

How You Store Information Affects How You Can Retrieve It: A Fundamental Principle For Business Students Studying Information Systems And Technology

Mark S. Silver, Fordham University, USA

ABSTRACT

During the current period of rapid technological change, business students need to emerge from their introductory course in Information Systems (IS) with a set of fundamental principles to help them “think about Information Technology (IT)” in future courses and the workplace. Given the digital revolution, they also need to appreciate the role of information in business as well as how to meet the challenges involved in managing information effectively. This paper addresses both those needs by presenting a fundamental principle concerning information management: How you store information affects how you can retrieve it. The paper commences by presenting the principle in a manner that can be used to introduce it to the class. It continues by providing numerous concepts and examples that draw on the principle and that students are likely to encounter in the core IS course, subsequent courses, and their real-world use of technology. The paper concludes by raising a set of issues suitable for class discussion or exam questions.

Keywords: Information Retrieval; Information Systems; Information Management; Information Technology

INTRODUCTION

Because businesses are so dependent on Information Technology (IT) in the digital era, business students at both the undergraduate and graduate level must emerge from their academic programs with a managerial understanding of Information Systems (IS) and Information Technology essentials. Business faculty often debate; however, the precise content of the core IS course (Topi, et al., 2010; Hershey, 2002; Ives et al., 2002; Salisbury, Huber, Piercy, & Elder, 2004; Silver, Markus, & Beath, 1995) and content varies greatly across educational institutions. While it is easy to agree that the course must cover “what every business student must know about IS and IT,” there is no consensus on what that is, especially as technology and its business applications change rapidly.

In the realms of IS and IT, what business students need to learn more than anything else in an age of rapid technological advances is how to “think about information technology.” Learning how to approach and contemplate information technology gives students something more long-lasting than just an understanding of the latest technologies, applications, or IT-related business trends. Knowing how to think about IT enables students to cope with a world in which specific technologies and their business consequences come and go. Providing students with a small set of fundamental principles that are widely applicable and likely to have ongoing significance can create an important foundation for such knowledge. These principles, ideally, should manifest themselves repeatedly in the core IS/IT course, more advanced courses, and students’ daily encounters with technology.

Information Management is one consensus topic with which business students must be familiar and there is a fundamental principle that underlies this subject. Because information is the lifeblood of the organization, managing it properly is essential to business people. Indeed, managing and using information effectively is critical just for getting the job done and getting it done well—that is, for operational efficiency and effectiveness. Information is also typically a key asset for businesses that use IT strategically. Similarly, “data-driven decision making,” which is the

business benefit of Business Analytics and Big Data, clearly depends on having the data. Moreover, almost everything business people do today is mediated by technology, so their personal productivity depends on managing vast volumes of email, documents, spreadsheets, and other forms of digital information.

Each of these business uses of information raises numerous practical concerns about managing information properly. Among the challenges are (1) how to secure information against theft or accidental loss, (2) how to leverage the sharing of information while protecting confidentiality and personal privacy, (3) how to cope with information overload, and (4) where to locate information for both accessibility and cost-effectiveness. The key to addressing all these concerns is found in a single fundamental principle:

How You Store Information Affects How You Can Retrieve It.

Put differently, this principle states that because businesses care a great deal about *retrieving* information, they must pay attention to how they *store* and manage it. Inattentiveness to the various elements of storing information is likely to preclude or impede retrieving it when, where, and in the form they need it.

Surprisingly, this principle is generally not found in the textbooks and is therefore missing from the typical business student's education. Yet, understanding this principle and its implications helps students appreciate why they are asked to learn many concepts and topics that may otherwise not seem essential to their business training. Framing course material in terms of this principle helps make topics that may otherwise seem overly technical relevant to business and management. Further, establishing this principle as a conceptual foundation may enable students to understand more clearly and appreciate more fully the material related to information and database management they encounter in their coursework. Additionally, internalizing this principle is likely to make them more effective at managing information in their personal and professional lives.

The purpose of this paper is to introduce this fundamental principle of information management as a resource for teaching business students about Information Systems and Information Technology. The paper begins by presenting the concept in a way that can be used to introduce the idea to students. The paper continues by providing numerous examples of the principle in action that students are likely to encounter in their business studies or in their personal use of technology. The paper concludes with a number of issues the principle raises that are suitable for class discussion, assignments, or final exam questions.

INTRODUCING THE PRINCIPLE TO THE STUDENTS

Before presenting the principle, one can prepare the students for its message by helping them recognize that (1) information is critical to business and (2) managing information is difficult. While some may see these two points as so-called "no brainers," there is nonetheless significant value in getting explicit student buy-in to these two important observations. If students have already studied (to a greater or lesser extent) the role of IT in business, or if they come to the course with substantial business experience, simply asking the students to provide examples of the role of information in business should enable the class to conclude relatively quickly that information matters a great deal in business. In environments where students do not have such experience, waiting to introduce the principle until at least some of the business benefits of IT have been covered in the course may be preferable.

Students tend to offer relatively specific examples of the role of information in business, so it may be useful to generalize from their offerings, thus identifying a few broad types of information use and its related benefits. For instance, students may mention how information can help with managing inventory or making better pricing decisions. These can be framed as instances of "data-driven decision making." Similarly, if students point to specific examples of employing information to promote more efficient operations, this can be generalized to "reducing costs." Examples of analyzing social media postings may illustrate "improved product design" or "enhanced customer service." Similarly, Customer Relationship Management (CRM) systems are one way of using information to retain existing customers and attract new ones. Because students often focus on the firm level, they often omit the role of information in improving personal productivity in the workplace. Should they do so, it is useful for the instructor to call this benefit to their attention.

The second motivating observation, that information is difficult to manage, can be illustrated by presenting a few key statistics such as this one from IBM (n.d.; Redman, 2016):

The US economy loses an estimated \$3.1 Trillion annually due to poor data quality.

Indeed, this statistic reinforces the significance of information in business while highlighting the challenge of managing information well. Rather than rattling off statistics, posing questions (“What is the estimated annual loss to the US economy due to poor data?”) and allowing students to guess the answers is a good way to get the students’ attention and make a big impression. Students typically understate the true statistics by many orders of magnitude.

