https://doi.org/10.48009/4_iis_2019_45-55

Issues in Information Systems
Volume 20, Issue 4, pp. 45-55, 2019

THE ANCIENT SECRETS OF INFORMATION MANAGEMENT

Ronald Fuller, The Institute for Logic and the Public Interest, rgfuller@logicandbusiness.org
Matthew North, Utah Valley University, mnorth@uvu.edu
Peter Cardon, University of Southern California, cardon@marshall.usc.edu

ABSTRACT

The lost art of classical logic is the key to overcoming the most difficult and costly challenges of modern organizations. This becomes easy to understand when we recognize the central role of logic in both modern and legacy systems. We theorize that the study and implementation of foundational principles of logic will improve more than just data storage and retrieval—it will fundamentally alter for the better the management of unprecedented amounts of information found in modern organizations. In this paper, we layout the theoretical underpinnings for adding classical logic to information systems design and development, without dictating any specific implementation methodology.

We show that people have used classical logic to organize information for more than 700 years, and that the logical structure of information should be determined by its owner, not by a toolmaker or a technical specialist. Based on our findings from historical uses of classical logic to guide the development of non-computerized business systems, we argue that a renewed emphasis on logic education and logic literacy will bring significant benefits to modern organizations that struggle to manage massive amounts of information effectively.

Keywords: Logic, Information Management, Data Governance, Data Quality, Integration, Next Generation Systems

INTRODUCTION

People who manage information have known for decades that classical logic can be used to query databases, but the use of logic to *organize* information within those databases is not well understood in the field of information science. Fricke (2012) states, "The monumental and authoritative Encyclopedia of Library and Information Sciences, Third Edition, 2009, does not have an entry for logic in its 6,856 pages."

We will show that medieval business practitioners used classical logic to organize information, and that they understood that the structure of information should be determined by its owner, not by a toolmaker or a technical specialist. These facts are supported by the system of interconnected books used in manual accounting systems such as double-entry bookkeeping since the 13th century.

Primitive business institutions did not have the benefit of computers, but they enjoyed a particular advantage over their modern counterparts: a logic literate merchant class. Based on our findings, we argue that a renewed emphasis on logic education and logic literacy will bring significant benefits to modern organizations that struggle to manage information effectively.

The greatest challenges to information management from a business perspective include integration, data quality, and adapting to changing requirements with reasonable cost and effort. Difficulties in these areas are typically attributed to misalignment between business units and IT departments. From a technical perspective, the same challenges are often attributed to deficiencies of the SQL language, its implementation by vendors, and limitations of the relational model. This paper presents an alternative view that addresses both perspectives: The greatest challenges of effective information management are not a result of departmental misalignment or technical issues; rather, they are a result of the decline of classical logic education during the 20th century, and mistaken assumptions about who should decide how information is organized.

Logic literacy should not be seen as equivalent to general or natural intelligence. Classical logic is based on a set of laws and principles that can be taught, studied and learned. In some ways learning classical logic is like learning to read and write – both involve a set of rules, a set of symbols (or concepts that can be represented by symbols) and learning to recognize and construct patterns based on the rules and symbols (or concepts). For much of modern history classical logic was taught to young children, and was considered to be a necessary foundation for all further learning (Joseph and McGlinn). Logic was eliminated as a required subject in schools during the progressive education reforms beginning around the turn of the 20th century. This was a political decision (Prawat). Logic literacy declined during the first half of the 20th century and computers were developed during the second half. Ironically, computers are based on classical logic. Computers do not decrease the value of logic literacy, they increase it significantly, just as the Guttenberg press increased the value of standard literacy. There is a movement afoot today to restore classical logic as an important part of basic education (Genesereth and Chaudhri). We wholeheartedly endorse this movement and believe that the business education community will be among its greatest beneficiaries.

Logic and Information Systems

When the relational model was introduced in the 1970's it was falsely seen as a new computer-based method of organizing information, rather than a new computer-based way to *automate* old *logic-based* methods that had been used with success since Medieval times (Fuller, 2018). Consequently, the responsibility to decide how information is organized shifted to computer experts and architects and away from domain experts.

