, which simplifies and restricts the scope of spreadsheets, to make them capable of formal representation. The idea is based on an insight documented by Simon Peyton Jones, [Peyton 2003] but goes beyond this approach which considers only developments of Excel. A flavor of the concept is provided by this screenshot of a storyboarded interface.

Outliners. Outliners, which are typically used for computer programming and the management of task lists, are hierarchical ways to display related items of text to graphically depict their relationships. Microsoft Office Word supports a very useful Outline mode. The best examples of more specialized outline programs are NetManage ECCO Pro and Micro Logic’s Info Select 2007.

Relational databases. The currently dominant paradigm in the storage of structured data on computers, the relational paradigm [Codd 1970] enables arbitrary manipulation: that is to say that queries can be defined which will always have an answer. However, the data is constrained to appear in normalized relations or sets or entities [Chen 1976].
Many PIMs store their data internally as relational tables; a few make the contents of those tables visible directly for end-user querying.

**Object-oriented databases.** The object oriented data paradigm [Kim 1990] allows other kind of associations between entity types; further, the structure of an entity is much richer than the normalized model. However, the approach has disadvantages; there is no software known that permits end users to create and manipulate such data bases and it is not always possible to obtain an answer to a question. A hybrid object-relational database has been proposed and implemented. The best known object-relational database is Oracle.

**XML documents.** Extensible Markup Language (XML) is a simple, very flexible text format derived from SGML (Standard Generalized Markup Language). Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere. XML is an excellent data interchange mechanism, and is very widely implemented.

**RDF and OWL.** Based on XML, the Resource Description Framework (RDF) and Web Ontology Language (OWL) form the basis of the semantic web [Berners-Lee 1998]. The Resource Description Framework (RDF) integrates a variety of applications from library catalogues and world-wide directories to syndication and aggregation of news, software, and content to personal collections of music, photos, and events using XML as interchange syntax.

**The Semantic Web.** The semantic web is an evolving extension of the World Wide Web in which web content can be expressed not only in natural language, but also in a form that can be understood, interpreted and used by software agents, thus permitting them to find, share and integrate information more easily. It derives from Berners-Lee's vision of the Web as a universal medium for data, information, and knowledge exchange [Berners-Lee 1998].

The current state of the art for the semantic web is described as a set of emerging large-scale research prototypes aimed at the corporate level. The best-known example to date is MIT’s Simile project (http://simile.mit.edu/) and the Haystack prototype.

**The Semantic desktop.** In computer science, the Semantic Desktop is a collective term for ideas related to changing a computer’s user interface so that data is more easily shared between different applications or tasks and so that data that once could not be automatically processed by a computer could be now. It also encompasses ideas about being able to automatically share information between different people. This concept is very much related to the semantic web but is distinct: the semantic Web represented at the small-group level.

**Existing PIMs and GIMs.** PIMs (Personal Information Managers) and GIMs (Group Information Managers) are software programs that help individuals and small teams to manage their shared data. This family of software applications has existed almost since the birth of the personal computer age, in the early 1980s. It has given rise to a large number of programs, to which much of the world have remained indifferent. About the only personal and group information management package which sees widespread use is Microsoft Outlook. Poor [Poor 2001] presents a personal information manager (PIM) Info Select, in which the user stores a variety of items in an outline structure and adding Boolean logic operators makes easier to narrow the searches. Recent academic research in the area of Personal Information Management has been summarized in the first 2006 issue of the Communications of the ACM which was devoted to personal information management. The introductory article by Teevan [Teevan 2006] and the article by Whittaker [Whittaker 2006] are of note. However, as T.D. Wilson points out [Wilson 2002], much of what passes for knowledge management is no more than information management.

3. **CLASSIFICATION OF APPROACHES TO SMALL-GROUP KNOWLEDGE MANAGEMENT (KM)**

**Background.** Art administrators managing information perceive themselves to be following a highly intuitive process based on personal experience or expertise by means of unstructured processes [Zach 2005]. This experts approach, while serving well in some familiar situations, may be of more limited success in situations where the complexity and the potential impact of the task or decision are high.

**Thesaurus based approach.** De Vorsey et al [De Vorsey 2006] present the development of a local thesaurus to improve access to the anthropological collections of the American Museum of Natural History. They construct a poly-hierarchical, mono-
lingual local thesaurus based upon the terminology used to catalogue the collection.

**PIM approach.** Teevan [Teevan 2006] suggests that the history of personal information management (PIM) dates from 1945 [Bush 1945] when a device was envisioned for individuals to store books, records, and communications, and which was to be mechanized so that it may be consulted with speed and flexibility. Bush called this device “Memex”.