From these two initial observations, one concludes that because information is both (1) essential to business and (2) difficult to manage, business people must understand how to manage information well. And that is why this topic needs to be studied in business school.

Now that the students are engaged in the topic of information management, one can point out that the goal of information management is to ensure that people and companies can retrieve the information they need when they need it, where they need, and in the form they need it. Here are a few informational wants and needs—stated in the second-person (“you”) to engage the students more personally—that can bring the point home:

- Your customer service representatives must be able to pull up a client’s data within seconds of answering the call, because this is what customers now expect.
- You want to be able to provide the information your boss demands when he or she demands it—which is more or less immediately.
- You need the requisite information to seize an opportunity before your competitors beat you to it.

Next comes the pivot: While the focal business need is for *retrieving* information, this need places a demand on *storing* information well. Why? Because of the following fundamental principle:

How you store information affects how you can retrieve it.

This fundamental principle of information management applies to both personal and professional information management, ranging from filing one’s daily email to managing a corporate database. The more careful you are, the more sophisticated you are, the more effort you expend in storing information, the more effectively you will be able to retrieve it when you need it.

Storing clean laundry makes an apt analogy to which students can relate well. Storing clean laundry is not a problem: You can open a drawer or closet and just toss it in. But what happens when it is 6 AM, you are half-asleep, and you are trying to find a pair of socks that match? At this point you wish you had expended more effort sorting the laundry and had stored it in a more appropriate format. Had you paired the socks when you stored them, you could easily retrieve a pair now when it is needed.

This fundamental principle is not limited to information; it is a special case of a principle for life:

How you store “stuff” affects if, and how easily, you can find that “stuff” when you need it.

Consider, for example, car keys. Many of us are always searching for our car keys because our carelessness—or, at least, our casualness—in storing those keys hinders our ability to find them later. But those people who place their keys on a hook near the garage, rather than just tossing them anywhere, are able to locate them more easily when needed for at least three reasons:

- By hanging them up, they will be more visible. They will be less likely to disappear between the couch cushions or be covered by something else that was just tossed anywhere.
- By always storing them in the same place, recalling where the keys were placed will be easier.
- By locating them near the garage, they are close to where they will be needed.

The clean laundry and car keys examples illustrate the principle in action. At its most basic level, the principle reminds us that storage and retrieval are connected. The way things are stored does bear on the way they are retrieved. This may seem obvious, but too often businesses neglect the connection when storing information. Many businesses expend significant resources storing data, but if they do not anticipate how that data will need to be retrieved they may find it difficult or impossible to retrieve the data as needed.

Put differently, one's ability to retrieve information may depend on how that information was stored. Indeed, how information is stored may affect (1) how easily it can be retrieved, (2) if it can be retrieved at all, and (3) the specific means by which it must be retrieved. In particular, ***greater care and greater effort in storing information generally yields significant rewards when retrieving that information.***

- **Greater Care:** The better the choices one makes when storing something, the easier it will be to retrieve. We saw this with the car keys. Less effective storage hampers or limits one's ability to retrieve information, whereas superior storage facilitates greater ease and speed in retrieval. Choosing where to locate information physically and logically are two key storage decisions that highlight the need for taking such care—and the perils of not doing so.

While the concepts of physical and logical data views are more complex and subtle in the context of database management, when introducing the fundamental principle and the need for greater care, they can be presented in the more limited case of storing and retrieving "files," such as word processing documents, electronic spreadsheets, and so forth. A brief mention of these choices should resonate with students who may recognize their own lack of care in storing information. It should also suffice to convince them that greater care when storing leads to greater benefit when retrieving.

- The **Physical Storage Location** can be described as where files are actually stored. Today, the basic choice is between local storage and cloud storage. But local storage affords many possibilities including computers (desktops vs. laptops), mobile devices (smartphones vs. tablets), and removable media (USB memory sticks, CDs/DVDs, magnetic disks, magnetic tapes). And both local and cloud storage pose a variety of considerations that include cost, availability, and security.

Students can be asked for examples of where insufficient thought about these storage options caused them problems. Alternatively, the instructor can poll the students about specific difficulties they may have experienced. No doubt many students have carelessly "left" a homework file on a laboratory computer only to discover that either they no longer have access to it or that someone else stole their assignment.

- The **Logical Storage Location** refers to how people and businesses identify files that they wish to retrieve among the hundreds or thousands that are stored in a given physical location. Typically, this logical location includes two components: the file or resource name (an identifier for a given file, such as "mortgagedata.xls") and a directory that contains the file (sometimes called a folder). Given the vast number of files that are stored, too little care in naming files or assigning files to directories can make those files very difficult to find later.

Students can surely relate to this phenomenon, as well. How many have panicked when they completed a research paper a day early but couldn't locate it when they went to print it an hour before it was due?

- **Greater Effort:** The more effort one expends in storing something the less effort generally will be required to retrieve it because that added effort often facilitates the subsequent retrieval. For example, constructing indexes when storing data enables direct access when retrieving that data. Given the trade-off between expending effort at storage time or at retrieval time, several factors favor expending the effort up front.
 - First, often a small amount of added effort when storing may yield dramatic savings when retrieving. For this reason we often make the effort to create and maintain secondary indexes in large corporate

databases, allowing us to search the databases rapidly on a variety of attributes. And in our personal information management, the small amount of added effort involved in assigning files meaningful names, paying attention to the folder in which we place the files, and noticing where that folder is located can save us much time in trying to locate that file at a later date.

- Second, unlike car keys that are both stored and retrieved repeatedly, information is likely to be stored once and retrieved many times over the subsequent months and years. So spending the added effort the one time we store the information is clearly better than needing to spend added effort each time it is retrieved.
- Third, we often have the luxury of expending extra time when storing information but time tends to be more scarce when the information is needed. For example, there may be no urgency when information is stored but delays in data retrieval may lead to lost opportunities or disgruntled customers.
- Fourth, in many organizations different people perform the storage than perform the retrieval. Order clerks, customer service representatives, and sometimes even the customers themselves do primary data entry, whereas retrieval is often performed by managers and executives for decision-making purposes. The time spent by those doing the storage costs the firm much less than the time of those who do the retrieval.
- And finally, if we do not take sufficient care in storing the information, in many cases it will be impossible — or nearly so — to retrieve the information at all. Consider the added effort involved in backing up stored data. If we fail to back up, no amount of time and effort may replace the lost information should we suffer a cataclysmic failure.