Regarding the way information is organized, our focus here is on the initial organized state and the final states that can be derived from them. Intermediate states are important too but are not discussed here. The promise of the relational model from the beginning was that information could be stored in a way that allows it to be reorganized for various purposes while guaranteeing that outputs are consistent with the inputs. This ideal is rarely achieved in current practice, and it cannot be accomplished through technology alone. It is possible only when information is deliberately organized in a way that logically entails the range of desired outputs. Both logic literacy and domain expertise are necessary to organize information in this way (Fricke, 2012).

Structure is Significant

Codd wrote that the relational model "provides a means of describing data with its natural structure only" (Codd, 1970). The word 'only', Codd explained, means "without superimposing any additional structure for machine representation purposes." He did not explain the meaning of "natural structure," but the implication is a structure consistent with classical first-order logic, which is the theoretical foundation of the relational model. Perhaps he chose the word natural because classical logic is the basis of natural deduction and has some characteristics of natural language which allow information owners to precisely express their intended meaning. Structure is significant because the way information is organized determines how it can be used. Structure encodes meaning and intent. The initial state of relational information determines the full range of all possible subsequent states that can be derived from it.

Terms and Concepts

We will use the following terms and concepts to illustrate the applicability of classical logic to information management:

- *Information resource*: An information resource is information organized for some purpose. It can be any kind of information organized for any purpose, from a memorized phone number to the entire internet.
- *Relational information*: Relational information is information organized in a manner consistent with classical first-order logic, such that it can be manipulated using the standard operators and the classical rules of inference. The structure of relational information can be described in terms of a *logical signature* or *vocabulary* (defined below). Relational information can be organized in many ways to serve many purposes. Note: when logicians speak of *relational logic* or a *relational signature*, it means that no functions are used (and therefore much of the expressive power of classical logic is not used). But the relational model almost always implies the use of functions. In terms of logic, a relation can take one or many inputs and produce one or many outputs, and a function produces only a single output for any given input. All functions are relations, but not vice versa. Codd's choice of the term *relational model* to describe a theory of information management based on the full expressive capacity of classical logic is an unfortunate accident of history and it can lead to confusion when discussing it in terms of logic. The term relational information can likewise be misunderstood, but it is familiar and we use it here to emphasize its connection with the relational model.

- *Account*: An account is a kind of information resource. It is a descriptive record of the resources and transactions of an enterprise which is organized for the purpose of management and analysis. Many accounts are relational. Many others are intended to be relational but are not because they do not conform to the rules and principles of the relational model.

The key to understanding an information resource is to recognize that the way information is organized determines how it can be used. Deciding how information should be organized requires a degree of subject matter expertise specific to the kind of information resource at hand. The concept of *sophotaxis* will add clarity to further discussion (Fuller & Cardon, 2016).

Sophotaxis
Sophotaxis (Greek: wisdom + order) is essentially the process of organizing information in a way that allows it to be used according to the intent of its owner. The degree of subject matter expertise required to achieve this condition can be described in terms of *sophotaxy*. The following usage examples illustrate these terms in contexts of business, information science, and logic:

- In business, the process of sorting documents in a file cabinet does not require any understanding about the files or their purpose, only an awareness of the sorting key; this is a *non-sophotaxical* task. Deciding whether the cabinet should be organized by last name or customer ID requires some knowledge of how the files will be used, but not any understanding of their content; this is a *semi-sophotaxical* decision. Determining what kind of information will be recorded on the forms, how they will be arranged, and the number of carbon copy sheets to be attached to each form requires a detailed understanding of the business process involved, the purpose of the forms, and how they will be used. Such business forms are *highly sophotaxical* information resources.
- In library and information science, classification schemes are defined by subject matter experts. After a scheme has been defined, a librarian can use it to catalog an information item with bibliographic references. After an item is cataloged it can be placed on a shelf by a volunteer or a machine according to the catalog identifier. These are highly, semi, and non-sophotaxical processes, respectively.
- In logic, sophotaxis combines the notion of entailment with that of intent. If a logical signature can be used to express an intended set of statements and their antecedents, then it is sophotaxically complete; if it cannot, then it is sophotaxically deranged or incomplete (Fuller and Cardon).

The *sopho* in sophotaxis implies a certain meaning for wisdom: Unlike knowledge, wisdom cannot be readily transferred from one person to another. Wisdom combines knowledge with perspective, experience, and exposure to consequences – all of which are needed when determining a logical vocabulary for an enterprise organization. Therefore, logical vocabularies should be created not by engineers or architects, but by subject matter experts who possess wisdom that can only come from significant firsthand experience in the organizations that own, produce, and consume the information.

