Reflections on Knowledge, Communication and Knowledge Organization†

Alan Gilchrist

32 Friar Road Brighton BN1 6NH, United Kingdom,
<alangilchrist77@gmail.com>

Alan Gilchrist started his career in an industrial information department before joining the Research Department of Aslib (Association of Special Libraries and Information Bureaux) working mainly on systems analysis and manually operated information retrieval systems. In 1977 he left Aslib to set up his own consultancy practice working for, and through, international agencies such as the EEC, UNESCO and OECD. Alan Gilchrist was the founding editor of the *Journal of Information Science*. He has written and lectured extensively and in 2006 was awarded an honorary doctorate by the University of Brighton for services to information science.

Gilchrist, Alan. **Reflections on Knowledge, Communication and Knowledge Organization.** *Knowledge Organization. 42(6)*, 456-469. 23 references.

Abstract: Moments after mankind started to make moveable physical records describing the world about them people started to store and arrange the records in logical order. Thereafter, records were copied and translated and simple catalogues of collections were compiled. Others, working from these records and from oral sources compiled lists, dictionaries and encyclopaedias. As the means of communication developed from the first revolutionary invention of writing to other revolutions in communication methods, notably printing with moveable type and the computer, techniques of knowledge organization became more sophisticated and powerful. In the first half of the twenty-first century we are faced with an unprecedented communications overload and the full range of knowledge organization techniques need to be deployed, further developed and applied.

Received: 31 July 2015; Revised: 20 August 2015; Accepted: 20 August 2015

Keywords: information, knowledge organization, classification.

† I would like to thank the Conference Organizing Committee for the honour of inviting me to give this keynote paper; and to the referees for their useful comments.

1.0 Introduction

This paper is the fuller version of the keynote address to the ISKO UK Conference of July 2015 commissioned by the organizing committee with the wide brief to "review the importance that knowledge organization has had in the past." This is a dauntingly wide brief and as the title implies this is not a thorough academic review but a subjective, superficial and inevitably personal survey of what this author sees as some of the major milestones of concern to knowledge organization (KO) from its beginnings to the present. Man has communicated since very early times, initially with grunts leading to speech, with cave paintings and signs, before the emergence of writing, followed by the invention of printing by moveable type, right up to the computer, advances that sometimes incurred unease and even hostility. Soon after records were made it became useful to store them for later use and reflection, and as these stores grew in size it became necessary to introduce some order into their arrangement and even to create aids for the location of specific items. Thus, knowledge organization is a fundamental activity that is thousands of years old, keeping up with advances in knowledge and developing in tune with the advances in methods of communication.

2.0 Definitions

Those working in the field of KO are well aware of the opacity of the words "information" and "knowledge," but in a world that trumpets phrases such as the "information society" and the "knowledge economy" it is worth reminding ourselves of the abstract nature of these two words. The historian Roszak said in his book *The Cult of Informa-*

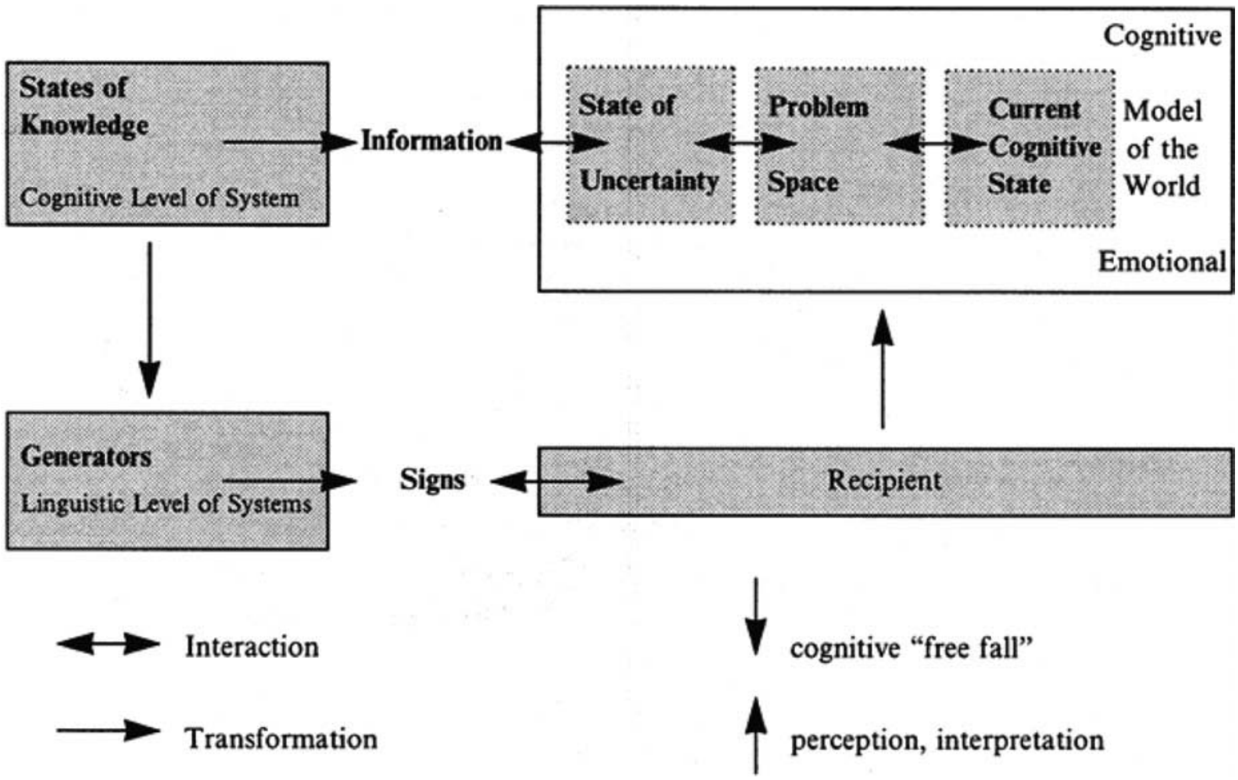


Figure 1. The cognitive communication system for information science and information retrieval (Ingwersen 1996, 6) .

tion (1986), "Information has come to denote whatever can be coded for transmission through a channel that connects a source with a receiver, regardless of semantic content."