So, taken together, added care and added effort in storing information enables better retrieval. This implies that how you store information is critically important because it affects how you can retrieve it.

At this juncture, immediately after introducing the principle, it may be propitious to put on the table several of the chief considerations that arise when choosing how to store information or, put differently, some of the factors that will matter when trying to retrieve it. These include the following:

- Cost — the financial cost of storing, maintaining, and subsequently retrieving the information.
- Availability — how easily, if at all, the information can be retrieved when, where, by whom, and in the form it is needed.
- Compatibility — whether the information is stored in a way that is compatible with the hardware and software systems that will be used to access it.
- Veracity — whether or not the information is correct.
- Security — how vulnerable the information is to theft or loss.

These considerations, which drive storage choices, may often conflict one with another. For instance, storage choices that promote availability, compatibility, and security may increase the financial costs involved. Storage choices that make information more readily available—for example, for sharing within the organization—may make that information more vulnerable to security breaches. Encouraging students to keep these considerations in mind as they contemplate information management may be a good way to conclude the principle's introduction.

Equipped with a basic understanding of the fundamental principle as well as the considerations that often come into play when applying it, the students are now prepared to recognize the principle when they see it in action and to apply it when analyzing information management issues.

THE PRINCIPLE IN ACTION: COURSE CONCEPTS AND EXAMPLES

The core course in IS/IT offers many concepts that can be presented as instances of the fundamental principle in action. And there is a duality at work when course concepts are presented in light of the principle: Appreciating the

fundamental principle enriches understanding of the course concepts, and studying the concepts reinforces the principle.

For those core IS courses that include a database management module, many of the examples of the principle are found in that module, but other examples can be found throughout the term. A few of the more sophisticated examples may apply more to advanced information systems courses, where they can reinforce the principle and demonstrate its ongoing value. Here is a set of topics (some narrow, some broad) where the principle is in evidence (and sometimes a driving force). They each illustrate how some combination of greater care and greater effort when storing information affects retrieval of that information. The topics are grouped into four major themes, but the order in which the individual topics and examples are presented would likely vary based on the design of a given course.

Concepts and Examples from Students' Daily Lives

Organizing Class Notes

If students are instructed to reserve some pages at the front of their notebooks for the fundamental principles of the course, these pages themselves serve as an example of the principle. By storing their notes about the fundamental principles in a special, easily accessible portion of their notebooks, students can retrieve this material easily when reviewing. Of course, increasingly, students are taking notes electronically. But the principle still holds. For easy access, notes relating to the fundamental principles can be placed in their own document (or highlighted in some manner within a single comprehensive notes document). In fact, since these ideas appear repeatedly throughout the term, students could go so far as to tag subsequent references to these ideas or to hyperlink those future references back to the original discussion of the principles. While most students may be unlikely to go the extra mile, the instructor can nonetheless suggest they do so, pointing out that such tags and hyperlinks are additional instances of greater effort in storage yielding added benefits when retrieving. And this application of the principle need not be limited to the core Information Systems class. As many of the better students have likely already learned for themselves, greater investment in organizing one's notes for any course typically yields dividends when retrieving those notes for study and review.

Personal File Organization

Another activity students can relate to easily is managing their files and file folders on a computer, mobile device, removable storage medium, or the cloud. Creating a clear directory structure, taking care when placing documents in folders, and using consistent naming conventions enables finding those documents later with relative ease. In contrast, paying little attention to the directories (folders) in which documents are stored or what they are named makes finding them when they are needed far more difficult. What might "Termpaper.3.doc" contain? A history paper? English literature? Economics? Is it the latest version? Is it even from this term? And in which folder did the document end up? The last one that was open, so the economics term paper is in a folder called "English Lit?" Or maybe it is on the desktop (or in the dropbox root directory) with 200 other uncategorized files.

Personal information management also illustrates, in a way that resonates for many students, that time is often more pressing when retrieving information than storing it. As an earlier example illustrated, one can often spare the extra minute or less to store a file carefully after working on it (even if one feels too tired or lazy to do so), but when retrieving a course deliverable there is often no time to spare.

In addition to the logical storage locations — the organization of files and folders—physical storage locations matter, too. Even better than just having a consistent naming and directory structure on each device is having a coherent approach to what gets stored where physically — for instance, on a computer, mobile device, removable medium, or the cloud. Finding information is that much more difficult if one does not even know which device to search. Indeed, that observation leads to the next example.

The Cloud

As noted earlier, one of the critical storage decisions that demands great care is where to place information physically — that is, deciding where the bits will reside. One element of this decision is choosing between storing information

locally on a piece of hardware that the individual or the firm has in its direct physical possession and storing it remotely on “The Cloud” — that is, somewhere on the Internet. Today, the cloud is all the rage, and even the least technical of core IS courses needs to address this subject to help students cut through the hype, confusion, and complexity that surround cloud computing. A complete treatment of cloud computing includes both the software cloud — employing applications available from third-party providers in a Software as a Service (SaaS) model — and the hardware cloud.

A significant component of the hardware cloud is cloud storage. It can be argued that not only does this aspect of cloud computing illustrate the fundamental principle but that referencing the principle is essential for presenting cloud storage to business students properly. Despite all the hype surrounding the cloud, it has its pros and cons and the challenge is how to trade them off one against another in any given situation. For instance, the benefits of cloud storage include better information access and sharing, since data can be retrieved anywhere there is an Internet connection, as well as the various advantages of outsourcing, such as less investment in hardware, reduced obsolescence, potentially better service levels, scalability, and fewer IT management issues. While these are compelling reasons for going to the cloud, these benefits can be offset by a few powerful drawbacks, which include dependence on the third-party cloud provider, security concerns as sensitive information becomes vulnerable by traveling over the Internet and residing on someone else’s server, and lack of availability if there is no Internet connection. All of this material, which is focused on choosing an appropriate storage location, can be framed in the context of the fundamental principle, since the principle highlights that storage choices need to be made in the context of retrieval needs. In particular, notice how the liabilities of cloud storage align with two of the storage considerations identified earlier: availability and security.