Logical Vocabulary
In terms of logic, a *logical vocabulary* describes the properties of a given system and determines the range of statements that can be made about the objects in discussion, and the range of conclusions that can be inferred from those statements. A logical vocabulary consists of four parts (Genesereth & Kao, 2013):
- a nonempty set of object constants
- a possibly empty set of function constants
- a nonempty set of relation constants
- an assignment of arities for each of the function constants and relation constants

The Fundamental Four

In terms of information management, the same notion of a logical vocabulary is established by answering the following questions:

- What tables will be created?

- What columns will each table have?
- What rules will apply to the values in each column?
- What relationships will exist between and within the tables and columns?

These are the *fundamental four*: tables, columns, rules, and relationships. Objects are expressed as tables. Their attributes are expressed as columns. Relationships to other objects are expressed as dependencies. Types and integrity constraints are expressed as rules and relationships, as are identity management and classification schemes. In common practice, these critical owner-management decisions are delegated to technical specialists such as architects, database administrators, or programmers. This is a mistake which assures bad outcomes from the beginning. It comes from the belief that the fundamental four represent some kind of abstraction or model of the owner's requirement. Not so – the fundamental four represent the minimal and essential elements of any requirement expressed for the purpose of defining a relational information resource. If the owner does not provide them, then no requirement has been expressed – at least not for a relational information resource. The requirements that organizations typically do express when they want an information resource are instead those for automated systems, which are not the same as information resources. They are, rather, tools used to access and maintain information resources, as well as automate various processes and workflows. Information resources and automated systems are closely related, but the distinction is important.

Quantification

Quantification means expressing information in its smallest meaningful components so it can be analyzed and transformed, using the classical rules of inference, into new expressions which can reveal new information while preserving the original values and meaning. Quantification is at the same time both highly subjective and perfectly objective in different respects. It is subjective in the sense that what constitutes the smallest meaningful component of something can vary between people, circumstances, and requirements. It is objective in the sense that it creates precise expressions that can be consumed by logical operations or computer programs. Figure 1 illustrates the subjective nature of quantification:

Phone number
(425) 245-6586
(206) 867-9204
(303) 521-7348
(970) 356-4775
(741) 622-9438

CC	Area	Excg	Line
+1	425	245	6586
+1	206	867	9204
+1	303	521	7348
+1	970	356	4775
+1	741	622	9438

Figure 1. Subjective Quantification

The table on the left expresses telephone numbers as single values. For some organizations this is a perfectly reasonable way to represent phone numbers. When the only purpose of the numbers is to make phone calls, the numbers here are already expressed in their smallest meaningful form and there is no need to break them down into smaller parts. For other organizations, however, the individual components of each number might be meaningful. Users may want to count or group numbers by area code, for example. The table on the right represents the smallest meaningful components in this case. Breaking them down even further, perhaps into 11 columns with only a single digit each, not only would serve no purpose, but would add unnecessary complexity and conceal valuable information. Knowing how to express information in its smallest meaningful form is highly dependent on specialized subject matter expertise and a thorough understanding of how the information will be used. Quantification in this sense is the essence of sophotaxis.

Data Literacy vs Information Literacy

Euclid defined data as *that which is given, and which can be determined from what is given*. If we were to receive a business requirement that specifies "A implies B", we can derive from this the logically equivalent term "If not B, then not A." The ability to interpret and manipulate data in this way is what comes to mind when we consider the term *data literacy*. Information literacy adds to this the recognition that the structure and content of such statements each express different kinds of information (see Quantification). Logic tells us that the 2 statements above are equivalent, but logic cannot tell us if A really does imply B in any given situation. Domain expertise and classical logic are both necessary. Research and experience show that most people can gain a practical level of logic literacy with 2 to 3 weeks of study (Newcombe et al., 2015; Attridge et al., 2016).

An Account, Not a Model

Hugh Darwen has said “A database is an account of some enterprise, not a model of it” (2018). This distinction is critical to understanding the difference between an information resource and an automated system. An account is a kind of information resource, while an automated system is a tool which follows a model of various processes and workflows. When information is organized to reflect the owner’s intent, use, and valuation, it reflects logical creation of information resources rather than dogmatic adherence to a specific systems design standard.

