Division of Research   December 1988
School of Business Administration

IT'S 10 A.M. DO YOU KNOW WHERE YOUR
DOCUMENTS ARE? EFFECTIVE MANAGEMENT
OF CORPORATE INFORMATION

Working Paper #567b

Michael D. Gordon
The University of Michigan

ABSTRACT

Much of the information important to organizations is "unstructured,"
taking the form of reports, memos, newspaper or journal articles, electronic
mail, a photograph, etc. rather than a concise, "structured," tabular form
amenable to database management. Yet management of such unstructured
information is rarely planned and sorely lacking.

This paper presents the results of interviews conducted with thirty nine
people from fourteen large firms with keen interest in information management.
From these interviews we see how failing to manage unstructured information
results in loss of intellectual product, duplication of effort, wasted time,
and wasted space. The broader consequences for business are impaired planning
and decision-making, reduced effectiveness during litigation, and an overall
decrease in competitiveness and profitability. We organize the interview
findings into four problem areas: searching and losing, sharing, information
overload, and excess storage. For each problem area, we suggest remedies
supported by the technical literature or undertaken by firms themselves.

It's 10 a.m. Do You Know Where Your Documents Are?

Effective Management of Corporate Information

1. Introduction

Could the accident at Three Mile Island have been avoided if the information anticipating it had been better managed? What is the cost to a firm when its senior level managers can't find reports and memos they require in their decision-making? Can an organization stay well informed without being swamped by information?

Much of the information important to organizations is "unstructured," taking the form of reports, memos, newspaper or journal articles, electronic mail, a photograph, notes on conversations with suppliers, etc. rather than a concise, "structured," tabular form amenable to database management. Yet management of such unstructured information is rarely planned and sorely lacking.

This paper presents the results of interviews conducted in organizations to examine unstructured information handling behavior and its impact on the firm. While much work has been directed at the design of information retrieval and office information systems, little attention has been paid to the actual practices, procedures, problems, and perceptions organizations have concerning management of unstructured corporate information. We document organizations' failings in dealing with unstructured information and present preventive and corrective measures to deal with them.

From these interviews, we learn that failing to manage unstructured information results in loss of intellectual product, duplication of effort, information overload, and excessive storage. The consequences for business are impaired planning and decision-making, reduced effectiveness during litigation, and an overall decrease in competitiveness and profitability. In addition, individuals take longer to accomplish less and may face stunted career growth.

Sections 5 and 6 of this paper describe, in interviewees' own words, the problems (and associated costs) with unstructured information that organizations face today. For each set of problems we identify, we suggest improvements involving both changes in procedure and behavior as well as deployment of computer technology. Sections 2 and 3 place these problems in context by emphasizing the importance of unstructured information to organizations. Section 4 describes the way in which the interviews were conducted.

2. Background: A trend toward information management

Modern business depends on information and supporting technology to function, let alone compete effectively. Larger organizations, more complicated organizational structures, as well as escalating environmental turbulence and hostility result in an increasing volume of information with which an organization must deal. Around the globe, information work comprises the largest single employment category in industrial societies, consuming, for example, two-thirds of labor expenditures (measured in either costs or hours) in the United States.[1]

Working with information is the central activity for many people in various organizations. Managers at all levels depend on information.[2] Consultants, lawyers, technologists, those in research and development or strategic planning, etc. comprise "knowledge centers" of specialists who have important roles in defining and solving problems.[3]

Over the last thirty years, organizations have changed their information management concerns and are beginning to focus on unstructured information.[4] Some three decades ago, concerns about physical controls over paperwork and about procedural efficiency developed into an urge to automate transaction processing and office systems. Deployment of such computing systems usually considered technological capabilities independently of and prior to assessment of information needs. As a result, the user became separated by technology from the information he needed in his job, often requiring an intermediary to gain access to information. Within the last decade, a confluence of technological and functional business concerns has begun to integrate communications, data processing, and office automation. The user is getting closer to the information he needs, strategic business planning is paying more attention to information technology, and treating information as a manageable resource is becoming a strategic objective of many organizations.

In the coming decade, management of unstructured information will come to the fore, by paying increased attention to techniques for selecting, analyzing, and synthesizing such information from disparate sources, and by better articulating users' information needs and concerns. Unstructured information management will better support decision-making, management, and operations by "focus[ing] on the content of information itself and how it is used and valued in the organizational setting[5]." The findings and recommendations in this paper are important steps in this direction.

3. Unstructured information in organizations

Database management systems are appropriate and powerful tools for storing, manipulating, and retrieving data, but much of the information in an organization is not data. Before we examine the interviews results (sections 5 and 6), it is important to understand how "data" (i.e., "structured" information) and "unstructured" information differ. In this way, we recognize that unstructured information is more prevalent in most organizations than is structured information and that it must be managed by different methods. Such recognition is prerequisite for proper information management.

Consider personnel records, manufacturing and inventory data, and accounting information. Both the high degree of structure of such information and the determinacy of its values make database methods appropriate for its management. The record structure of such information captures all its essential characteristics. By assigning one set of fields to personnel records (employee id number, employee name, department, date of hire, etc.) and a different set to inventory records (part number, quantity on hand, etc.), we can precisely frame requests for information. For instance, we can ask to list employees by date of hire, a request that makes no sense for inventory records. Additionally, the fact that data have determinate values provides us with "content addressable" access to information. For instance, we can use the "key" of an inventory record (part number=112233) to locate any or all data about that part. Similarly, we can find all part numbers with inventory below 100 units. Together, well structured information and determinate values provide us with "all-and-only" access: all the records we need and only those records.

But, most information is not so highly structured, nor does it have equally determinate values. A stored document (stored electronically or in paper files) is ordinarily represented by several fields. Some provide "content" information, usually in the form of keywords (such as, "Fourth generation language," or "microcomputers"). Others give determinate information about the document (such as document author, date of publication, etc.).