The trade-offs between the cloud and local storage can be highlighted by the following anecdote, which describes what is likely a common occurrence in higher education:

One semester, I was teaching back-to-back classes at opposite ends of the campus so I needed to dash from one to the other. At that time, I stored the PowerPoint slides for my lectures on a USB memory stick. One day in my haste I left the memory stick in my first classroom and I was without the PowerPoints for my second class. The following class day I made the move to the cloud and stored the PowerPoints online. And, of course, the Internet was down in my second classroom that day and I was again without my PowerPoints. In each case, how I *stored* the PowerPoints limited how I could *retrieve* them. In fact, in each case, it ended up preventing me from retrieving them during class. Now I store my PowerPoints both on the cloud and locally for every class.

With this story as motivation, students are often able to suggest their own examples of the trade-offs. After such examples are presented, it is important to note that if these trade-offs have significant consequences for individuals (professors and students), one can only imagine how consequential they are for major firms storing large volumes of essential corporate data.

Finding Information on the World Wide Web

One of the most common tasks students perform online is searching for information on the web. What makes the web powerful and unique as an information repository are the hyperlinks that connect one online resource (for instance, a webpage) to another. Whether students follow the links themselves or rely on a search engine such as Google to do it for them, these hyperlinks are critical to information retrieval on the web. One factor that can make finding what they need difficult for info-seekers, and can make being found challenging for businesses on the web, is that while creators of new documents create links to existing documents, they have no control over which pages link to them. So, for the most part, new pages are stored without any paths to reach them. This limitation of how pages are stored constrains the ability to find them later. Web page creators and administrators do, however, employ various methods to increase the likelihood of their pages being found. This practice is known as Search Engine Optimization (SEO). In addition to creating alliances with other businesses and websites, most web page authors who are concerned that their pages be found provide their pages with meaningful URLs and carefully selected metadata (such as page titles and keywords). They also submit their pages to those search tools that accept registrations. So, if one thinks of these activities as part of storing a web page, then how the web pages are stored affects whether they are found.

Backups

Given the typical student's poor record of backing up files, one cannot mention backups too often in the introductory course — or any course, for that matter. And backups provide another example of greater care in storage yielding rewards down the road. Regular backups certainly place an added burden on the user and the computer, with no immediate benefit. But when bad things happen—a complete disk crash, the accidental deletion of a single file, a lost USB memory stick, and so forth—data retrieval is not just impeded without a backup, it is made impossible.

If time permits, it may be worth pointing out to the class that just backing up one's files regularly is not sufficient. There are correct and incorrect ways of performing backups. Here are two ways that a person or business can get in trouble even with regular backups:

- Replacing the previous backup with the new one is a common mistake. For instance, suppose a student spends several weeks working on a term paper and diligently backs it up each night in a file with the name "English.TermPaper.Spring17.backup.doc." While this might seem like a very descriptive filename—which is a good thing—each night's backup replaces the previous one. So if during the course of the day an important part of the term paper is accidentally deleted, and the student does not notice it, that night's backup will replace yesterday's good backup with a copy of the damaged file. It is important, therefore, to use a backup scheme that does not destroy previous backups.
- Storing the original data and the backup copy in the same physical location is also problematic. Suppose the student is working on the hard drive of his or her laptop computer and places the backup on the same hard drive. This is better than nothing, but if the disk crashes or the laptop is stolen, the backup is lost with the original. Indeed, even if the backup is made on a different device or medium, the backup should be kept offsite to protect against fire, flood, and theft.

Both of these caveats regarding appropriate backup methods further reinforce the main idea. In this case, how you *backup* information affects *whether or not* you can retrieve it. Although the examples illustrate situations students are likely to confront, businesses often fall prey to the same deficiencies when individuals are responsible for backing up information (if they back it up at all). When businesses store information on the cloud or in a centralized database, thus taking responsibility for backups out of the hands of rank-and-file employees, these problems can be diminished or avoided.

Representation and Compatibility

In addition to deciding where to place information physically and how to organize it logically, a third storage decision is how to "represent" it. Some might call this process "formatting" or "encoding." For example, web pages are represented using HTML. The choice of representation scheme raises various issues, one of which is compatibility, since the same information can be represented in different ways and one representation might be incompatible with another.

Consider the many file formats available to users of Microsoft Office applications when they invoke the "Save as" option for storing their documents, spreadsheets, presentations, and so forth. The very availability of all these formats highlights the problem, that one application might store information in a format that cannot be read by another. And that is why the Microsoft Office suite (and other applications) support so many formats—to enable users to store information from one application in a format that other applications can open. Adobe's PDF representation was also invented to overcome compatibility issues, originally between applications and different printers.

But some software applications store information in proprietary formats that are incompatible with other software applications. In the best case when such incompatibilities occur, information stored by one application must be converted to a format that works with the other. In the worst, case, there is no technology available for performing the conversion and the data are inaccessible by the other program.

XML affords another illustration of representation and compatibility issues. Because XML (eXtensible Markup Language) has become such an important component of modern IT, business students need to emerge from their programs with some rudimentary knowledge of what it is all about. But in a course for general business students,

making the significance of XML meaningful can be challenging. Once again, the fundamental principle can play a role. And even in a more advanced course, where students actually learn to write in XML, the principle can enrich their appreciation of it.

The purpose of XML is to facilitate the semantically rich exchange of documents among applications. Using XML as a universal language to represent and store information can eliminate the compatibility problems that accompany using propriety formats. For instance, XML has been used to represent web pages, word-processing documents, electronic spreadsheets, and so forth. Accounting and Finance students, in particular, may be interested in knowing that XBRL, a widely used standard for exchanging financial information, is XML-based. In short, XML—like the various file formats that applications use—illustrates a special case of the principle: How you *represent* information affects how you can *use* it.