If an automated tool will be used to access and maintain an account, the account should be defined first. When the owner of an account only expresses a requirement for a tool, and allows the tool designers to define the account, the account becomes a subcomponent of the system. Problems with data quality, integration, and adaptation to change are inevitable because those objectives are highly reliant on a sophotaxically correct relational vocabulary.

When accounts are sophotaxically correct they are easy to adapt to new requirements and easy to modify when requirements change. Subsequent logical derivations (in the form of queries, ETL jobs, reports, etc.) can be made with no reliance whatsoever on domain expertise. In contrast, when an account is sophotaxically deranged, domain expertise must be reintroduced at every point of change to ensure (or attempt to ensure) that information content and meaning are preserved. Working with sophotaxically deranged information is very much like playing a game of charades while blindfolded.

No More ‘Modeling’

With an understanding of classical logic and the principle of sophotaxis, the notion of data modeling has no further value. All the various modeling techniques developed over the previous decades such as object role modeling, entity relation modeling, etc., are entirely subsumed and surpassed. Sophotaxis is not a model or methodology, it is a set of principles that can be applied to many aspects of information management. In terms of logic, a model is an interpretation. In other words, a model is an assignment of values to variables. In terms of IS, a model in this sense corresponds to the values of a populated database, in contrast to the structure of an empty database (which corresponds to a logical signature).

Logical vocabularies can be expressed in many ways, including with text (such as SQL), symbolic (such as standard logic notation), and graphical (such as conceptual graphs), all of which are fully substitutable without any loss of information. A signature expressed using one notation can be translated to any other with perfect precision and accuracy. The ability to express information in many different ways using different systems of notation while preserving its meaning is a powerful aspect of classical logic.

The Relational Approach

Among the choices organizations have when deciding on an approach to organize information, the relational approach is the most highly sophotaxical of all. That means direct, firsthand subject matter expertise is required more urgently when organizing relational information than any other kind. The reason is simple: Relational information represents the logical vocabulary for an enterprise; and classical first-order logic is the most expressive form of knowledge representation that can be computationally evaluated by any person or machine. It provides a means of encoding wisdom and intent that is more reliable and precise than any other mode of communication. Properly derived outputs are guaranteed to be logically consistent with the inputs (Sowa, 2000).

Experts Only

Subject matter experts are native speakers of their own wisdom and intent, with their attendant perspective, experience, and sensitivity to consequences. Expertise in a given subject can take many years of firsthand experience to acquire. Such expertise comprises the vast majority of the total understanding necessary to define and create a sophotaxically correct relational vocabulary – in other words, to decide the fundamental four. In contrast, the rules of logic are only a small portion of the total understanding needed. They can be learned in a few days and mastered in a few weeks.

The common practice of delegating responsibility to define the fundamental four to technical specialists – relying on interviews with experts, or requirements expressed for automated systems – leads to predictably bad results for the following reasons:

- Subject matter expertise changes from organization to organization, job to job, and project to project. It can take years to acquire and it comprises the vast majority of the total understanding needed to define a vocabulary
- The rules of logic are static. They do not change from organization to organization, job to job, or project to project. They are easily acquired and represent a tiny fraction of the understanding needed to define a vocabulary
- The rules of logic must be applied according to the needs of the owner and subject matter experts

For these reasons it is not a good idea for technical specialists to decide how information is organized, unless it is information relating directly to their own field of expertise

Programmers have often imagined that classical logic, with its great expressive power, is too complex to be of much use to programmers, to say nothing of business professionals. But this perceived complexity actually comes from the difficulty of expressing, in great detail, complex information relating to areas they are unfamiliar with. This difficulty carries the illusion of technical complexity, when it is really nothing more than unfamiliarity with a given field of expertise such as the complex rules and processes of a large business organization. Chris Date (2011) explains this well in the following passage. We use the term *classical logic* in parentheses below to replace the term *predicate calculus* in the original text. Predicate calculus is a common notation for classical logic, which can be substituted for many other notations, including SQL, as noted above.