At the individual level an understanding of the interaction of information and knowledge is vital to the understanding and work of KO. Many authors have discussed this, supported by diagrams of the interaction; the one here proposed by Ingwersen (1996) neatly bringing together the two concepts within a framework that also includes the vital concepts of communication and language. To be even clearer, the word communication may here be alternatively considered as "a message" as proposed by Belkin and Robertson (1976).

There is, of course also consensual knowledge, where for some aspect of reality it can be said "It is generally believed that ...," indicating that there is a group of people who at some moment agree on some hypothesis or set of beliefs. However, such consensus is not always permanent as new evidence and new hypotheses emerge, giving rise to such statements as "It was previously thought that ...." The task faced by KO across the board is to organize all these historical and current messages into a logical framework. Not an easy task.

The task is made even more difficult by the complexity of language and its constantly changing nature, so that extracting meaning from messages may not be straightforward and unambiguous.

The linguistic problem has been succinctly summarised by Liddy (1998) in her work on natural language processing. When Liddy published a list of semantic levels a software developer estimated that putting into practice her ideas for advanced information retrieval would require computing power, which was rarely to be found in the USA at that time. More recently her ideas have become more viable with the greater sophistication of computers. Her seven levels are:

1. **Phonological**: interpretation of speech sounds within and across words
2. **Morphological**: componential analysis of words, including prefixes, suffixes and roots
3. **Lexical**: word level analysis including lexical meaning and part of speech analysis
4. **Syntactic**: analysis of words in a sentence in order to uncover the grammatical structure of the sentence
5. **Semantic**: determining the possible meanings of a sentence, including disambiguation of words in context
6. **Discourse**: interpreting structure and meaning conveyed by texts larger than a sentence
7. **Pragmatic**: understanding the purposeful use of language in situations, particularly those aspects of language, which require world knowledge.

It can be seen that the levels become increasingly complex in terms of potential meaning of a piece of text, with "discourse" working at any level larger than a sentence (which may be long) and "pragmatic" which includes metonymy and all the oddities of specific and local languages such as the "White House" to indicate the seat of the American government. Ordinary KO, of course, does not work with accuracy at such fine levels of granularity as can be inferred from Liddy's list and I think it was Vickery who coined the expression "aboutness" to reflect the level at which it can and does work.

Finally in this section, the SECI Model proposed by Nonaka and Takeuchi (1995) is reproduced in Figure 2 to show four transformations in communication, based on the two modes of tacit (oral) and explicit (recorded). The transformations are "tacit"→"tacit," called Socialization; "tacit"→"explicit," called Externalization; "explicit"→ "explicit," called Combination; and finally "explicit"→ "tacit," called internalization. Two of these have a direct bearing on KO, a third is of interest and the fourth, in the domain of knowledge management and more recently electronic social media, is also of concern.

3.0 Knowledge and communication–a brief history

3.1 Writing

Following the oral tradition ("tacit"→"tacit"), the first life-changing revolution in communication arrived with the invention of writing around 2600 BCE. The Sumerians, who had settled in the fertile crescent of Mesopotamia some millennia before, improved on an earlier and primitive prototype incising characters on soft clay tablets before baking them hard. Initially, it was thought that this writing was a gift from the gods so that only the king and his priests were allowed to write. It was not very long, however, before this restriction was lifted as it became clear that writing was extremely useful for the production of records in trade and administration.

The technique then spread to the Akkadians, friendly neighbours of the Sumerians who were of a different ethnicity and spoke a different language. In the seventh century BCE the Akkadian king Ashurbanipal established a vast library at Nineveh stocked with records derived from previous rulers, the collection including the famous *Epic of Gilgamesh*. Ashurbanipal then ordered his librarians (the Sumerian word for librarian was "man of the written tablets") his translators and scribes, all of whom were well regarded and paid accordingly, to compile a bilingual dictionary of the Sumerian and Akkadian languages ("tacit"→"explicit") one which, moreover, incorporated a rudimentary classification. Sample entries, as noted by Finkel in his fascinating book (2014), included under the heading Sheep such entries as: "Sheep with arthritic hips," and "Sheep given to butting." (This looks like a very early application of the twenty-first century ISO Standard 25964 on thesauri and interoperability).

Clay tablets with cuneiform writing spread to the

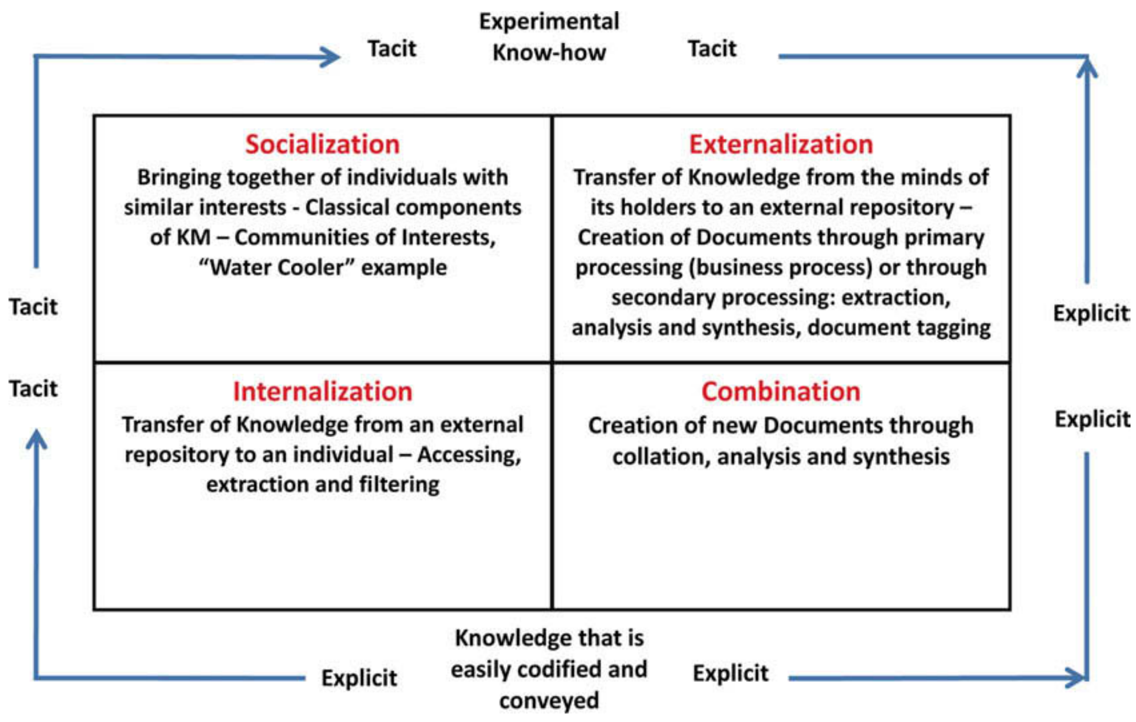


Figure 2. The SECI model (Nonaka and Takeuchi 1995).