Media Compatibility

Before leaving the topic of compatibility problems, here is an “oldie but goody”—an example that is largely outdated technologically but very poignantly highlights the compatibility problem as well as the storage/retrieval connection. Some instructors may, for good reason, choose not to include examples that are technologically obsolete for fear of putting off students or wasting class time. But older examples such as this can be useful both to demonstrate the longevity of the fundamental principle and to provide students with an historical context that is useful for appreciating today’s technologies.

Once a very popular form of secondary storage, floppy disks have been made largely obsolete by the move to the cloud as well as the benefits of USB memory sticks. But the history of floppy disks teaches an important lesson regarding compatibility. First there were 8” floppies, followed by the more widely used 5.25” variety. The first of these were single-sided, but soon there were double-sided floppies and double-density floppies (that stored twice the information in the same space). While the disks all looked alike, the newer varieties (that had more capacity due to double the surfaces or double the density or both) could not be read on drives designed for the older types. Then along came the 3.5” floppies, but over time their density kept increasing, again causing backward compatibility problems. And, of course, a 5.25” floppy was useless on a computer with only 3.5” drives and vice versa.

This situation—an artifact, largely, of technological advancement—created two very real compatibility issues. At any point in time, there were typically multiple floppy disk formats in practice and people who recorded information on one type were out of luck if they needed to retrieve it on a machine that did not support this format. In addition, people often stored information on floppy disks that they might need in the future. But today any of these drives are hard to find. Clearly, floppy disk users—and at one time that was nearly everyone—often experienced problems stemming from the incompatibility of the media they used for storage with the devices available for retrieval. This problem was often encountered in universities, where the computers in one lab might be incompatible with those in another or where the lab computers might be incompatible with students’ personal machines.

Compression

Data compression reduces a given set of information to fewer bits by applying an algorithm to the original representation (compressing the information). The compressed data are then expanded (uncompressed) to return the data back to its original form when the information is needed. In the case of files, compression is often referred to as “zipping” the files. Compressing information is advantageous not only because the stored information occupies less space but also because less transmission capacity is required when retrieving the information. This retrieval benefit, however, may be partially or fully offset by the need to uncompress the files, which may inconvenience the users and require some processing time by the computer. So, while data compression may be primarily about saving storage space, it does have implications for retrieval, as well, as the fundamental principle would suggest. Compression can also raise compatibility issues. In particular, self-extracting “zip” files intended for the Windows PC cannot self-extract on the Apple Macintosh, inconveniencing those who retrieve them. This example is especially vivid in courses where teaching materials are made available as self-extracting zipped files and many students have Macs.

Starfield Displays and Dynamic Queries

Shneiderman and his students (Ahlberg & Shneiderman, 1994; Ahlberg, Williamson, & Shneiderman, 1992) developed a novel interface for visual information seeking (VIS) that has now become standard practice for many websites. Users manipulate such interface controls as sliders, buttons, and check boxes to filter rapidly information depicted in a starfield display (two-dimensional scatterplot). For instance, this feature might be useful when searching a real estate database for homes satisfying various conditions or when searching a parking database for garages in Manhattan. While the key innovations are the direct manipulation widgets, the two-dimensional starfield display, and the connection between them, the rapid update of the display in response to user manipulations (upon which the system depends) is made possible by cleverly storing the data in a manner consistent with the various query capabilities. This powerful retrieval system illustrates the fundamental concept, since its effectiveness depends, in part, on how the information is stored.

Database Management Concepts

Database Design

Many core IS courses include hands-on use of a relational database management system (DBMS) such as Microsoft Access or MySQL. The purpose of this module is twofold: (1) to make the concepts more real and meaningful for students by having the students apply them to actual data and (2) to provide the students with hands-on skills that they may find useful in the workplace. Nonetheless, many students do not fully appreciate this material and some resist it. By having already bought into the fundamental principle, however, that principle can be used to help them understand the value of this material.

Some students question why they need to care at all about how databases are designed. The answer is that they do not have to care inherently about design for its own sake, but because they do care about retrieving information, the principle tells them that they must also care about design. As future business people, they need to ensure that the databases supporting their work are well-designed in the sense that they will provide them with the information they need, accurately, conveniently, and in a timely matter. They may therefore need or want to participate directly in the design of the databases and they can only do so effectively if they understand the issues.

Students also sometimes wonder why they need to learn a DBMS when they already know how to use an electronic spreadsheet. “Why can’t we just use Excel?” Indeed, the designers of Lotus 1-2-3, at one time the de facto standard for spreadsheet software, originally envisioned the program as supporting database management. But a spreadsheet—consisting of just a bunch of rows and columns—does not afford enough structure to manage data effectively. Because spreadsheets have no notions of “entities,” a given row might describe two different entities. For instance, each row might capture one course registration—information about a given student and a specific course he or she is taking. Each student would appear in potentially many rows, as would each course. This *redundancy* is problematic. If a student’s contact information appears with each course he or she is taking, and if that student moves, the changes might take place in some rows and not others, leading to inconsistent and inaccurate information. And if a student is enrolled in only a single course and drops that course, deleting the row will lose all of his or her contact information. So—referring to the fundamental principle—*storing* information in a spreadsheet is a bad way to manage data because, in the end, it is likely one will be unable to *retrieve* the information one needs accurately. An instructor can demonstrate this to the class with a large spreadsheet, pointing out the problems of redundancy, inconsistency, and the potential for lost data.

More generally, it is important for students to understand the problem of redundant information — the same information being stored in multiple different repositories within an organization. In some contexts "redundancy" is considered a good thing in IT — for example, having two identical processors—because if one system fails then the other is available as a replacement. But redundant data is problematic because, sooner or later, data updates are made in one place and not the other, leading to inconsistency in the data. So, once again, the lesson for students is that how you store your data matters!