> "Codd's 1970 paper had proposed, among other things, that (classical logic) might reasonably be used as a basis for a data language (or data sublanguage, as Codd called it), and it gave examples of what expressions in such a language might look like. Unfortunately, it did so using the formal notation of (classical logic), notation that many people in the computing world at the time were quite unfamiliar with. So the idea arose that (classical logic) was just too difficult for ordinary mortals… It is true that the syntax of (classical logic) might look a little strange and complicated; but the concepts expressed by that notation… are actually quite simple, and can be explained quite simply too. In other words, it is not difficult to wrap some nice syntactic sugar around those concepts and make them very palatable indeed.
>
> Do you happen to be familiar with Query-By-Example (QBE)? QBE is exactly (classical logic)! In my opinion, it's a very user friendly syntactic sugar coating of the semantic ideas embodied in (classical logic)." (Date, 2011)

The principles of logic that make QBE an intuitive way to query information are the same principles that will allow logic-aware managers and logic-literate business professionals to solve their most difficult challenges managing information.

Plausible Desirability

A data architect with a small army of business analysts can spend hundreds of hours playing 20,000 questions with business experts to determine what they consider to be a plausibly desirable range of outcomes, and still not come up with a complete list. Scope creep is unavoidable with this approach. But when business experts define their own vocabularies the problem is significantly reduced or even completely eliminated. What is plausibly desirable to an owner or expert may change over time, but it will always change far more slowly than the *perceived* requirements when IT professionals must play the guessing game.

Even when a vocabulary is defined by a 3rd party, a logic-literate expert can readily determine if it will meet their needs. For this reason, a junior manager or analyst can take on the legwork of creating a vocabulary on behalf of an information owner, and the owner can then validate it when complete. This is the way financial budgets are managed: they are prepared by subordinates, and then approved by an accountable manager. Budgets are generally not accepted without careful review and firsthand approval from the responsible leader. It should be likewise with information resources. But this can only become possible with logic-literate business professionals and logic-aware management.

When integration is required, resources must be defined in collaboration with professionals within each of the related processes and subject matter areas. Deciding the fundamental four requires no knowledge whatsoever of any technical machine or software (except for advice from specialists on cost benefit issues, which are discussed in "Cost-Benefit Concerns" below).

Sophotaxical Derangement

Technical expertise can actively and severely impair one's ability to decide the fundamental four, aside from and exacerbating the fact of not having firsthand domain expertise, regardless of how well one may understand database software, which is irrelevant here.

The following example shows how eminently qualified technical experts at a world leading company, who are also themselves in this case the domain experts, are blinded by their technical perspective from obvious and severe problems with the fundamental four. They are the database architects of one of the world's largest technical content publishing systems. The system is built around two types of content items: articles and tables of contents (TOC's), each of which can participate in many-to-many relationships with the other. However, in this system all articles and TOC definitions are stored as XML-typed data in a single base table. The rationale for this decision is that "it's all XML data" therefore, it is all of the same type and therefore the same domain (no, it's not). As a result, the leaders of the organization are frequently unable, despite tremendous expense and effort, to answer simple questions about their content, such as "Which authors have written articles about a given topic?" Even the seemingly simple task of building a path of breadcrumb links for topical navigation requires an enormously complex stored procedure. The joke goes (even within this team) that the easiest way to find something within in this system is to Google it. As a result, they may end up just creating their own web-crawled key-value indexes like a search engine and forfeit any hope of achieving the richness, power, and flexibility that is only possible with truly relational information.

The problem here is that programming requires a mode of thinking that is quite different from the sophotaxical processes necessary to organize information. They can be difficult to clearly separate when working with both simultaneously – even in rare cases when the engineers are also the subject matter experts. This dilemma is further exacerbated because programmers and database engineers are just as unlikely to have learned classical logic as business professionals and other experts.

GIGO is Overrated

'Garbage in, garbage out (GIGO)" is a common explanation for poor data quality. But a far more frequent problem exists when information within a system is known to be complete and correct but cannot be used in a desired way, or formed into desired outputs, because it is not organized in a way that allows it. In other words, when it is sophotaxically deranged. This is also a primary root cause of integration failure. Information systems cannot be integrated unless the combined logical signatures are sophotaxically correct. That is possible only when the individual component systems are themselves sophotaxically correct and compatible with one another. Integration can only be achieved through the coordinated efforts of logic-aware management. No amount of effort by architects, engineers or consultants can succeed without this.