Babylonians, Hittites and Assyrians before dying out in the first millennium BCE, to be replaced by papyrus, which had been made for centuries in Egypt, from the extensive reed beds of the Nile and widely exported. Then came those great international traders the Phoenicians who invented the first modern phonetic alphabet in about the seventh century BCE, used for Aramaic and soon adopted and adapted by the Greeks, followed by the Romans, Hebrews and Arabs, and eventually by all the languages of modern Europe.

3.2 Greeks and Romans

It was papyrus that was used for the collection of the Library of Alexandria, established by Ptolemy I to be the greatest in the world. To achieve this, any books carried by ships docking in the port of Alexandria were temporarily impounded, copied and translated into Greek. When his library was rivalled by that at Pergamon, it is said that Ptolemy cut off the supply of papyrus, which library then switched to using parchment. We even know that the chief cataloguer of the library of Alexandria was called Callimachus, who arranged the stock according to genre and the subjects: rhetoric, law, epic, tragedy, comedy, lyric poetry, history, medicine, mathematics, and natural sciences; each of these subjects was then subdivided by author. There was also that safety class of miscellaneous–reminiscent of the cartoon in the *New Yorker* showing a row of filing cabinets with the first drawer labelled A-Z, and all the others miscellaneous. Each scroll carried bibliographic details and the containers in which they were kept carried tablets called *Pinakes* each with a set of index lists.

The library was destroyed either in a single operation, or more likely, over a period of years into the first century CE; scholars still debate the cause ascribing it variously to Julius Caesar, the Arabs or restless Christians uncomfortable with pagan writings. The Romans followed the Greeks in the building of library collections and it was possibly Julius Caesar who introduced public libraries in Rome. Private libraries were so popular that teams of twelve or more scribes would take dictation from a speaker reading from popular scrolls. Recent discoveries have been made of scorched papyri at Pompeii and it is hoped that with modern analytical techniques it may be possible to decipher parts of these scrolls.

3.3 The monasteries

Rome finally fell to the Vandals in the fifth century CE and many of the libraries were destroyed. However, some were removed to safe places by members of the Roman aristocracy in their flight, and more found their way to the growing number of monasteries springing up across Europe, notably one at Vivarium established by a Roman aristocrat called Cassiodorus, and, later, another at the famous Monte Cassino by Saint Benedict, founder of the Benedictines who were to become widespread and influential in Northern Europe as educators and technical instructors. Many of these monasteries were, in time, equipped with scriptoria, and manuscripts were often lent to other monasteries secured by a deposit of money or another book so that their monks could make copies. Consequently, some of these collections became increasingly large, even containing scrolls by pagan authors such as Aristotle and Plato. Copying later became "secularised" as commercial scribes took commissions, and there is a twelfth century manuscript *Derivations* by Hungaris of Pisa, of which 200 copies still exist. After the monasteries came the universities, largely devoted initially to theology; Bologna in Italy in 1088, Oxford in England in 1096 and Salamanca in Spain in 1134, among others. The Vatican Library was officially founded in 1475, though it had been in existence for many years previously. This library is one of the most significant collections in the Western world and now holds some 75,000 codices and 1.1 million books including 8,500 incunabula on subjects including history, law and science as well as, of course, theology. Originally, the library was organized by shelf lists recorded in notebooks, but in the period 1927-1932 a card catalogue system for a part of the collection was instituted using the Library of Congress *Classification*. More recently a programme of digitization has been launched, presumably supported by search software.

3.4 The golden age of Islam

Within a relatively short time there came an astonishingly rich period of some few hundred years, between the middle of the eighth century to the middle of the thirteenth, when the Arabs dominated the Western world intellectually, scientifically and culturally in what became known as the Golden Age of Islam. In Baghdad a centre of learning and a library, similar to the destroyed Library of Alexandria was established called The House of Wisdom, where philosophers, astronomers and scientists gathered to debate and to translate ancient and modern works into Arabic from Farsi, Hebrew, Aramaic, Syriac, Greek, Latin and even Sanskrit. With a spread to the West, Cordoba in what was then known as al-Andalus, became a centre of learning comparable to Baghdad. Scholars and translators came from all over Europe to work in the relatively free society of Cordoba open to Jews and Christians alike, all busy translating the classics into Arabic, Hebrew and Latin. Greatly aiding this industry was the importation, some centuries after its inception, of the Chinese invention of paper and improved by the Arabs by the addition of starch in their paper mills of Baghdad and al-Andalus, making it tougher for the Arabic pens preferred to the Chinese use of brushes. It was an extraordinary age for the advancement of science in a wide range of subjects including philosophy (partly in the classical sense of natural philosophy), astronomy (including the invention of the astrolabe), mathematics (including the invention of algebra and the introduction of the Hindu-Arabic numeral system), optics and the medical sciences. Cordoba boasted hospitals open 24 hours a day, every day, and which were free to all. One of the leading experts in medicine was the polymath Avicenna, living in the late tenth and early eleventh centuries, Avicenna compiled a huge work called *The Canon of Medicine* (“explicit”→“explicit,”) which brought together all that was known from the works of the Greek, Roman, Persian and Indian authors together with the latest thinking by contemporary Arabic authors including his own. This extraordinary work was in use for some hundreds of years in the main universities of Europe. With some 300 libraries in Cordoba, the largest is estimated to have had a collection of some 400,000 books compared with the largest in Christian Europe with a meagre 400. Such a collection must have been organized for easy access to the scholars working in three languages, though little is known of the details. It is only fairly recently that wider and proper credit has been given to this productive era which some now say laid the foundations for the Italian Renaissance. In 1258 the Mongols sacked Baghdad completely destroying the Grand Library, while after a period of decline al-Andalus was finally secured in 1492 by King Ferdinand II of Aragon and his wife Isabella I of Castile, patrons of Christopher Columbus.