Normalization is often a difficult concept for students and is typically covered in a more advanced class. While the fundamental principle does not make the intricacies of the normal forms any simpler, it does provide a motivation for the need to normalize. Database design, in general, and normalization, in particular, are all about structuring data in such a way that the database successfully captures the real-world phenomena of interest. One way to think about poor database design or a lack of normalization is that these weaknesses lead to information being stored in such a way that it cannot be retrieved properly later. This insight may be a useful introduction to teaching normalization.

Magnetic Tapes vs. Disks

Back in the days when magnetic tapes and magnetic disks were alternative media for storing currently active data, choosing between them was a practical issue of significance for businesses as well as a simple and classic example of the fundamental principle. Tapes are sequential access devices, whereas disks are direct access devices, so if one needs to pluck information for a given entity (a customer, supplier, product, or whatever) from the middle of a dataset, one can do so relatively easily if the dataset is stored on disk but not if it is stored on tape. In short, since a magnetic tape is one long strip, it requires the relatively slow process of winding through much of the tape to reach the desired location. In contrast, since a magnetic disk rotates, with any given sector passing under the read-write arm thousands of times per minute, the latency for reaching the desired piece of information is dramatically reduced. This difference makes disks ideal when direct access capabilities are needed and makes magnetic tapes impractical in such situations.

Today, due to this deficiency, magnetic tapes are no longer used for storing active data. Nonetheless, they do provide a striking and easy-to-understand example of the principle. Moreover, while some claim that tapes are entirely obsolete, others point to their ongoing value for backup and archival purposes (Nuncik, 2016). What tapes lack in speed and direct access capabilities, they make up for in cost and reliability. Cost is, of course, always important in business and reliability is also a significant business consideration, especially since data are often archived for compliance purposes. So the trade-off between tapes and other media lives on as a practical concern and an example of how storage choices affect future retrieval (since such retrieval depends on the reliability of the media). Indeed, this example can be used in the context of, or even to raise, a discussion about the need for reliable data archiving in the current climate of compliance concerns.

While the basic differences between tapes and disks are no longer so important as they once were, the distinction does serve as an easy-to-understand example of the principle. In fact, it may be worth using as one of the first examples. If you store information on tapes, you cannot easily access a given piece of information directly. So, you must store information on a direct-access storage device (DASD) such as a disk if you want to be able to do so.

Physical Storage Structures

Once students appreciate the key difference between tapes and disks they generally conclude that if you store a student database on disk you can pluck any given student's data from the middle of the disk when he or she has a problem with a bill or with registration. But they are often stumped when asked how one knows where in the middle of that disk to look for a given student's data. Indeed, just storing the information on a DASD such as a magnetic or optical disk or USB memory stick is not sufficient to be able to retrieve information directly. Absent any further effort in storing the information, one would still need to search the device sequentially.

It is by employing additional storage structures, such as indexes or hashing algorithms, that one is able to retrieve information directly from the middle of a disk. This observation provides a second, even better, example of the principle. By undertaking extra effort each time a record is stored—updating an index or applying the hashing algorithm—the computer is able to retrieve that information much more rapidly when it is needed. Since information is likely to be stored once and retrieved many times, and since retrieving it rapidly may be important, the extra effort in storing the information is worthwhile. And, if it is desired to retrieve information from the student database rapidly based on various criteria—majors, hometowns, and so forth—secondary indexes are required, once again illustrating that how you store information affects how you can retrieve it.

Today, core IS courses tend to be less technical than in the not too distant past and indexes and hashing algorithms are admittedly on the more technical side. Nonetheless, some students are eager for some more technical material. Alternatively, this application of the principle might be better suited for an upper-level course.

Replication of a Distributed Database

When one chooses to store information on a network rather than locally, other storage choices still remain. For example, should the data be centralized (in one location on the network) or distributed (over multiple locations)? One form of distributed database is the replicated database where identical information is stored in multiple locations to shorten the access path to the data. Maintaining replicated databases creates a challenge for updates because all the versions need to be kept in synch (otherwise one has the problematic situation of inconsistencies that occur with redundant data). Those who choose the added effort—and expense—of maintaining replicated databases do so to reap the retrieval benefits of having the data stored closer to the users.

Concepts with a Twist on the Principle

Passwords and Encryption

Normally one engages in extra effort when storing information to make accessing the information later easier. But given security and privacy concerns, more and more often these days we want to make it more difficult for unauthorized parties to access private or confidential information. This is true both in the corporate and personal realms. Password protection, encryption, and other security measures are implemented as barriers to the unauthorized retrieval of information and make it more difficult and effortful for everyone, both authorized users and intruders, to retrieve information. Nonetheless, the principle still holds — how you store the information affects how you can retrieve it—but now the added effort in storing the information is intended to make retrieving the information more difficult in order to restrict access. Ironically, extra effort is expended both in storing and retrieving the information, but the effort is justified by the objective of protecting the information. Thus, this example represents a twist on the general principle.

Archival Storage Off-Line

In our data rich and digital world, more and more organizations are needing to maintain large amounts of historic data that are rarely accessed. Despite advances in the capacity and economy of on-line storage of various kinds, given the vast volumes of such archival data, organizations typically make the choice to store such archival information off-line. This, too, is a twist on the basic principle. They are choosing to make some information more difficult to access in exchange for either reducing total storage costs or speeding up access to the information that remains on-line. This example might be used in conjunction with the example of magnetic tapes versus disks, since tapes are typically the media employed when archiving data.

Topics from the Information Systems Curriculum

Systems Analysis and Design

Many core IS courses cover the Systems Development Lifecycle (SDLC) — the so-called “Waterfall” methodology that breaks the lifecycle of an information system into a series of distinct and sequential stages. Certainly the more advanced course on “Systems Analysis” presents the SDLC. The information requirements analysis phase occurs early, at a pivotal point in the traditional design lifecycle, because we must first determine who needs to retrieve what information where and when in order to design its collection and storage accordingly. Once again, storage choices are driven by retrieval requirements.