However, there are no methods for assigning content terms (keywords) to documents (and other forms of unstructured information) in a determinate way: Perceptions of content vary greatly among individuals, and language provides countless ways to express similar ideas. For example, someone looking for documents on "office automation" won't be furnished with documents described by the keyword "local area networks," even though such networks are integral to automated office work. Nor can the determinate information associated with documents solve the problem. Though such determinate information is useful for dividing a document database into disjoint subsets, it is not complete enough, by itself, for selecting particular documents. For instance, in an electronically managed document database, one can restrict retrieval to documents published between October and November 1987. Still, to actually retrieve relevant documents, the inquirer usually must try to identify documents' subject contents. Thus, the result of retrieving unstructured information is often far from the all-and-only retrieval one expects with highly structured records which employ determinate values.[6]

This means that database management models which logically address the retrieval of documents in all-and-only fashion are not appropriate. Instead different models of retrieval must be used to govern retrieval of documents.[7] The field of information retrieval is that branch of computer and information science addressing such issues.

5

Though, classically, the information retrieval problem involves finding references to relevant journal articles or books in a library, information retrieval is also a business problem. For its research in supporting product development, marketing, or economic forecasting, for example, an organization may need to locate relevant journal articles. Equally, any kind of report, memo, or meeting minutes is also potentially manageable by information retrieval methods. So, too, are filed exchanges of conversations, electronic mail messages, passages within computer conferences, photographs, graphic or image data, written procedures for accessing certain tasks, etc. Information retrieval techniques apply to strategic information as well as to more mundane information; internal information and external information; electronically-stored information and paper-based information; information that has been generated some time ago and information about which one wants to be notified when it is produced (or published). In fact, viewed this way, "unstructured" information occupies a much larger position in conducting business than does the "structured," tabular data manageable by database management methods. Whether or not firms have explicit plans, procedures, and software support, they are constantly dealing with unstructured information to keep their businesses going. As Keen said addressing the International Conference on Information Systems, "Most of our data is (sic) documents."[8]

Rather than addressing the technical aspects of information retrieval,[9] the remainder of this paper focuses on unstructured information retrieval as actually practiced in large corporations. By observing these practices, we see the penalties organizations pay for failing to manage better their unstructured information and can recommend solutions to this problem.

4. Method

Fourteen firms participated in this study. Each was contacted because of its commitment to better understanding information systems (as evidenced by its membership and active participation in the Information Systems Executive Forum at the Graduate School of Business at The University of Michigan). Collectively, these firms represent both manufacturing and service industries, and publicly and privately held companies. Participating firms were major organizations, with revenues often exceeding a billion dollars annually (see Table 1).

Altogether, thirty-nine people were interviewed, and all interviews were (audio) tape recorded. Participants held jobs from the executive level down to the clerical level, though most interviewees were upper level managers (see Table 2). Various functional areas were represented, including legal, purchasing, planning, and information systems.

Although an attempt was made to interview a cross section of people at a variety of firms, participants interviewed constituted neither a random nor stratified sample from the companies or industries represented. These interviews were conducted in an effort to elicit detailed reports which, by themselves or in combination, reveal the current state of practice of information management, especially problems in need of more attention. As is appropriate in exploratory research, this methodology permitted the in-depth exploration of important issues that could not be fully anticipated. Interviews ran from forty-five minutes to several hours, and each interview was completed in a single session. Generally, we follow the suggestion[10] that research on information systems must be willing to embrace newer, non-quantitative methods, including those requiring description, interpretation, and argumentation.

5. Interview results

The results of the interviews are now presented. Findings are grouped into four partially overlapping categories: problems with searching and losing; problems with sharing; information overload; and problems related to physical volume. For each of these concerns, we first examine the problem and then suggest remedies. The remedies we consider are directed at unstructured information management facing individuals as well as departments and corporations, and we consider both procedural and behavioral solutions as well as those which are technology based.

5.1 Searching and losing

People (and organizations) store documents so they can later make use of them.[11] But searching for missing information takes time, and losing information wastes ideas, evidence, and know-how. Both can severely affect a firm's performance and even threaten its survival. Unfortunately, these costs are too common an occurrence.

5.1.1 Losing time

Professional workers are estimated to spend 25% of their time distributing, filing, and retrieving documents[12], and a difficulty in locating information seems to affect all functions and all strata. A systems analyst conceded that "in a busy week, I could spend four hours looking through files." A manager of research and development described spending an entire day looking for a strategic planning document he had completed about six months earlier. An attorney summed up the situation in his firm concisely: "We have a significant problem locating documents. We waste a lot of time."

Even the executive level is affected. One vice president explained that, in his work, certain searches were simply too complicated to be delegated. As he explained: "Today I looked with my secretary for all documents [on a particular topic] for over an hour. We wanted more [text, drawings, and diagrams] but just didn't know where to look. [Such occurrences] arise several times per week."

The reason for all this effort is that certain information is mandatory for business to be conducted effectively. Consequently, a manager of end user computing described conducting the following search with another high level manager: "The two of us looked for a letter from my boss to the manager of operations. We each looked for one [entire] week. ... And his secretary also helped for half a day."

5.1.2 Loss of proof, fact, or experience

Observed a purchasing manager: "There have been many instances where you've got to prove it by pulling hard copy. The problem begins when it's the [three year old] paperwork you're looking for, not the [current paperwork]." One of the failed corporate-wide paper hunts he mentioned involved a $100 million anti-trust finding against an industry from which his company bought raw goods. All U.S. customers of that industry had to prove how much they had purchased during the seven years covered by the finding. "We had to send in reams of documents. Without the [purchase] orders, you're lost. And [we couldn't] find everything [we needed] ... from [our] archives."

The biggest costs to organizations of unlocatable documents involve short-circuiting or delaying the ability to refer to needed information. A document contains dates and facts which must be known for proof or verification. A highly placed director of operations explained the importance to him of having his files available. "I [am] very interested in files ...

9

not only [for] my correspondence, projects, etc., but [for monitoring] technology... I [am] convinced that after a long enough time, having a good enough set of files [is very important]."