3.5 The Renaissance and printing

Between the fall of Baghdad and the ending of the Moorish dominance in Spain the Italian Renaissance was fast developing, starting in the thirteenth century and spreading throughout Europe in the years to the seventeenth. A key feature in the birth of this Renaissance was the Humanists, a loose school of followers of the scholar Petrarch who had discovered a collection of letters of Cicero. This inspired many other scholars to search out classical Greek and Roman manuscripts, travelling to remote monasteries in Europe accompanied by their scribes. The resulting influx of classical prose and poetry influenced the basis of Renaissance thinking, and with that came art and sculpture as personified by Michelangelo and science by his contemporary but older Leonardo Da Vinci. At the same time, mediaeval universities, taking over some of the activities of the monasteries were developing and broadening their coverage from the theological to a broader curriculum influenced by the classical scholars such as Plato and Aristotle. A new educational framework consisted of the “seven liberal arts,” the trivium (grammar, logic and rhetoric) followed by the quadrivium (arithmetic, geometry, music and astronomy). Now came an acceleration of progress in science and technology as the Renaissance spread throughout Europe. Following Leonardo Da Vinci, known as “the father of modern science” came Copernicus in Poland contesting the geocentric view, Francis Bacon in England (credited with being one of the originators of the scientific method), and Galileo who narrowly avoided execution for his heretical beliefs following Copernicus. The most important innovation, from the point of view of this essay, came in 1455 with the invention of moveable type, the second major revolution in communication technology after writing. Eventually causing a huge increase in literacy and reading beyond the aristocracy, it coincided with the teachings of Luther and the Reformation with the Bible now made more widely available in Latin. Paper had been used in Europe for some time for use in handwriting and block printing and was now used by Gutenberg in printing the Bible. It is interesting to note that he printed 150 copies on paper and it is estimated that the 30 he printed on the traditional parchment each required 300 sheepskins, a clear example of cost-effectiveness.

The European Renaissance with its huge advances in science and learning was followed immediately by the so-called Age of Enlightenment. This was the age in which, to give it its full name, the “Royal Society of London for Improving Natural Knowledge” was formed in England in 1660, publishing one of the first scientific journals in 1665, the *Philosophical Transactions of the Royal Society,* very soon after the French *Journal des sçavans.* One of the early Presidents of the Society was Isaac Newton, one of many illustrious scientists of the time, including Leibniz and Lavoisier.

The advent of printing spurred the growth of libraries and the seventeenth and eighteenth centuries were what has been called the “Golden Age of Libraries” with many being established all over Europe, including the famous Bodleian Library of Oxford University established in 1602 on the basis of much older collections and open to scholars world-wide. Public subscription and lending libraries became common, including the Manchester Chetham Library in England which, founded in 1653, claims to be the first public library in the anglophone world; though the Malatesta Novella library established in Italy in 1452 claims to be the first civic library belonging to the commune rather than the church. Then came the national libraries; for example the British Museum in 1753 (later incorporated into the British Library), though the forerunner of what was to become known as the Bibliothèque nationale de France has a longer and more complex history. Again, all or most of these libraries had useful shelf arrangements backed by usually handwritten lists or catalogues.

3.6 The industrial revolutions

By now, one new movement was overlapping the last and so the two industrial revolutions started around 1760 and continued into the early to middle years of the nineteenth century. Coal, iron and steam power were the driving forces, the last providing the energy for the mass production of newspapers, thus further increasing communication and literacy. Almost immediately, the first industrial revolution was overtaken by the second, introducing steel and railroads, chemicals and mass production. New sources of energy became available: oil and electricity, the second leading to the invention of electric communication by the telegraph and the telephone, inventions, which were not always greeted with enthusiasm. When Edison visited London in the 1870s to promote interest in his apparatus, no less a person than the Chief Engineer of the Post Office exclaimed (author's personal notes): "I fancy the descriptions we get of its use in America are a little exaggerated; but there are conditions in America which necessitate the use of instruments of this kind more than here. Here we have a superabundance of messengers, errand boys, and things of that kind." In fact, this was the start of the third communications revolution heralding the arrival of the radio and television, computers, the Internet, emails and the World Wide Web all having a fundamental effect on KO. Many government libraries were established in the 1800s and in the wake of the industrial revolutions a growing number of special and industrial libraries were set up from 1900 onwards displaying more advanced knowledge organization.

3.7 Computers

Computing as a process is, of course, far older than our modern idea of a computer. The abacus is several thousand years old and analogue devices were used in astronomy from that period on, being advanced and refined by Indians, Persians and the Arabs from the tenth century and continued by mediaeval scientists in Europe. Babbage, from 1822 to 1837 was one of the first to develop the idea of the mechanical computer with his Difference Engine and Analytical Engine, neither of which were completed, but were demonstrated. At one such demonstration, a Member of Parliament asked "Pray, Mr Babbage, if you put into the machine wrong figures, will the right answers come out?," which prompted Babbage to reflect that "I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question" (author's personal notes). Consequently, the idea of the processing of data was firmly entrenched before the idea of the possibility of processing text; hence the surprise of computer scientists when early information scientists introduced the idea of the inverted file as an alternative to the data dictionary for text processing. Even today, in the U.K. enterprises, the chief information officer should more properly be called the chief information technology officer and will probably be more concerned with data than with unstructured information. Alan Turing is credited with being one of the first to conceive of modern programmable computing with his hypothetical Turing Machine proposed in 1938, a machine that printed symbols on paper tape in a manner that emulated a person following a series of logical instructions. After 1938, developments in computer design, accelerated in part by World War II were rapid, notably with the set of Colossus computers used at Bletchley Park to decipher German codes. In 1946 the Americans launched the ENIAC computer, the first large scale, general purpose, fully programmable digital computer. This machine had 20,000 vacuum tubes and weighed some 50 tons, so that Thomas J. Watson, IBM President, can be forgiven in retrospect for saying in the early 1940s (Watson 1943, cited in Remenyi 2003, 55): "I think there is a world market for about five computers." Regardless of this apparent pessimism, the euphoria persisted as can be seen in the prescient article by Vannevar Bush, published in 1945, in which he said (108): "Wholly new forms of encyclopaedias will appear, ready made with a mesh of associative trails running through them, ready to be dropped into the memex and there amplified."