A classic failure in this arena raised the issue naturally in my course one term. At that time, our students’ user IDs for the BlackBoard course management system were the same as their e-mail IDs but the default passwords were different. Many of the first-term freshmen taking the introductory course did not even know their IDs, let alone the two different passwords. When our associate dean requested a list for faculty of their students’ user IDs, he was told by the Information Systems tech team that such a list could not be generated by the system. Clearly the information is stored somewhere — how else could the system log them in successfully? — but no one anticipated a very natural need for this information. Ironically, individual instructors using BlackBoard could do what the tech team could not: produce a list of user IDs for the students in their courses. This salient example is one that students remember.

Data Warehouses

Gallaugher (2015) defines a data warehouse as a “set of databases designed to support decision making in an organization” (p. 347). In particular, data warehouses are often the repositories of information for Business Analytics, a prominent subject (and “hot topic”) in many core IS courses and in the broader IS/IT and business curricula. The fundamental principle can be used to explain the existence and role of data warehouses. After all, students who have already learned about the value of large centralized corporate databases may wonder why one would have a separate data warehouse (and possibly several smaller data marts).

One reason that businesses maintain data warehouses is that Business Analytics and decision support require data to be structured differently from the way data are stored in corporate databases that support transaction processing or record-keeping. This practice can easily be explained (perhaps a bit simplistically) to students who are familiar both with relational databases and with pivot tables (or, more generally, OLAP), a common form of data analytics where data are structured as a multi-dimensional data cube. Clearly, the data in the relational databases are not structured as multi-dimensional data cubes. A second reason that businesses employ data warehouses is that Business Analytics and decision support draw upon data that must be integrated from various sources inside and outside the company, including current and legacy systems. These systems may be incompatible and structure data differently. So, for both reasons, data requires restructuring. This process is typically referred to as ETLing the data —Extracting it from its source location, Transforming it as necessary, and Loading it into the data warehouse.

What does this have to do with the fundamental principle? Without the data warehouse, the data are not stored in a way that facilitates Business Analytics. By storing the data in a Data Warehouse via the ETL process, the data are now stored in a way that they can be accessed as needed. Data marts—mini-data warehouses, if you will—further support this concept by storing the data specifically needed for a given functional area or a given problem such as Finance or Marketing.

Information Architecture, Enterprise Systems (ERP), and Data Warehouses

Today’s emphasis on corporate information architectures and ERP systems reflects the same principle—storing data in a way that they are available for later use—with yet another twist. In the old world of IT, we designed specific applications and designed our systems to collect and store the specific information that would be required by those who used the application. Now we are recognizing an organizational need for flexible information retrieval not necessarily targeted at a particular application. Our information architectures, ERP systems, and data warehouses are designed to provide support not only for today’s information needs and portfolio of applications but for those of the future as well. The challenge of designing these systems is to provide sufficiently generic support that both meets today’s needs and anticipates those of the future. The fundamental principle is unchanged: We need to store information in a form that we can retrieve it later as needed. But while in the past that meant very specific storage targeted at specific types of retrieval, today it often means the even greater challenge of designing more generic storage that can meet a variety of unspecified retrieval needs.

DISCUSSION POINTS

Beyond shedding light on many specific concepts in the course, the fundamental principle also raises several broader issues:

- Like most topics in business and technology, and as noted early in this paper, storing information with an eye toward retrieval involves a number of considerations that often conflict one with another. A critical challenge for business people and technologists, therefore, is making the trade-offs among considerations and arriving at appropriate storage decisions in any given situation. It follows that important career preparation for business students in a core IS/IT course is understanding those trade-offs and how to make them. Introducing the consequential factors at the time the principle is first introduced—cost, availability, compatibility, veracity, and security—positions the students to appreciate the nuances associated with the various information management topics covered in the course. By the end of the term, once the various concepts and illustrations have all been placed on the table, returning

to the issue of how to trade-off these considerations can lead to a fruitful and enlightening capstone discussion. One could also turn this topic into an in-class debate surrounding the trade-offs in a given situation or a case-based assignment. Trading-off the pros and cons of cloud storage might be an especially effective theme. Table 1(a) summarizes the storage choices and Table 1(b) summarizes the factors and considerations that influence those choices.

- As seen in the examples, we are experiencing a weakening of the storage-retrieval link in at least two arenas. First, those who store information on the world wide web are often independent of those who retrieve it. From the perspective of web surfers, the matter of how web pages were stored is largely unseen. Second, the old model of developing applications was to develop files or a database specifically for the application. Increasingly, however, organizations are developing information architectures intended to support current as well as future applications. In both instances, the data store is not designed with specific retrieval needs in mind. But, as already noted, while these phenomena may change the nature of the storage-retrieval connection, they do not eliminate it. The fundamental principle remains intact, although its implications have changed. Indeed, the looser coupling of storage and retrieval makes the problem of storing information effectively for future retrieval even more challenging.
- While the changing nature of the storage-retrieval relationship may not undermine the underlying principle, it does lead us to ask, “How enduring is this principle that today is so fundamental?” The explosion of information on the Internet and the concomitant difficulty of finding what you need within the enormous worldwide information base have led to much interest in retrieval technologies (often referred to as “search”). Might the combination of improved retrieval techniques, dramatically increased computing power, and networked (cloud) storage reduce the impact of how information is stored on how it is retrieved? If so, how would a disturbance of this basic principle ripple through the rest of the Information Systems field?

Table 1. Major Storage Choices and Considerations

Panel A. Major Storage Choices	
Physical Location	<ul style="list-style-type: none"> • Computer (Desktop or Laptop) • Mobile Device (Smart Phone or Tablet) • Removable Media <ul style="list-style-type: none"> ○ Floppy Disk ○ Optical Disk (CD/DVD) ○ USB Memory Stick • The Cloud
Logical Location	<ul style="list-style-type: none"> • Directory (Folder) Structure • Naming Conventions
Representation Format	
Panel B. Major Storage Considerations	
	<ul style="list-style-type: none"> • Cost — the financial cost of storing, maintaining, and subsequently retrieving the information. • Availability — how easily, if at all, the information can be retrieved when, where, by whom, and in the form it is needed. • Compatibility — whether the information is stored in a way that is compatible with the hardware and software systems that will be used to access it. • Veracity — whether or not the information is correct. • Security — how vulnerable the information is to theft or loss.