The reasoning behind some action, a carefully reasoned policy statement, or an explanation of technical operation may be found in passages of certain documents. The director of operations added that "Some of that stuff [that I've filed] I will get called on a couple years later and [asked] 'Why did we do that?' or 'What was the approval process?' And I need [to know]...I would be very sad if these [files] went away because I periodically have to call on them [to address building problems, agreements, systems], and it helps keep things from being innuendo and keeps things down to the fact." At the CIA, where great volumes of unstructured intelligence information is processed, intelligence analysts consider their personal files their most important source of intelligence information.[13]

A document can encode intense, sustained intellectual activity for which individuals are highly trained and well paid. Such knowledge is part of the information backbone of an oganization. An analyst described "dissecting a computer program for [many] hours to solve some really critical problems" because the documentation was lost or thrown out. With the documentation, the problem could have been solved much more quickly.

Information also serves an important reminding function. A director of Information Systems observed that many documents he found very important to his work he had actually forgotten he had. He only recognized their relevance to a job at hand by browsing through his files, and he could make such assessments quite easily, even though he was not looking for these documents. Without proper attention, such information is effectively lost.

5.1.3 Contorted work

Patterns of work develop around the knowledge that finding documents is difficult. A high level manager was discussing trying to assemble all the relevant information from others in his firm in trying to decide an issue: "I don't do that too much. A lot of that may be that we know it would be almost impossible to do." Thus, the technical decisions he makes for the company don't benefit from pooled, collective experience and wisdom. More subtly, a search may be concluded on the mistaken assumption that all relevant materials have been located. As an attorney said: "[The documents we need to support our work] generally turn up where they should. But, then, we only look there."[14]

A manager of strategic planning described his difficulty in preparing plans, since 25% of the information he needed came from past history he had to dig out of other people's drawers and files and 50% more from interviewing people. He described his uneasiness about not knowing if his plans reflected his organization's true concerns since he was never sure if he had all the information he needed.

Failing to locate needed information can cause after-the-fact recollection from memory to replace more accurate documentation of experience. One individual explained his need to redo from memory a missing, six month old project plan needed by a corporate Vice President. After several days of attempting to reconstruct missing facts and figures the document was reproduced: "Let's just say I got away with it," although the re-creation was nowhere as complete or accurate as the original. "That does happen with regularity, [though] often times when that happens you can find another copy of it... though you end up with five or six people searching through their files, too." Another manager concurred, explaining that he usually had to re-

think and rewrite a lost document every few months. To anyone who has ever re-read something he has written, the point is clear: There is an immediacy to thought that makes even your own thinking seem foreign a few months later. To recreate lost or missing documents by relying on memory, often without notes, is risky business.

5.1.4 Remedies

Procedures can be instituted and software developed to make searching for unstructured information take less time and finding unstructured information more likely. These remedies can address the needs of individuals, departments, or entire organizations (see Table 3).

At the heart of the problem of searching and losing is the inappropriate and/or inconsistent naming of documents (and other unstructured information). A secretary spoke of being exasperated trying to search for documents for her boss, explaining that, to her, his file system (four fully crammed file drawers) was incomprehensible: To file, he "picks a different name every time [for the same topic] or simply circles the name on a memo he receives from various people on some project ... [Retrieval becomes] really difficult because ... there may be ten files in there [on the same topic] under ten different file headings."

This situation can be made far less severe by drawing up a controlled, pre-defined list of subject descriptions from which all subject designation of unstructured information must be selected.[15] Controlled vocabularies can be developed for an individual, department, or corporation. Once in place, improved consistency will help make it easier to locate a more complete set of needed information with reduced search effort.

Secondly, a variety of contextual descriptors (date of authorship, recipients of document, etc.) can be attached to documents (or other

unstructured information), thereby making it possible to isolate more easily

needed information. ("I'm not sure exactly how I named it, but I know I wrote

it last February or March, just before the project deadline, and Sachnoff was

sent a copy.") Such "contextual" information is especially helpful when

searching for electronically stored files with retrieval software capable of

searching for strings of text. By combining such contextual information,

("must have a creation date between last February and last March, and must

have been copied to Sachnoff"), a searcher can, effectively, reduce a corpus

of thousands of documents (or more) to a much smaller number which meet these

search criteria.

Also, responsibility for storing and searching is too often delegated.

Acknowledging that his files are almost useless, a director of business

systems admits he doesn't want to be involved in filing and retrieval.

According to his secretary, he later "can't find a thing" that he has had her

file using the subject designations of her choosing. Clerical costs mount not

only by looking for unlocatable documents but unnecessarily re-typing needed

documents that can't be found. Greater expense comes from failing to find

information that is needed.

In another organization, the responsibility for naming documents was

expressly vested in the clerical staff. One disapproving manager explained

that when he needed to locate electronic documents maintained on his

secretaries' computers the task was hopeless. "Memos [that they] put on

[their machines] I can find five times as fast in my paper files, ... because

... if they don't have that [document] number on the bottom of the page

they're lost. The heading may be 'Subject' [using] the first ten letters and

that's it. Without those numbers they have to go through all the subjects and

try to find it. It takes them forever. It may take a whole hour. [So I use

them] only when all else fails, only if I've exhausted my [paper] files and can't find it."

It is not necessary in all cases for managers and executives to personally file their own documents. But, they should form better habits of communication: telling their secretaries or assistants the subject phrases most likely to be useful in the future in retrieving a document, assigning names to electronic documents and electronic directories, and indicating the contextual information to attach to a document (or other piece of unstructured information) to help locate it in the future.

Four other prescriptions can help individuals improve their control over their personal electronic documents: 1) Using hard disks rather than floppy disks for file storage so that all files are online and, thus, electronically searchable at the same time. This is especially useful with software tools that search a device for particular strings of text. 2) Using information retrieval software developed for personal computers to manage personal document collections. These tools are designed for small to medium scale searches and use both subject and contextual information for retrieval. 3) Using hierarchical computerized directories to induce a useful organization (for instance, by project, topic, type of document, or period of time) for locating documents. 4) When floppy disks are used instead of hard disks, devoting a given floppy disk to a given project, topic, etc. Such partitioning, by directories or floppies, effectively reduces the number of files one must examine to locate a needed file by allowing a searcher to rule out all files not meeting known search criteria. Application of these rules can reduce the twenty minutes, on average, some interviewees say are required every time they need to locate an electronic file. Too often, a sequential search of all their stored files is the only way they are able to locate needed electronic documents.