In 1946 the Royal Society held a meeting under the title "Royal Society Empire Scientific Conference" which consisted of papers not only on scientific research but on information services. One of the key players in the second of these themes was the scientist and Royal Society Fellow J.D. Bernal, who also attended the seminal Washington Conference of 1958 devoted entirely to scientific information. This Conference heard 75 papers in seven areas, of which three were given to aspects of (author's personal notes) "organization for information search and retrieval," presented mainly by speakers from the U.S.A and the U.K., but a few from other European countries. Bernal was now sufficiently enthused to persuade the Association of Special Libraries and Information Bureaux (Aslib) that had been established as long ago as 1924, to establish a Research Department and to further persuade the British government to provide funds from the same pot that financed a range of scientific research associations. The first Director of the Aslib Research Department was C. W. Hanson who had previously been Information Officer for the Scientific Research Instruments Research Association.

The Department lasted for over 20 years and produced three professors: Stephen Robertson (an originator of Bayesian information retrieval), Blaise Cronin (promoter of Social Informatics) and Brian Vickery (renowned expert in information retrieval and honoured by the ISKO-UK biennial Conference of 2011). In the U.K., the Institute of Information Scientists was formed in 1958 by Jason Farradane and colleagues, followed in 1967 by the first course in information science in the world at what is now the City University in London. Both the Institute and the course recognized information as being central to information science and what was later to be known as knowledge organization to be central to information retrieval.

4.0 Knowledge organization before the computer

4.1 Dictionaries and encyclopaedias

Not long after the invention of writing people were compiling lists, lists of their deities and of natural objects in the world around them; and these developed into more complex forms such as the bilingual dictionary produced in Akkadia mentioned earlier in this paper. Thereafter dictionaries (of words) and encyclopaedias (of subjects) were increasingly produced. An early and massive encyclopaedia was produced by Isidore of Seville in the sixth century CE, consisting of 448 chapters in 20 volumes, much of it written by himself or taken from Greek and Roman sources. Though much of it was reasonable for the time, he was occasionally carried away by the exotic as can be seen from an entry that read (author's personal notes): "The Cynocephali are so called because they have dog's heads and their very barking betrays them as beasts rather than men."

The Age of Enlightenment saw a marked increase in the compilation of dictionaries and encyclopaedias in most of the European countries. Perhaps the most well known lexicographer was Dr. Samuel Johnson, poet, essayist and biographer who compiled his famous *Dictionary of the English Language* in the eighteenth century. This was an enormous task as Johnson wryly noted in his definition of Lexicographer (author's personal notes): "A harmless drudge that busies himself in tracing the original, and detailing the significance of words." He was also most conscious of the often ephemeral nature of words and their meanings. As an indication of the huge advance made by this dictionary one can compare it to what at the time was a best-selling dictionary compiled by one Nathan Bailey earlier in the eighteenth century. Two entries in this work read (author's personal notes):

Black–A colour
Dog–An animal well known

In eighteenth-century France a group of some hundred intellectuals (including Voltaire) contributed to the *Encyclopèdie, ou dictionnaire raisonné des sciences, des arts et des metiers* aimed at the advancement of science and rational thinking. This work, inspired by the earlier *Cyclopaedia, or Universal Dictionary of Arts and Sciences* (compiled by the Englishman Ephraim Chambers) was taken much further by Denis Diderot, editor, contributor and co-founder with Jean d'Alembert. It was ambitiously intended to incorporate all of the world's knowledge and in Diderot's words (author's personal notes) "to change the way people think." At the start of the nineteenth century, Peter Mark Roget, a doctor by training, created his ground-breaking work with the long title *Thesaurus of English Words and Phrases Classified and Arranged so as to Facilitate the Expression of Ideas and Assist in Literary Composition*. The work was created in 1850 but not released till 1852 since when it has greatly expanded and seen many editions. Roget suffered from depression and it is said that he worked on his thesaurus to occupy his mind. Now there are very many encyclopaedias and monolingual and bilingual dictionaries in all countries with such famous publishers as Larousse in France and the Oxford University Press in Britain. *The Oxford English Dictionary* (the *OED*), started in 1888 was last published in 1989 with 21,728 pages in 20 volumes, though electronic versions have since been published; the online version quarterly. It is a prescriptive, rather than descriptive, dictionary recording not only usage but a record of the development of the English language over time and the ways in which words subtly change their meanings. Many of these later works have been digitized and are available on the World Wide Web, some, like the *OED*, with clever interactive features.

4.2 Classification

Not long after lists were being created, their contents were being sorted into logical and helpful order: a natural intellectual activity with a very long history. In the fourth century BCE Aristotle worked, as a philosopher and in natural philosophy, both on the principles of logic and classification and most famously on their application to a hierarchical typology of animals. Of the many attempts at creating a universal classification over the years one created by Francis Bacon in the seventeenth century is shown at Figure 3.

The discovery and colonisation of other countries outside Europe created a great interest in the collection and classification of plants, leading to the greatest name and achievement in this area: Carl Linnaeus the Swedish botanist and his classification of plants laid out in several publications in the eighteenth century. Now, his classification framework contains nineteen hierarchical levels ending in