These issues offer fertile ground for discussions at the end of the term, assignments, or final exam questions.

CONCLUSION

Business students studying Information Systems and Information Technology could benefit from the following fundamental principle, which is a foundation for much of what they learn about these topics:

How you store information affects how you can retrieve it.

The significance of this principle lies in its implication for storing information:

Because being able to retrieve and use information when we need it, where we need it, and in the form we need it is important, *how we store information matters!* Exercising care and expending effort when storing information is essential to facilitate the all-important retrieval.

While most visible in the context of database management, this fundamental principle represents one of those transcendent ideas seen throughout the course. Why do we care at all about the design of information systems? Because we want to ensure that organizations—and the people in them—are equipped to use information effectively to meet their needs. And one key factor that enables or constrains how organizations use information is how it is stored.

Since this IT principle is a special case of a broader principle for life, it can be introduced at the beginning of the term before any IT material has been covered. Clean laundry and car keys can be used to illustrate the more general principle. Then it can be applied as each of the IT concepts and examples comes along. Those presented in this paper are summarized in Table 2. Examples of the principle are so widespread that students can be encouraged to look for, and present to the class, examples from their own use of IT. They can also be asked to find examples of the principle in the business, technical, or popular press. Identifying and analyzing such examples can help the students understand the principle better and, most importantly, help them apply it in the future. Examples identified by students are often especially effective when presenting the principle in subsequent terms.

As we have seen from some of the examples, the principle is richer than the first version presented at the outset of this paper. More fully, one might state it as follows:

How you represent, organize, and physically store information affects if and how you can retrieve and use that information.

The students would benefit from confronting the more complete (but less pithy) version at some point during the term.

Why is this principle important for thinking about Information Technology? Because the principle is so widely applicable, it can contribute to students’ appreciation of course material, to the technological choices they are making currently, and to their information-related behavior in the future. Many IT practitioners and educators have internalized this principle over years of experience. Focusing on the principle in the introductory IS course for business students will enable novices to incorporate it in their emerging perspectives of the subject.

Table 2. Concepts and Examples that Draw on the Fundamental Principle

Concepts and Examples from Students’ Daily Lives	<ul style="list-style-type: none"> • Organizing Class Notes • Personal File Organization • The Cloud • Finding Information on the WorldWide Web • Backup • Representation and Compatibility • Media Compatibility • Compression • Starfield Displays and Dynamic Queries
Database Management Concepts	<ul style="list-style-type: none"> • Database Design • Magnetic Tapes vs. Disks • Physical Storage Structures • Replication of a Distributed Database
Concepts with a Twist on the Principle	<ul style="list-style-type: none"> • Passwords and Encryption • Archival Storage Off-Line
Topics from the Information Systems Curriculum	<ul style="list-style-type: none"> • Systems Analysis and Design • Data Warehouses • Information Architecture, Enterprise Systems (ERP), and Data Warehouses

ACKNOWLEDGEMENTS

M. Lynne Markus, Dominik Molitor, Jie Ren, David Salisbury, Erika Schwartz, Tamilla Mavlanova Triantoro, and Sidne Ward provided valuable feedback on earlier versions of this paper.

AUTHOR BIOGRAPHY

Mark S. Silver is Associate Professor and former Area Chair of Information Systems in the Gabelli School of Business at Fordham University. He was previously a faculty member at the Anderson Graduate School of Management at UCLA and the Stern School of Business at NYU. Professor Silver received his B.A. in Mathematics from the University of Pennsylvania and his Ph.D. in Decision Sciences from The Wharton School. Silver is the author of a book, *Systems That Support Decision Makers: Description and Analysis* (Wiley, 1991), among other publications.

REFERENCES

- Ahlberg, C., & Shneiderman, B. (1994). Visual information seeking: Tight coupling of dynamic query filters with Starfield displays. In *Proceedings of CHI '94* (313-317). New York: Association for Computing Machinery.
- Ahlberg, C., Williamson, C., & Shneiderman, B. (1992). Dynamic queries for information exploration: An implementation and evaluation. In *Proceedings of CHI '92* (619-626). New York: Association for Computing Machinery.
- Gallaugh, J. (2015). *Information systems: A manager's guide to harnessing technology*. v. 4.0, Flatworld Knowledge.
- Hershey, G. L. (2002). A different focus and content for the core information systems course for business school majors. *Communications of the Association for Information Systems*, 12, 479-493.
- Ives, B., Valacich, J. S., Watson, R. T., Zmud, R. W., Alavi, M., Baskerville, R., Baroudi, J. J., Beath, C., ... Whinston, A. B. (2002). What every business student needs to know about information systems. *Communications of the Association for Information Systems*, 9, 467-477.
- Nuncik, M. (2016, February 12). Why we still use magnetic tapes today. *Kroll Ontrack*. Retrieved from <http://blog.krollontrack.co.uk/concepts-explained/companies-have-difficulties-keeping-their-old-legacy-tape-infrastructure-alive-and-need-a-better-solution/>.
- Redman, T. C. (2016, September 22). Bad data costs the U.S. \$3 trillion per year. *Harvard Business Review*. Retrieved from <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>
- Salisbury, W. D., Huber, M., Piercy, C., & Elder, K.L. (2004). The AMCIS 2003 panels on IS Education-I, let us not throw out the baby with the bath water: Information, systems, and technology. All matter in the Core IS course. *Communications of the Association for Information Systems*, 14, 128-146.
- Silver, M. S., M. L. Markus, & C. M. Beath. (1995). The information technology interaction model: A foundation for the MBA core course. *MIS Quarterly*, 19, 361-390.
- The Four V's of Big Data. (n.d.). IBM. Retrieved from <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>
- Topi, H., Valacich, J. S., Wright, R. T., Kaiser, K. M., Nunamaker, J. F., Jr., Sipior, J. C., & de Vreede, G. J. (2010). IS 2010: Curriculum guidelines for undergraduate degree programs in information systems. association for computing machinery and association for information systems. Retrieved from <http://www.acm.org/education/curricula/IS%202010%20ACM%20final.pdf>.