To remedy problems of searching and losing at the level of the department, division, or corporation, efforts to investigate larger scale information bases for unstructured information should be encouraged. Several of the organizations interviewed are developing or investigating such information systems. Applications range from an online corporate-wide information base providing shared access to a Controller's manual covering corporate financial procedures, to shared text bases providing departmental or corporate-wide access for managerial decision-making. Non-text projects under consideration include information bases for retrieving engineering drawings and a facility for locating and retrieving appropriate statistical databases (or subsets of them) to provide input for graphics-aided decision making.

Corporate and department-wide information bases differ from individual information bases in that they must accommodate a number of information searchers, each with different perspectives and information needs. The most effective retrieval will be provided when searchers can access unstructured information using a variety of searching methods.[16] Among these are: full-text searching (which allows retrieval based on actual words contained in a document's text), keyword searching, and contextual searching. An interesting alternative, used in one organization interviewed, is to specially compose synopses of legal cases using only terms selected from a controlled vocabulary. These documents are created for subsequent use in litigation in the hopes of providing highly effective retrieval with reduced storage volume.

In addition, large scale unstructured information bases should evolve and permit customization. In fact, adaptation should be considered a necessary ingredient in an effective retrieval system for unstructured information.[17] Feedback should be solicited from system users concerning the "best" subject designation or most appropriate contextual information for particular stored

items. Such feedback can be used to improve subsequent access. Individuals can also be allowed to electronically annotate shared electronic documents, with retrieval software to keep track of who annotated what.[18] In this way, people will more easily be able to locate needed information and make use of it when they find it ("I want the document on which I included an annotation saying 'apply this formula for the Apollo project'.").

When applicable, paper documents can be digitized or converted to electronic format via O.C.R. (optical character recognition, which puts text into machine readable form) to allow more convenient or more effective management. In addition to yielding space savings, all documents can be put online, thus eliminating questions of what medium was used for storage (paper, microform, or electronic) and allowing searching by use of a variety of strategies (keywords, date ranges, full-text, etc.).

As we have seen, individual and organizational productivity is reduced by spending time hunting for information. What is more, unlocatable information fails to do its job: substantiating fact, instructing, explaining reasoning or operations, and reminding. The impact of yesterday's documents on today's work is illustrated by the remarks of a senior level manager responsible for business tactics in several technical areas. When asked if there was ever a document he had looked for but been unable to find, he replied, "I never found it many times. Some would have made my life a lot easier."

5.2 Sharing

Falsely, some people feel they can have under their control all the information they need without sharing. For example, as Lancaster described:

CIA analysts charged with producing "finished" intelligence require
comprehensive information about some topic in which they specialize: some
aspect of science and technology, medicine, economics, etc. Despite their
need for comprehensive information, these analysts felt they could personally
assemble, file, and locate all the information they would need. Studies
showed they were wrong: analysts' personal files were not complete, and
"major value" items were actually stored elsewhere in the organization.[19]

An organization can do what no individual possibly could, and sharing of
information is vital to coordinate and execute its various activities, from
operations through planning. However, as a result of size, specialization,
and diversification, the individual is separated from the information   needed
in his or her job. Large organizations make it difficult or impossible to
know who has produced or has access to certain information. Specialization
fragments that part of the world with which one has contact, and
diversification can mean that various sources may unknowingly require
overlapping information (such as information about shared markets for two
different product lines within one business). As a result, information
sharing is very difficult. An information system that can support
departmental or corporate storage and searching for information becomes a tool
for information sharing. We focus in this section on problems surrounding
sharing information and recommend remedies for them.

5.2.1 Out there ... but where?

Few organizations maintain active, departmental or corporate-wide files
to support workers' overlapping information needs. An attorney said: "In the
legal department here, everybody writes documents of some type that can be
used again. A sales agreement, a distribution agreement, you name it. An
agreement for the purchase of sale of the business. But I'll be damned if I

know who wrote what, on what subject, or where they are. But I know it's out there. So I've got to hunt these guys down." A director of administration, whose job includes being a disseminator of information, gets information "catch as catch can." Different departments inform her differently, some with regular reports, some erratically, and some not at all. Several times per week she must spend over an hour tracking down information that others failed to provide her so she can answer someone's request.

In fact, sharing can be impossible even when one knows who has already done work that could have direct bearing on a current situation. A vice president interviewed said, "Today I needed some files maintained by [someone who was no longer with the organization]. I couldn't use his files. Yet, you could have immediately gotten what you wanted from them if he was there." Such problems arise through hirings, firings, job reassignments, and corporate reorganizations. To a person, interviewees who had attempted to retrieve information from others' files reported rarely being successful doing so — especially for paper files. Organization of information is largely a personal matter, and the way an associate files things may make no sense to you. Given the difficulty many interviewees had in gaining control over their personal store of information, sharing becomes more difficult still. A manager of 95 people remarked: "I have one of those [horror stories concerning organizing and locating information] every day — as you can see by my office."

5.2.2 Consequence: replicated resources

Some firms recognize that duplicating information wastes resources. This sentiment was expressed by an information systems executive in a manufacturing organization: "We have replicated information systems groups [within our organization], and even within one of our operating companies at different locations. It would be nice [for one group] to say 'Is there any other work

going on in this area?'." Actually, such duplication of effort may never even be suspected. In one large company, several separate groups independently formed committees to evaluate the acquisition of main frame-based text retrieval software, at considerable cost to each group. When such duplication was discovered after the fact, the possibility of sharing knowledge and, thus, reducing expense had passed. "We never even knew it," one committee member said. With information systems designed for broadcasting and archiving information, such costs are avoidable. Further, projects requiring too great research expense can be undertaken by exploiting the knowledge already existing within the organization.