Moving data from an old platform to a new one is not the problem here. The real challenge is reorganizing information that is sophotaxically deranged, which has nothing to do with technology, or the age of any system or programming language. It has only to do with the way the information has been organized.

The U.S. Department of Defense spent 10 years and more than a billion dollars trying to integrate its payroll systems, and the project ended in complete failure with no change in capability or any useful outcome (Reuters). Sophotaxical derangement may also provide cover for fraud and abuse. When a system is unable to detect or verify errors in legitimate transactions, it seems unlikely that it can be used to detect fraud.

Without logic-aware management it is nearly impossible to integrate information systems, even for highly funded government agencies and world leading technology companies. Logic-literate business professionals and logic-aware management are necessary to create vocabularies to describe the accounts of any large enterprise.

Sophotaxical derangement can contribute to other factors relating to poor data quality as well. For example, the problems identified by Francis and Prevosto in *Data and Disaster: The Role of Data in the Financial Crisis* were cases of failure to obtain or analyze data. Sophotaxical derangement can be a significant factor in such failures. A 2010 Thomson Reuters and Lepus report claims that fragmented data platforms contributed to the 2008 financial crisis (FinExtra Research). Likewise, disjointed systems prevented the intelligence community from connecting the dots prior to the 911 terror attacks. Fragmented systems that cannot be integrated, but should be, are common evidence of sophotaxical derangement and present enormous opportunities for logic education in business.

Cost-benefit Concerns

In the process of defining a logical vocabulary, engineers and architects can provide valuable input on cost-benefit decisions involving issues like performance, storage, and scale. For example, an owner might define the fundamental four in a way that accomplishes the critical business process of identity management using alphanumeric codes rather than sequential integers. An experienced architect might advise that implementing this requirement could cause problems with performance because of the way some database software works. The owner then has a decision to make. Options might include: 1) accept the performance consequences, which could be very significant; 2) change their approach to identity management and use sequential integers instead of the codes; or 3) add extra columns with integers (surrogate keys) to make the software work faster, and accept the additional complexity and long-term management overhead of maintaining two (or more) separate identity management systems. Whatever the decision, it should be informed by specialists, but decided by information owners in view of the overall leadership objectives and priorities.

The need for cost-benefit analysis and related tradeoffs when defining the fundamental four is not new. Don Chamberlin, co-inventor of the SQL language, explained that the original IBM System R team designed SQL to give customers the option to enforce certain rules only when they choose to. One reason he cited was that enforcing the rules takes additional processing and storage, which in the mid 1970's was quite expensive. He explained, *"When the original SQL designers decided to allow users the options of handling nulls and duplicates, they viewed these features as minor conveniences, not as major departures from orthodoxy, taken at the risk of excommunication"* (Chamberlain, 1996).

But that decision did not force anyone to do or ignore anything. Rather, it gave users the freedom to make their own cost-benefit decisions. Unfortunately, to this day those kinds of decisions do not often fall to the people who bear the cost of poor choices and are properly positioned to make good ones. So, not only are highly sophotaxical decisions being made by non-subject matter experts, but – adding insult to injury – they are often implemented in a defective manner that makes the problems even worse. Sophotaxical derangement is a pervasive aspect of production information systems everywhere, and presents a vast frontier of research opportunity for logic oriented IS scholars. The first such efforts should focus on characterizing and cataloging the various modes of failure.

Triumph of 'the Systems of Men'

Thomas Haigh (2001) explains why information management is treated as a technical discipline rather than a management one:

> "The systems men were the members of the Systems and Procedures Association (SPA) during the 1950's and 60's. They offered corporate executives an implicit bargain: 'You put us in charge and we'll deliver to you more power over your firms than you've ever dreamed of". The papers presented at their International Systems meetings…exhibit a fixation on questions of status and power. But executives for the most part were not convinced that technical skills should translate into management authority. The systems men claimed to possess a body of objective knowledge and techniques qualifying them to make superior decisions within a particular technical domain. But their task of legitimation was uniquely difficult because their claimed domain was management itself. Only during the 1980s did the term MIS become so tainted by failure, reflecting the persistent reality of computer work's low status in the eyes of management that academics and management writers flocked to alternatives rather than to redefinition."