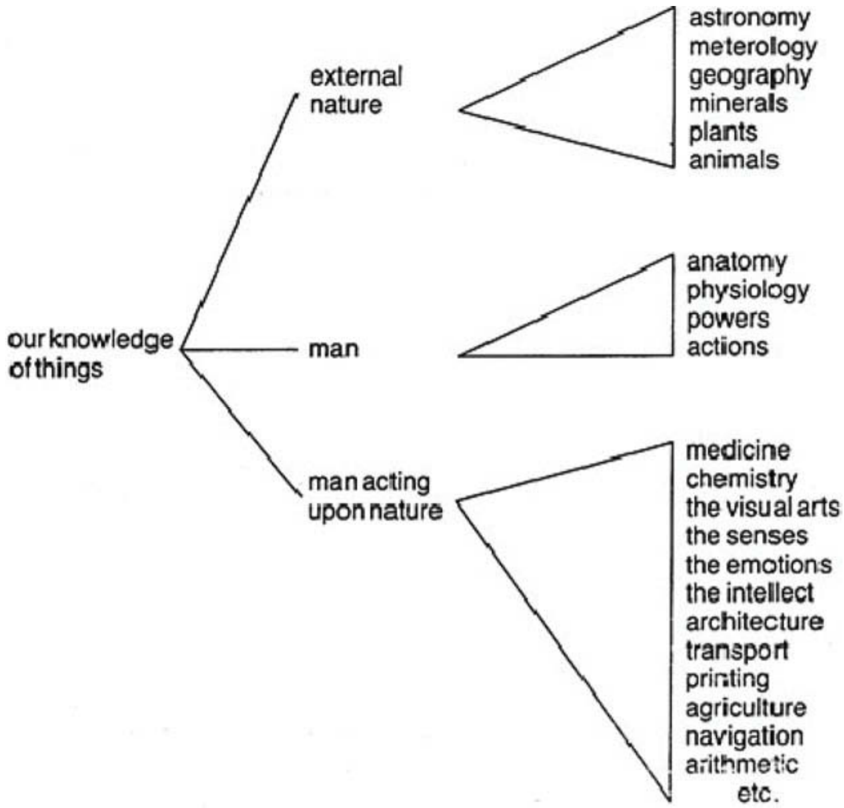


Figure 3. Francis Bacon's taxonomy of knowledge (reproduced from McArthur 1986).

the binomial Latin construct of, for example, *Rosa canina* for the Dog-rose. In the following century, the Russian chemist Dimitri Mendeleev devised a two dimensional classification of the elements which correctly predicted several elements that had not at that time been identified, thus guiding the work of contemporary chemists.

Though bibliographic classification, often in crude form, was practised since the establishment of the first libraries, it was not till the nineteenth century that the first great general classifications were compiled, four of which are briefly described below. The first of these, published in 1876, is the *Decimal Classification* by Melvil Dewey (his given name was Melville, but as a proponent of abbreviation and phonetic spelling he changed it). Dewey was a librarian who, as a young man, was given the task of rearranging the stock in the Amherst Library in America. Starting apparently from the taxonomy devised by Francis Bacon and reproduced above, Dewey constructed a hierarchical decimal classification starting with ten main classes:

0.0 General works
0.1 Philosophy
0.2 Religion
0.3 Social sciences
0.4 Philology
0.5 Natural sciences
0.6 Useful arts
0.7 Fine arts
0.8 Literature
0.9 History

The scheme became extremely popular and widely used with the result that fundamental changes to the structure had to be avoided, while creating space at lower levels for the explosion of classes to accommodate such subjects as "plastics" and "computers" in their modern contemporary sense. The custodians over the years, now the Online Computer Library Center (OCLC), must be congratulated on their ingenuity as can be seen from the few amendments to the ten main classes made in the current 23rd edition: 0.0 has become computer science, information and general works, 0.1 has added psychology, 0.4 has updated to language, 0.5 has modernised to plain science and 0.6 to technology, 0.7 to arts and recreation and geography has been added to 0.9. A simple but important addition to the scheme was the Relativ [sic] Index, which guided readers and cataloguers from alphabetical subject entries in the catalogue to the numerical notations given to the classes. Then, in Belgium at the very end of the nineteenth century two lawyers Paul Otlet and Henri La Fontaine founded the International Institute of Bibliography (IIB), originally financed by the Belgian government, with the aim of collecting published bibliographies from around the world and to build up an extensive classified catalogue of their contents. This bibliography grew to an enormous twelve million entries at which stage the IIB changed its name to the International Institute of Documentation and later to the *Fédération Internationale de Documentation*, finally settling in The Hague and housed by the National Library of the Netherlands. Before this final move it was realised that a new classification scheme would be required, one which would be more detailed, complex and flexible than any that existed, particularly for the sciences, as may be deduced from the somewhat archaic "useful arts" of the Dewey scheme, though this scheme was used as a starting point for the first edition of the Universal Decimal Classification (UDC) published in French in 1905. Work continued on the scheme attracting international attention with editions being published in German and English in the 1930s following the second French edition. As with the *DDC*, the UDC is a hierarchical scheme based on a first division of the field of knowledge by discipline. An important innovation, prescient of faceting was the introduction of common auxiliary schedules detailing such aspects as time or place which could be used in conjunction with the main classes. It was now realised that international collaboration was required in the maintenance and revision of the scheme and so a number of revision committees were set up consisting mainly of members working in the growing number of special libraries around the world. The UDC became increasingly influential and widely used with, for example, the English translation as a British Standard and the adoption of the scheme by VINITI, the abstracting service of the Russian Academy of Sciences covering the world's scientific literature. Eventually, this management structure became too unwieldy, leading to such problems as the *impasse* in the social sciences class where it was found impossible to reconcile the Western approach of the chairman to the Marxist approach of the vice-chairman from the Soviet bloc. Eventually, the control of the UDC was handed to a group of publishers who formed the UDC Consortium (UDCC) which has re-established the scheme as a leading international classification published in 40 languages and used in 130 countries.

Almost in parallel with these developments in hierarchical classification came new ideas from the U.S.A and India. Henry Bliss, a librarian at the College of the City of New York for 49 years worked tirelessly on his ideas for a new method of bibliographic classification, one that would have general applicability, rather than, for example, the Library of Congress *Classification* that he believed was created for a specific collection. He thought hard about the theoretical foundations of classification, studying other schemes and publishing his thoughts in 1933 in a book titled *Organization of Knowledge in Libraries,* which also contained an outline of the scheme on which he was working. The full tables appeared in four volumes between 1940 and 1953. The book also contained 32 principles summarizing his approach to the theory and practice of classification. It was probably unfortunate for Bliss that the first volume of his classification appeared at the onset of World War II, and the final volume in the euphoria surrounding the birth of the age of the computer. At the same time, in faraway India, a mathematician and librarian named S. R. Ranganathan was closely following the work of Bliss and working on his own classification known as the *Colon Classification*. It was apparently inspired by his discovery of the toy Meccano set and the belief that subject classes should be built from the bottom up. As Hjørland (2008, 91) has pointed out, quoting Ranganathan:

1. That enumerative classifications have a superficial foundation.
2. That the discovery of new knowledge cannot be anticipated in an enumerative system; and
3. That the discovery of new knowledge can be anticipated in a faceted system (based on the view that new knowledge is formed by combination of *a priori* existing categories).