5.2.4 Consequence: losing opportunity and courting disaster

Companies fail to share documented findings of studies (for instance, the conclusions concerning the evaluation of different text retrieval packages) and changes in policy or practice (engineering change notices are often communicated erratically). Information from the field (obtained by salespeople, customers, etc.) has a difficult time finding its way to those who can make use of it.[20] On occasion, lack of shared information can potentially precipitate a catastrophe. Senior engineers' formal memoranda to management about sticking cooling valves possibly leading to a meltdown at Three Mile Island pre-dated the actual accident by two and a half years.[21] Their memoranda, submitted on the wrong form, were misdirected and ignored, as were their further urgings to take action. So were notices that the poor human-factors design would be a severe handicap during a crisis — another prediction that was borne out during the accident.

5.2.5 Consequence: Lost work product

Organizations have their own way of doing things, procedures they follow, and resources they use to get their work done. Often, this knowledge is

tacit, a fact which can painfully come to light during times of change. An attorney, describing the situation in his firm, concluded, "We do not do a very good job of capturing the work product. It's hard to know how to do your job if you're new or someone leaves. It would be very helpful just to know where to look." An information system that lets you "know where to look" can alleviate a good amount of disruption during transition, even though it will never make people completely interchangeable in their jobs.

5.2.4 Remedies

Improved sharing must involve stronger efforts to spread fact, opinion, and expertise within an organization (see Table 4). In some cases (such as capturing work product), stronger effort must be made to document and update practices and procedures, instead of allowing this know-how to leave the job along with departing employees. Wider broadcasts of news concerning ongoing projects, requests for information, new undertakings, etc. can help prevent duplication of effort and wasted resources. Intra-corporate communication networks and distribution lists can help achieve these ends. Training can help disabuse susceptible individuals of thinking that the information that they possess is complete.

Sharing of information has been given a boost by electronic mail. It was noted in several firms that this is the primary means of communication, especially where the communication is frequent and across organizational boundaries. "[With] lots of brief ... daily [messages] I used it all the time. For maybe a six or seven month period we had a major ... project. You needed to communicate. [But] as the project ... got done ... communications went from ... four or five times a day to once a week. [At that point,] people quit sending and start calling." The need to communicate with many people about the same topic is well supported by electronic mail distribution

lists or computer conferences, which make it as easy to communicate with many people as with just one. Electronic communication closes the gap geographically and temporally, allowing easy communication among those who don't work in close proximity.

But, sharing by electronic mail is instantaneous. It serves to alert colleagues to upcoming events, to notify about certain facts or milestones, and to organize (such as by helping to quickly arrange a meeting date). These communications, as helpful as they are, serve to disseminate information but fail to make it available for (convenient) future use. Even when long documents are routed electronically (rather than through the U.S. or intra-office mail), provision is rarely made for storing and indexing them -- which would regard them as a resource with future value. Rather, most companies interviewed demand that electronic information (mail) be purged from the system within several months of its distribution or receipt.

Ultimately, successful sharing implies an ability to thoughtfully, deliberately, thoroughly, but easily look through an accumulation of information. We have argued (section 5.1) that information bases for unstructured information should be used for departmental or corporate-wide storage and retrieval. To promote sharing, online access should be provided to widely used routine information. Additionally, electronic browsing should also be provided to support less anticipated information needs.

Recurring information needs can be supported by keyword "user interest profiles" used together with "current awareness systems." Incoming information is matched against users' profiles, thus alerting selected employees to internally or externally generated information pertinent to their work. Organizations are beginning to offer in-house current awareness systems for managerial and executive level employees to help keep them apprised about

the economy, government regulations, and other issues important to their work. Similarly, one can establish a profile in order to automatically keep informed about certain companies or industries that are tracked by the Dow Jones News Retrieval.

As we have mentioned, some of the organizations interviewed are beginning to provide shared information bases which support their legal departments, technical divisions, comptroller, etc. These efforts at sharing costs firms money, but ultimately can save them much more. To move ahead quickly, attention must be paid to: information quality, information selection, and size and users of the shared information base.

Quality filtering of information may take either of two forms. Information to be shared can be examined for its factualness, comprehensiveness, clarity of presentation, timeliness, and other factors. Alternatively, quality filtering can be inferred from usage patterns: High quality documents tend to refer to other documents of high quality, as do reputable authorities.[22] Quality, measured in either way, may be monitored. Singly or together, these measurements provide guidance concerning what information to maintain for sharing and what to exclude.

In light of its quality standards, an organization must next select the materials to which it will provide shared access. It is not necessary to store high quality information that no one will want to use, and low quality information is not of value. An insight that can help steer selection is that information users are more strongly influenced by ease of access than by quality of information.[23] In other words, higher quality information is used over lower quality only when it is equally accessible. This means that an organization that solicits from its employees information to be shared in a departmental or corporate information base is susceptible to making available

exactly what its members use in their work: easy-to-access, rather than quality, information. Particularly when electronic storage and retrieval can make all information equally accessible, quality should be made a stronger consideration. Organizations should ask managers, executives, and other information workers what information they need for their work but cannot obtain easily. Providing such information should be strongly considered.

Accompanying considerations of quality and selection are those of size of a shared information base. By virtue of their size, larger, more comprehensive information bases increase the likelihood of a searcher retrieving non-relevant information.[24] On the other hand, smaller information bases, used to satisfy local information needs of fewer people, may fail to reach those who can best make use of them and fail to contain certain relevant information. A careful assessment must be made to develop a system of the appropriate scale. Equally important, first efforts at sharing unstructured information must be appealing to the people who will use them, be relatively easy to prototype, develop and adapt, and have management's full-hearted support, lest they die a death from lack of use, inflexibility, or choked off funding.

In summary, facts, clarifications, procedures, recommendations, and warnings are available in the form of reports, memos, position papers, sketches, photographs, or graphs. Yet, this information is not easily shared. As a result, organizations incur a cost: duplication of effort from re-capturing and/or re-analyzing raw data; or ignorance on the part of its members with needs to know.