By the 1970's the SPA was defunct; but the systems men had left a stubborn cultural legacy that persists still today: the idea that managing information is a job for technicians. Haigh continues:

> "For better or worse, to speak of something as an information system continues to imply that it should be engineered by an information specialist and built using information technology. It seems unlikely that the idea of information can ever truly be separated from these roots: it is just too historically and culturally charged."

To overcome this illusion, we must look to the past.

Sophotaxis in the Middle Ages

Double entry bookkeeping (DEB) has been used in commerce since around 1296. It has two defining characteristics:

1. A system of interconnected books which are cross-referenced and correlated. We'll refer to this as *relational bookkeeping* for reasons explained below
2. A process of dual entry where every transaction is recorded twice, once as a debit and once as a credit, for the purposes of checking math. This aspect is not significant for our analysis here

Hoskin, Ma, and Macve (2015) explain that these two characteristics were used separately at different times and places through history. Accounting historians disagree on which aspect is more significant. However, the use of both together is always indicated by the term double entry bookkeeping. This unfortunate fact inevitably clouds the debate. We propose the term *relational bookkeeping* to specify the system of interconnected books to eliminate this ambiguity.

Relational bookkeeping (i.e., the system of interconnected books), even as practiced in the Middle Ages, is a practical application of classical first-order logic, and every object and process can be described in terms of modern relational databases. Below are some DEB terms with their corresponding modern counterparts:

- Journals and inventories are relations (transaction tables)
- The repertory or finding key is a relation (reference table)
- Book entries are n-tuples (rows)
- Columns are domains (attributes)
- Ledger accounts are filtered views of journals, and therefore they are also relations
- Ledger accounts refer to journal entries by way of page reference numbers, which are essentially foreign key constraints for each account
- Balance sheets and other financial statements are filtered or aggregated views of the ledger, capital and cash accounts
- The rules of DEB provide consistency and referential integrity
- A Chart of accounts is a system catalog, or metadata dictionary

Every aspect of the relational model is satisfied. With this it becomes clear that medieval merchants in northern Italy invented fully functional, manually operated relational database management systems, almost 700 years before the theoretical foundations of those systems were to be formally described and understood. Their leather-bound books and manual techniques served the same purposes and accomplished the same kinds of outcomes as modern databases and query languages – with perhaps even more reliable results due to higher rates of logic literacy and observance of sophotaxical principles.

The same paper-and-ink techniques could have been used to manage other kinds of information as well, but manual accounting systems are painstaking and time consuming, so it is easy to understand why their use were limited to managing the kind of information of greatest concern to early merchants.

Implications

These facts stand as evidence that computers and technology have not fundamentally changed the underlying nature of information management, as has been commonly assumed since the 1970's. They reestablish the previously long-understood fact that managing information is a management discipline, not a technology or even a hybrid discipline. They further establish that the discipline of information management is based on fundamental principles. And not just any ordinary fundamental principles, but the principles of classical logic, which many mathematicians, philosophers, and scholars consider to be the most fundamental of all principles. Kurt Gödel, the best friend of Albert Einstein (and whom their colleagues considered his intellectual equal) described logic as *"*a science prior to all others, which contains the ideas and principles underlying all sciences.*"* Compare his use of words to the following statements by economist Werner Sombart in 1919:

> *"*Double-entry bookkeeping came from the same spirit which produced the systems of Galileo and Newton, and the subject matter of modern physics and chemistry. . . "Double-entry bookkeeping is based on ... the basic principle of quantification which has delivered up to us all the wonders of nature, and which appeared here for the first time in human history in all its clarity." (Most, 1970)

Sombart and others were roundly criticized for supposedly overstating the nature and significance of double-entry bookkeeping; but now, with recognition of its congruence with classical first-order logic, those criticisms should be reconsidered.

CONCLUSION

The ancient secrets of information management consist of two important lessons:

Ancient Secret #1
Classical logic can be used to organize information. The modern field of library and information science, for which the organization of information is a primary concern, has only recently begun to investigate this useful fact. Fricke argues, *"It should be possible to give an adequate account of classification using only First Order Logic."* And a review of Fricke supports our claim that the task of using logic to organize information should fall to subject matter experts rather than information specialists (Rasmussen , 2013).