**Methodological issues.** Data management is normally regarded as a domain of computer science, information management as a domain of information systems, and knowledge management is variously claimed because there is comparatively little accord on what it constitutes. There is more agreement on knowledge representation as the discipline which most clearly concerns itself with the presentation and management of information and knowledge stored on computers. Management Information Systems (MIS) is a still-young field in rapid change and turmoil [Palvia 2003].

**Case study methods.** Benbasat et al [Benbasat 1987] defines and discusses the case research strategy. Criteria for the evaluation of case research are established and several characteristics useful for categorizing the studies are identified. This study concludes with examples of research areas that are particularly well-suited to investigation using the case research approach.

**Participatory action research.** Baskerville & Wood-Harper [Baskerville 1996] review the origins, techniques and roles associated with action research into information systems (IS). The paper summarizes the rigorous approach to action research and suggests certain domains of ideal use (such as systems development methodology).

**Laboratory and other experiments.** Jarvenpaa [Jarvenpaa 1985] considers methodological issues in experimental IS research.

**Ethnographic approaches.** Harvey [Harvey 1995] introduces ethnographic research. Myers [Myers 1999] applies it to information systems.

As presented at the beginning of this section, we feel justified to establish that the current state of the art is largely practitioner-based use of more-or-less useful tools.

4. **KIMSPAG**

KIMSPAG (Knowledge and Information Management Support for People And Groups) is the authors’ initiative to address communication of data, information and knowledge in small groups that share a common language of discourse, with meanings which are agreed by the members of the group. The underlying assumption is that meaning and order will emerge as the members of the group build individual classifications and then cooperate to attempt to merge those classifications.

Methodologically it is possible to identify different approaches to the issues involved. One is to create a unifying “super-app”; one program which does everything, bundling the world into a super PIM/GIM. Two major research prototypes have emerged which take this approach and use semantic web techniques. They are MIT’s Haystack (http://simile.mit.edu/hayloft/index.html) and the Gnowsis project (http://www.gnowsis.org/). A more conventional “super PIM” approach is being taken by a Californian start-up company called NeoTech systems with their SQLNotes product [SQL 1986].

Another approach is to take a federating approach in which minimal assembly of emerging building blocks is undertaken [Sauermann 2005]. The work of two research centers is crucial in this context. One is the German Research Centre for Artificial Intelligence DFKI Gmbh based in Keiserslautern. The other is DERI notably at Galway in Ireland. DERI states that its mission is “to exploit semantics for people, organizations, and systems to collaborate and interoperate on a global scale” (http://www.deri.ie/).

In this context we do not feel that there may be much to be learned from surveying past users of unsuccessful PIMs, and identifying a sample would be difficult in the extreme. We currently favor a more qualitative, ethnographic approach, backed up by specific experiments and/or detailed ethnographic research, perhaps involving the time-honored tradition of “sitting with Nellie”. By extension, it is likely that action research will have relevance, it being difficult to conceive of an ethnographic approach to this problem domain which did not explicitly involve the observer or observers as agent. Summarizing, methodological approaches will be based on a combination of: experimentation, action research and ethnographic research.

Potential contexts of use are described by two typical scenarios, research and project management.
Research. Research is archetypal knowledge creation, and therefore management, carried out by individuals and small groups.

In this context, it should be noted the recent paper by Brown and Sice [Brown 2005], which restates the fundamentals of the systems approach. Some of their insights will form a leit-motif for our own subsequent work. In particular, we will recursively use the systems approach firstly to scope the research, then to arrive at a composite research methodology applicable in the scoped domain, and in the systematic evaluation of specific approaches and software.

Project Management. Projects are typically represented by the complex interaction of multiple activities and sub activities which are hierarchically related. Also, in projects, functional activities represented as vertical relations interact with project activities represented as horizontal relations in a complex resource sharing. Our project is a meta-project for itself that can be use in a recursive relationship with itself as a part of this research.

5. SUMMARY, CONCLUSIONS AND FUTURE RESEARCH

Perceived deficiencies in current data management paradigms and other paradoxes motivated the authors to look at and reflect on knowledge and information management tools with emphasis on small groups. An approach is proposed to be further developed in future research. The proposed research aims to be a grounded and relevant examination of why what should be widely-used techniques, are not. Some of the techniques that will be encountered and evaluated have been identified as likely to extend the scope of systems in the life of the individual, the group and even the organization [Evans 1997]. Our research will explore the possibility of achieving new levels of richness in the communication of knowledge and information without compromising attainable reach (see Fig. 3).

![Figure 3. The Blow-up of the Richness/Reach Trade-off [Evans 1997].](image)

6. REFERENCES