5.3 Overload and filtering

Our age of information has spawned an overload of information -- both paper and electronic. American business deals with 400 billion paper documents, a number which is growing by 70 billion a year.[25]

5.3.1 Toll

The purchasing manager who "start[s] every day with a six to eight inch lump of paper...You just don't know where to start," and the information planning executive who says "It takes me at least an hour a day just to sort my paper" experience the same problem: it takes time and effort just to filter irrelevant incoming information so that work can begin. This filtering can exact its toll. A business planning executive said, "I'm buried by paper ... [I've got] too much reading." His comments reflect the feelings of a telecommunications manager: "[I experience] definite paper and information [overload]. There are a couple hours a day reading time. It would be nice to 'net through' that." The problem is particularly difficult, he continued, since "It's tough to tell in advance [what information is important]. Something that's not hot today is hot tomorrow."

5.3.2 A worsening problem

Overload by paper is being accelerated by the ease with which one can prepare, revise, and print manuscripts. The advent and widespread use of technologies such as word processing software, computer photo typesetting, and laser printers on local area networks mean that a greater number of easy-to-produce, nice-to-look-at documents will be distributed internally in the form of reports, memos, etc. In addition, professional quality newsletters, solicitations, product announcements, etc. will flow increasingly from the outside to further overwhelm business with paper.

Similarly, electronic transmission of documents is an accepted method of communication in many businesses today. Electronic mail, especially in conjunction with distribution lists which automatically send a single message to a pre-selected group of people, has spawned a new kind of overload: electronic junk mail.[26] Posting of messages to computer bulletin boards and in computer conferences create new sources of information to contend with.

5.3.3 Remedies

Ackoff has discussed the "misinformation" that comes from too much information.[27] Filters can help sort information by subject or context and are under development for electronic communications.[28] These tools follow the general design of information systems for finding or sharing unstructured information by matching users' stated information needs against descriptions of incoming information. In principle, such filters apply to any information repository through which one wants to browse.

Operating effectively, information systems can locate additional relevant information while helping insulate a user from an overwhelming clutter. When not, the relevant becomes indiscernible from the non-relevant, and one can spend all day assembling the right information so that work can finally begin.

5.4 Storage Volume and Cost

How much unstructured information do people and organizations store? And what are the costs of their doing so? As we will see, excessive storage costs add to problems we have already discussed surrounding losing, failing to share, searching for, and being overloaded by information.

5.4.1 Guilty parties and costs

At all levels of an organization, people store documents. Secretaries maintain documents for those for whom they work in addition to memos, schedules, procedural manuals, etc. they require in the conduct of their personal work. One vice president interviewed maintains 80 file drawers of documents, each about three-fourths full. Having examined document storage in his own organization, he concluded that there were "more square feet of [active] file space than most people would believe." Another organization that measured what it was maintaining discovered five years worth of paper records totaling eleven linear miles.

This is a business age in which it costs four cents to file every document, $30 to process a purchase order, and $68 dollars for a single misfile.[29] As we have seen, American business deals with 400 billion paper documents, a number which is growing by 70 billion documents per year. This proliferation of paper is accompanied by the increasing ease with which one can electronically create and distribute multiple copies of documents using word processors and electronic mail networks.

5.4.2 Information hygiene

Training in the area of storage and retrieval is almost nonexistent. In response, some people tend to over-retain. One planning manager reported being "afraid to throw away a couple of old filing cabinets [worth of information]. And a lot is duplicates of others' files." Estimates are that at least fifty percent of a company's records are duplicates, twenty five percent of remaining records are worthless, and 85% of all filed documents are never referred to again.[30] Such behavior comes from hoping that personal retention and organization of large amounts of information will make needed information accessible, despite the penalties: increased storage and, what is worse, more difficult retrieval due to increased volume.

5.4.3 Remedies

Storage volume can give one pause, especially since companies interviewed have discovered that electronic storage is no longer more expensive than paper storage. Careful records management at a major bank have been accompanied by significant cost savings: over half a million dollars from consolidating inactive records; nearly a quarter million from reducing categories; over seven million by replacing paper with microfilmed records; and almost $600,000 by shortening forms.[31]

Emerging optical technologies must be given very serious consideration as a storage medium for replacing both paper and magnetic memory.[32] Digital scanning and OCR capabilities increase the attractiveness of optical storage.

Individuals must make efforts to weed out-of-date and little used materials from both their paper and electronic files. The argument that storage is cheap and weeding takes time is strongly rejected by the searching problems people have from looking through too much. A needle in a shoebox full of hay can be found: The odds of finding information go up as storage volume goes down.

Departments that maintain shared information bases must follow suit. Statistics should be maintained concerning the frequency and recency of use of centrally stored information. These can be used as a basis for reducing information volume. (See Table 5).

Certainly, savings from reduced storage are admirable. But we must not measure the effectiveness of information management solely in terms of cubic feet of storage and costs of physically filing documents. We must always consider the uses to which documents (and other unstructured information) are put and the ways in which inattention to their management can disrupt people's work.

6. Discussion

By design or default, companies manage their unstructured information. The impetus for improving information management in most organizations is the realization that neglect costs money. Remedies can take the form of training, procedures, and better computer-based tools for management of both text and other forms of unstructured information such as image and graphic data. Attention to unstructured information can be directed at the level of the individual worker up to the organization as a whole.