Ranganathan published his *Colon Classification* in 1933 and his *33 Canons* in various publications, but like Bliss, his scheme has not been widely applied in practice, partly because it did not lend itself easily to the practice of shelf arrangement of books. However, both schemes and their principles had an enormous effect on the thinking and work of later classificationists, particularly initially in Great Britain. The revolutionary difference between the hierarchical *DDC* and UDC, and the new schemes of Bliss and Ranganathan was that while the former two regarded knowledge as an integral whole the latter worked from the bottom up identifying individual concepts and clustering them logically within fundamental categories. Ranganathan initially proposed these fundamental categories in his PMEST formula where the letters stood for personality, matter, energy, space and time. Later thinkers, while finding the concept useful, considered that other more detailed fundamental categories were equally possible, being particularly critical of the somewhat vague category of Personality as the 'core' of any subject. For example one of several versions of the general scheme initially advanced by Vickery and others is taken from Aitchison, Gilchrist, and Bawden (2000, 65-6):

Entities/things/objects
(By **characteristics**)
Abstract entities
Naturally occurring entities

Living entities, Organisms
Artefacts (man-made)
Attributes: properties/qualities, states/conditions
Materials/substances, constituent substances
Parts/components
Whole entities/Complex entities
(*By* **function**)
Agents (Performers of actions–inanimate and animate)
Individuals, personnel, organizations
Equipment/apparatus
Patients (Recipients of actions–inanimate and animate)
End-products
Actions/activities
Processes/functions (internal processes, intransitive actions)
Operations (external, transitive actions)
Space/place/location/environment
Time

The excitement generated by these advances in Great Britain led to the establishment of the Classification Research Group (CRG) in 1952, which was further strengthened by a visit from Ranganathan. In 1952 the CRG issued a memorandum to the Library Research Committee of the U.K. Library Association with the confident title *The need for a Faceted Classification as the basis of all methods of information retrieval*, and subsequently were instrumental in organizing the International Conference on Classification, later known simply as the Dorking Conference where it was held. This was a seminal meeting nicely summarized by R. A. Fairthorne in a celebration of the conference edited by Gilchrist (1997):

> Dorking was necessary, if only to show that devices by themselves were not enough to achieve inverse communication. It was successful because it set in motion the understanding that classification is an essential ingredient of information activity.

Then the Bliss Classification Association was formed in 1967 with the objective of updating the old edition of the scheme. With some fundamental changes the first volume of a second edition was published in 1977 with some financial support from NATO. Work on the other volumes has continued, but the energy slowly dissipated and the work is slow as the principal authors retired and wider interest was submerged in the advance of computerized information retrieval. However, the Bliss Classification Association is still in existence and Vanda Broughton is still working on the scheme.

Before closing this section it is worth pointing out that OCLC has introduced a significant amount of faceting in recent editions of the *DDC*.

5.0 Knowledge organization and computer-based information retrieval

While computers were being developed there was quieter work in the background applying the use of punched cards for document indexing and retrieval. In the late 1940s, Calvin Mooers designed a method of coding subject terms for maximum use of the space on machine-sortable edge-notched cards, while Mortimer Taube (1953, 5-6) used the term co-ordinate indexing to describe the use of his uniterms (unit terms), again with punched cards as: "The analysis of any field of information into a set of terms and the combination of these terms in any order to achieve any desired degree of detail in either indexing or selection." In the UK, W. E. Batten of the Imperial Chemical Industries Company produced the "feature card," originally known as the "peek-a-boo card," a manual device which inverted the approach of the edge-notched card by having the card represent a subject and drilling a hole at a numbered position on the card, which later came to hold 10,000 positions. This was followed in the U.S.A by the machine-sortable 80-column Hollerith card.

There was increasing excitement with these developments, such that Calvin Mooers first coined the term information retrieval. This was deemed by R. A. Fairthorne in England to be exaggerated, he preferring the term reference retrieval. However, Mooers is also credited with the realisation of the concept of the concept. Initially, the uniterms proposed by Taube were taken from the document itself, similarly to the rotation of title words in a KWIC (key word in context) index. This threw up certain problems concerned with meaning such as those compound terms that had a unity, such as "artificial respiration" (which was followed later by the realisation that even the two words in combination were inaccurate and that a term such as "induced ventilation" might be preferred.

These thoughts led to better understanding of the semantic problems, leading to the idea of semantic factoring, the analysis of words into their semantic primitives; for example that "father" might be constructed as "male + parent," a device used by many indexers seeking to limit vocabulary size. These early attempts at getting to grips with semantics laid the foundations for what was to become known as the thesaurus (a term borrowed loosely from the work of Roget, and attributed to Helen Brownson of the U.S. National Science Foundation).

With the recent application of punched card methods there followed a debate regarding the optimal potential sizes of a thesaurus, some arguing for limited vocabularies, others relying on the capability of computer processing, though early applications in this respect required the construction of complicated Boolean equations to be run overnight. As expected, the computer adherents won and large thesauri started to appear, one of the first being one produced by the American Institute of Chemical Engineers in 1960. This was followed by the *Thesaurus of Engineering and Scientific Terms (TEST)* compiled by a committee of some dozen subject experts on behalf of the Engineering Joint Council, which had well over 10,000 terms and later merged with the chemical thesaurus noted above.

Some organizations were now inspired to update their methods, notably the American National Institute of Medicine that had been using subject headings since 1879 in its *Index Medicus*. They initiated a computerization project to convert this into MEDLARS (medical literature analysis and retrieval system) completed in 1964 and involving the transformation of the subject headings into a thesaurus format called *Medical Subject Headings* (*MeSH*). By 2014 *MeSH* had 27,149 descriptors, MEDLARS had gone online as MEDLINE, and in 1997 was made available free to the public under the name PUBMED. Thesauri became increasingly popular in many countries and variations in design and display started to appear.