Practitioners of commerce have relied on the expressiveness and computational power of classical logic to organize information for many centuries. But logic is not enough. A vocabulary is also needed, which must come from logic-literate business professionals and managers, as expounded in secret #2.

Ancient Secret #2
Decisions about organizing information should be made by logic-literate subject matter experts. The medieval merchants who practiced relational bookkeeping certainly did not look to tool makers to define their ledgers and accounts. But that is what modern organizations routinely do. It makes no difference that the old tools were made from parchment, feathers, and dye, and the new ones from computers, software, and networks. The old tools served precisely the same purpose as the new with respect to the definition and management of information resources. The new tools serve an additional purpose of automating processes and workflows, but that is no reason to believe that the engineers who create the tools should also be the ones to define the accounts, and there are important reasons to understand why they should not.

REFERENCES

Attridge, N., Aberdein, A. & Inglis, M. (2016). Does Studying Logic Improve Logical Reasoning? In: *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education*, Szeged, Hungary, 3-7.

Chamberlin, D. (1996). *Using the new DB2*. San Francisco: Morgan Kaufmann.

Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM, 13*(6), 377–387.

Darwen, H. (2006, January). *The Askew Wall*. Retrieved from www.dcs.warwick.ac.uk/~hugh/TTM/TTM-TheAskewWall-printable.pdf

Date, C. J. (2011). *SQL and relational theory*. Sebastopol: O'Reilly Media.

FinExtra Research. “After Financial Crisis, Firms Look to Improve Data Quality - Thomson Reuters Report.” *Finextra Research*, 2 Mar. 2010

Fricke, M. H. (2012). *Logic and the organization of information*. New York: Springer.

Fuller, Ronald (2018). “First-Order Logic In13th-Century Accounting Systems.” *The Bulletin of Symbolic Logic*, vol. 24, 2018, p. 209, doi:10.1017/bsl.2018.12.

Fuller, R., & Cardon, P. (2016). Sophotaxis. *The Bulletin of Symbolic Logic, 23*(1), 128.

Geijsbeek, J. (1914). *Ancient doubleentry bookkeeping*. Geijsbeek, Denver. Includes translation of Pacioli, L. 1494. Summa de arithmetica, geometrica. Proportioni et proportionalita. Paganini, Venice.

Genesereth, M., & Kao, E. (2013). *Introduction to Logic (2nd ed.)*. Williston, VT: Morgan & Claypool.

Genesereth, M. & Chaudhri, V. Logic in Secondary Education. http://intrologic.stanford.edu/miscellaneous/intro.html. Accessed 28 January 2019.

Haigh, T. (2001). *Inventing information systems: The systems men and the computer, 1950–1968*. *Business History Review, 75*, 1561.

Hoskin, K., Ma, D., & Macve, R. (2015). *A genealogy of myths about the rationality of accounting in the West and in the East* (London School of Economics working paper). Cited with permission.

Joseph, Miriam, and Marguerite McGlinn. *The Trivium: The Liberal Arts of Logic, Grammar, and Rhetoric : The Understanding the Nature and Function of Language*. Paul Dry Books, 2002.

Most, K.S. (1970). *The role of accounting in the economic development of the modern state*. Ph.D. dissertation, University of Florida. Includes translation from Sombart, W. 1919. Der Moderne Kapitalismus. 3d ed. Duncker and Hurablot, Munich and Leipzig.

Newcombe, C., Rath, T. Zhang, F., Bogdan, M., Brooker, M. & Deardeuff, M. (2015). How Amazon Web Services Uses Formal Methods. *Communications of the ACM, 58*(4), 66–73.

Prawat, R. S. (2003). The Nominalism Versus Realism Debate: Toward a Philosophical Rather Than a Political Resolution. *Educational Theory*, *53*(3), 275–311.

Rasmussen, D. (2013). Review of Logic and the Organization of Information by Martin Frické. *Canadian Journal of Information and Library Science, 37*(1).

Reuters. (2013). Special Report: How the Pentagon’s Payroll Quagmire Traps America’s... *Reuters*, 9 July 2013. *www.reuters.com*, https://www.reuters.com/article/us-usa-pentagon-payerrors-special-report-idUSBRE96818I20130709.

Sowa, J. (2019). *Notes on the History of Logic.* http://www.jfsowa.com/peirce/hist_log.htm. Accessed 7 March 2019.