We have pointed out the efforts of some firms to begin making it easier to store, locate, disseminate, and share information. These activities parallel efforts abroad at providing retrieval for unstructured information. The Japanese are replacing the custom design and development of subcomponents of new products by searching for existing subcomponents catalogued along any of several retrieval dimensions, including: part geometry, the process used to make the part, the functionality of the part, etc.[33] Additionally, MITI has sponsored a major national project, Sigma, to create a national software library where Japanese companies may obtain free subroutines.[34]

But, effective retrieval of unstructured information remains a fantasy in most U.S. organizations. Companies typically employ computer systems for unstructured retrieval only after they dramatically demonstrate their benefits or because equally dramatic consequences could have been avoided by their use. A consultant to a manufacturing firm described "an unbelievably successful lobbying effort [using text retrieval software] ... which saved $700 million dollars ...[and] changed the way [they] did business. They scanned on O.C.R. ten years worth of documents and took them out of the file cabinets." Another

large company involved in a suit brought against the entire chemical industry had a surprisingly small settlement against it compared to its smaller co-defendants. Head counsel immediately attributed this savings to his company's ability to locate textual evidence necessary to support its legal claims. A less happy outcome surrounded the "uninformed lower level manager [who] incinerated some of the archives. [The company] found during litigation that they couldn't defend. They changed their business, [making] wholesale use of electronic information now."

The lines are drawn in many corporations over "old ways" vs. "new ways." An attorney in one corporation described "the enormous amount of textual information [we] manipulate," adding "It's absolutely essential to automate." Head counsel at another large firm stated he was still unconvinced that computers had any place in law. His feelings echoed those related by a information specialist at a facilities management organization describing a firm resisting the computer: "We've been doing business the same way for seventy-five years. Why change now?"

Several implicit, but fallacious, economic arguments hamper the use of information technology to make management, and business in general, more effective and efficient. First, information technology costs money, but non-computerized information systems are free. A consultant describing a firm with massive amounts of stored paper explained that they were keeping "eleven linear miles at three and one half cents a page ... and they had no idea what it cost to store, let alone ... retrieve, let alone the value of [being able to effectively retrieve it to help them conduct their business]." Second, the costs of information technology are real, but not the benefits. The information systems staff at one firm interviewed was asked if text retrieval software that could contribute a ten percent advantage in winning a $700

million lawsuit could be easily sold to management. "[No, because] we look to hard savings. The DASD [direct access storage device] requirements [would] choke our hardware planners who [would say] 'Can you really justify the cost?'." Third, benefits must have a strong local effect to be truly meaningful. However, information storage, retrieval, sharing, and filtering affect many people individually, diffusing their benefits across the corporation. In fact, as several information systems managers suggested, information storage/retrieval/sharing systems can have their greatest effects across departmental boundaries within a given organization by allowing better exchange among people who would ordinarily be isolated from each other. As a result, it is hard to find a vigorous champion for such systems, let alone someone to incur the expense of deploying one.

Information technology can support newer, streamlined patterns of doing work or, wrongly used, can automate obsolete patterns.[35] Properly used, information retrieval systems can provide information workers with more relevant information than they have dreamed possible (or considered economically feasible), and information in the form best suited to their needs: text, graphic, image, etc. At the same time, new patterns of work can develop from information systems that better support information sharing and reduce the information overload with which we too often deal today. There is opportunity for management to study the attention their firms pay to storage and retrieval of unstructured information, to realize that old ways aren't necessarily right, and that improved information technology can more than pay for itself.

We may only speculate about how improved information management might have prevented potential catastrophes such as at Three Mile Island. What is clear immediately is that organizations are devoting space to house

information they don't use, time to locate information they can't find, and
brain power in needlessly revisiting already solved problems. Though we are
overwhelmed by information, we can't seem to access what we need. By
appreciating these problems, assessing their costs, and developing information
systems for their solution, business today will hasten true information
management of tomorrow.

Annual Revenues of Participating Firms

Revenues	Number of Firms
$900 Million - 1 Billion	1
$1 - 2 Billion	2
$2 - 3 Billion	1
$9 - 10 Billion	1
$11 - 12 Billion	1
$22 - 23 Billion	1
Data Not Available	7

Table 1. Annual revenues for seven of fourteen participating companies.

Job Classifications for Subjects

Category	Number
Administrative	3
Attorney	5
Clerical	5
Executive	4
Manager	14
Technical	8

Table 2. Job classifications for thirty-nine subjects. Technical jobs include hardware and software specialists, corporate librarians, research scientists, archivists, and records retention personnel.

INDIVIDUAL	DEPARTMENT OR CORPORATE-WIDE
Improve naming:	Improve naming:
Use Controlled vocabularies	Use Controlled vocabularies
Use Contextual descriptions	Use Contextual descriptions
Avoid rote delegation of filing and searching	Establish shared, electronic information bases for unstructured information
Use hard disks for storage	Provide full-text access for retrieval
Use bibliographic retrieval software	Provide keyword access for retrieval
Use hierarchical file directories	Solicit feedback concerning subject designations and and contextual descriptions
Devote floppies to a particular project, date range, etc.	Allow personal electronic annotation of materials in shared information base
	Digitally scan or use OCR to convert from paper to electronic storage

Table 3: Remedies to problems of searching and losing

INDIVIDUAL	DEPARTMENT OR CORPORATE-WIDE
Document and update work practices	Establish corporate networks and distribution lists
Broadcast news and information over corporate networks	Use electronic information bases for widely used routine unstructured information
	Allow electronic browsing
	Establish user interest profiles
	Develop current awareness systems
	Establish quality standards for shared information
	Monitor quality of shared information
	Select materials to share in light of quality standards
	Ask employees what needed information is not accessible
	Carefully consider the size of a shared information base
	Make first efforts appealing to users and management

Table 4: Remedies to problems of sharing

INDIVIDUAL	DEPARTMENT OR CORPORATE-WIDE
Weed out-of-date, unimportant material	Consolidate inactive records
	Investigate optical technologies for high volume storage
	Maintain statistics on recency of information use
	Maintain statistics of frequency of information use

Table 5: Remedies to excessive storage volumes

References

1. Strassmann, Paul A. <u>Information Payoff: The Transformation of Work in the Electronic Age</u>. The Free Press. New York, 1985. p. 3.

2. Mintzberg classified the activities of managers into three roles, each of which depends on information for its proper execution: interpersonal (which includes "exchange" relationships connecting the company or unit the manager oversees to its environment); informational (which includes monitoring formally and informally relevant sources of internal and external information and broadcasting information to subordinates, superiors, and people outside the company or unit managed); and decisional (which involves using information to bring about controlled change in the manager's organization, exploiting opportunities, handling disturbances, allocating resources, and negotiating). See: Mintzberg, Henry. <u>The Nature of Managerial Work</u>. Harper & Row. New York. 1973.

3. Taylor, Robert S. "Information Values in Decision Contexts." <u>Information Management Review</u>. Vol. 1, No. 1, 1985. pp. 47-55.

4. Marchand, Donald A. "Information Management: Strategies and Tools in Transition." <u>Information Management Review</u>. Vol. 1, No. 1, 1985. pp. 27-34.

5. Marchand (1985), p. 32.

6. Some systems allow for the creation of different document types. Lens, a prototype information retrieval system, is the strongest example of such a system. See Malone, Thomas W.; Grant, Kenneth R.; Turbak, Franklyn A.; Brobst, Stephen A.; and Cohen, Michael D. "Intelligent Information Sharing Systems." <u>Communications of the ACM</u>, Vol. 30, No. 5. May, 1987. pp. 390-402.

7. Blair, David C. "The Management of Information: Basic Distinctions." <u>Sloan Management Review</u>. Fall 1984. pp. 13-23.

8. Keen, P. G. W. In "Proceedings of the Fifth International Conference on Information Systems," edited by L. Maggi, J. L. King, and K. L. Kramer. Chicago: Society for Information Management, 1984. As cited by Holmes, Fenwicke W. "The Information Infrastructure and How to Win with it." <u>Information Management Review</u>, Vol. 1 No. 2, 1985. pp. 9-19.

9. Two general probabilistic strategies for coping with the difficulty in assigning subject terms to documents are 1) assigning descriptions to documents based on the terms inquirers actually use in searching for relevant documents and 2) devising queries based on how relevant and irrelevant documents are described by different sets of keywords. See Robertson, S. E., Maron, M. E., and Cooper, W. S. "Probability of Relevance: A Unification of Two Competing Models for Document Retrieval." <u>Information Technology: Research and Development</u>. Vol. 1, No. 1, 1982. pp. 1-21. For a fuller discussion of information retrieval problems and approaches see VanRijsbergen, C. J. <u>Information Retrieval</u>, Second Edition. Butterworths. London. 1979, and Salton, Gerard, and McGill, Michael J. <u>Introduction to Modern Information Retrieval</u>. McGraw-Hill Book Co. New York. 1983.

10. Galliers, Robert D., and Frand, F. Land. "Choosing Appropriate Information Systems Research Methodologies." Communications of ACM. Vol. 30, No. 11, November, 1987. pp 900-902.

11. In addition, of course, various regulations issued by the Interstate Commerce Commission and other agencies also require certain documents be retained. Recent record-keeping violations have been uncovered in large firms by the Occupational Safety and Health Administration and the Nuclear Regulatory Commission.

12. Yourdon, Edward. "Paper Chase: Keeping up with Office Productivity." Computer World. July 21, 1986. pp. 53-58.

13. See Lancaster, F. W. Toward Paperless Information Systems. Academic Press. New York. 1978.

14. Searchers for information often believe that electronic retrieval systems provide all the relevant information on a topic. A study which reveals such perceptions, and their inaccuracy, is Blair, David C., and Maron, M. E. "An Evaluation of Retrieval Effectiveness for a Full-Text Document Retrieval System." Commumications of the ACM, Vol. 28, No. 5, March 1985. pp. 289-299.

15. F. W. Lancaster treats this topic authoritatively in Vocabulary Control for Information retrieval, Second Edition, Information Resources Press, Arlington, VA, 1986.

16. Two recent studies have independently shown that searching by a variety of methods will locate non-overlapping sets of relevant documents. See: Katzer, J. McGill, M. J., Tessier, J. A., Frakes, W. and DasGupta, P. "A Study of the Overlap Among Document Representations," Information Technology: Research and Development, 1(4), 1982, 261-274; and Tenopir, C. "Full Text Database Retrieval Performance," Online Review, 9(2), 1985, 149-163.

17. See Gordon, Michael D. "The Necessity for Adaptation in Modified Boolean Document Retrieval Systems," Information Processing and Management, 24(3), 1988, 339-347.

18. Lancaster (1978) explains that the CIA instituted such a facility for the CIA analysts using shared electronic files.

19. As a result, the CIA maintains company-wide intelligence files for both archival and current awareness purposes. Lancaster, F. W. (1978)

20. Gilad, Tamar, and Gilad, Benjamin. "Business Intelligence—The Quiet Revolution." Sloan Management Review. Summer, 1986. pp. 53-61.

21. Burns, Christopher. "Three Miles Island: The Information Meltdown." Information Management Review, Vol. 1, No. 1, 1985. pp. 19-25.

22. Warren, Kenneth S. Selectivity in Information Systems. Praeger Publishers. New York. 1985.

23. O'Reilly, Charles A. "Variations in Decision Makers' Use of Information Sources: The Impact of Quality and Accessibility on Information." Academy of Management Journal, Vol. 25, No. 4, 1982. pp. 756-771.

24. Lancaster, F. W. Information Retrieval System -- Characteristics, Testing, and Evaluation. 2nd Edition. New York: Wiley, 1979.

25. Yourdon (1986).

26. Denning, Peter J. "Electronic Junk," Communications of the ACM, March, 1982.

27. Ackoff, R. L. "Management Misinformation Systems." Management Science. December, 1967, pp. B147-B156.

28. Malone et al (1987)

29. Swartz, Herb. "For the Record." Skylines. Feb. 1986. pp. 10-11.

30. Swartz (1986).

31. Swartz (1986).

32. CD ROM: The New Papyrus. Lambert, S. and Ropiequet, S., eds. Microsoft Press, Redmond WA. 1986.

33. Gunn, T. Computer Applications in Manufacturing. Industrial Press, Inc. 1981.

34. Haavind, R. "Tools for Compatibility." High Technology, August 1986. pp. 34-42.

35. Strassmann (1985)