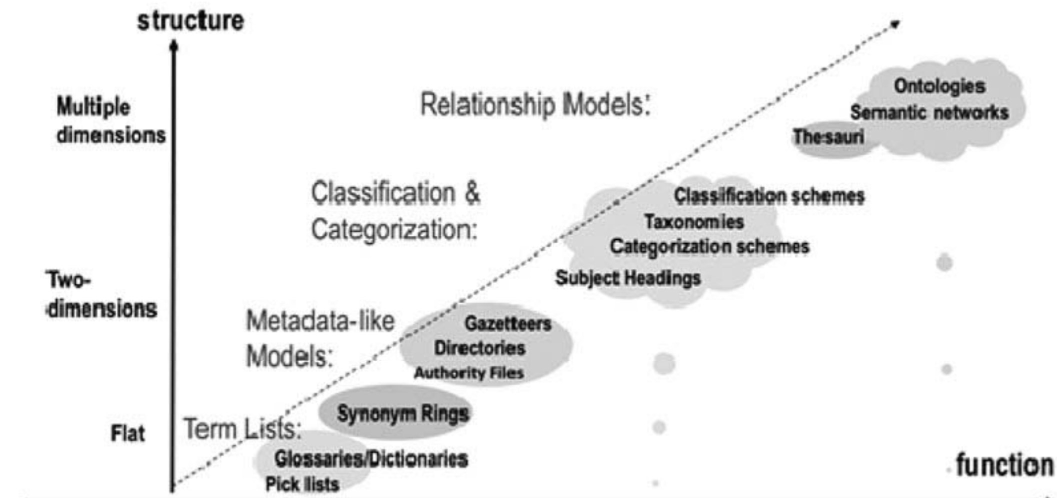
The Directorate General of the European Commission issued a directory of thesauri containing some 4,000 entries, including some that the Commission had produced; one being the EURATOM Thesaurus supporting the European Union's activities in the field of nuclear energy. This thesaurus used "arrowgraphs" to display terms in a set of diagrams somewhat similar to the concept maps used in expert systems and later still in ontologies. In 1969, Jean Aitchison and her team produced the *Thesaurofacet,* a new approach for which she was awarded the Ranganathan Medal. This scheme combined a faceted classification with an alphabetical display in thesaurus format, the entries in both having a 1:1 correspondence. This work was commissioned by the English Electric Company Ltd., but there were few organizations willing to finance such detailed work. A notable exception was UNESCO, although even here the second edition produced in-house abandoned the original detailed structure.

Extending the usefulness of thesauri, a number were produced in more than one language, introducing problems of inter-social meaning; for example a country that had no word for "strike" had to consider the alternative construct "withdrawal of labour in complaint against working conditions." Nevertheless, many international organizations have succeeded in producing multilingual thesauri in ten or more languages.

While all this extensive work was going on and the results being applied to online searching, information retrieval software was also becoming more sophisticated. Online searching was more widely available, using an inverted file to combine index terms in Boolean searches and tricks were introduced to support such things as phrase searching and word proximity. When distributed processing revolutionized the workplace, users were equipped with their own terminals and had direct access to databases. Information specialists now became instructors, undertaking searches as a back-up service. About this time and, as late as 1977, the founder of the Digital Equipment Corporation is alleged to have pronounced that (author's personal notes): "There is no reason why anyone would want a computer in their home." Users were offered in a trial the choice of "simple search," involving entry of terms in a single box as is now common with Google, or "expert search" where different metadata: subject, author, date range of publication etc., could be entered. The trial showed that users preferred simple search, sometimes in the ratio of more than 10:1.

Later came the World Wide Web and Google and many information departments were either downsized or closed altogether; but as is often the case old and new systems continued to co-exist. Attention was now concentrated on the end-user and in providing easy access and search. Google was the prime example, at least for the World Wide Web; and individual websites provided menus supported by simple vocabularies called "taxonomies," a reduced hybrid of classification and thesaurus. Users were also invited to supply their own index terms, initially called "folksonomies," later settling down as less formal "tags." Some organizations used the resulting folksonomies as sources for enriching their own schemes used both internally and on their websites. Another innovation in KO in this period was the pioneering work of Gene Garfield with his introduction of citation indexing leading to the *Web of Science* and the study of scientometrics.

Another big advance in KO came from the field of knowledge engineering, a term that largely replaced the more ambitious artificial intelligence though this earlier term is now becoming again more widely used. This advance is the ontology and, once again, a very old word is borrowed and slightly revised. Originally, the word ontology, in its philosophical and metaphysical sense means "the nature of being," but now can be used to mean simply a conceptualization. There is a further distinction to be made in that the knowledge engineer will use the ontology in conjunction with a scheme of symbolic logic allowing mathematical manipulation while others will regard it in a wider database context, to quote Frické (2012, 27), as: "a description of the types or kinds of entities, and the properties or attributes, that are assumed to exist for the purposes of the database." Horrocks (2012) in an



eliminating ambiguity	xxx		xxx	xx	xxxx	xx
controlling synonyms		xxxx	xxx	xx	xxxx	xx
establishing relationships: hierarchical			x	xxxx	xxx	xxx
establishing relationships: associative					xxxx	xxxxx
presenting properties						xxxxx

Figure 4. An overview of structures and functions of Knowledge Organization Systems (Zeng 2008, 161)

ISKO-UK event in which he described various aspects of ontology creation and application defined an ontology simply as "A model (of some aspect) of the world," followed by a slide with the simple statement "Developing and using ontologies is hard." Any of the definitions above places the ontology in a graph as suggested by Zeng (2008) in Figure 4.

The reason why creating ontologies is hard is that each ontology must create (or borrow) a vocabulary, establish complex relationships between the concepts (beyond the simple relationships found in most thesauri), and support these with rules and axioms. Furthermore, the ontology must be described in one of the various logic languages available. Interoperability between semantic languages and between logic languages used in different ontologies can be difficult, though standards and interoperability tools are being developed. For this and other reasons, most ontologies are confined to specific, usually relatively small, domains of knowledge.

The knowledge organization systems (KOSs) in Zeng's diagram not only show a range of schemes, but a collection of techniques that can be used singly or in combination, and with different packages of software. Knowledge engineers are well aware of this and are ready to use any or all of the KOSs that have preceded their own endeavours. Sheth et al. (2005) have identified the role of semantics in several disciplines: information retrieval (IR), information extraction (IE), computational linguistics (CL), knowledge representation (KR), artificial intelligence (AI) and database management (DB). They go on to group these approaches into those that primarily draw upon unstructured texts (IR, IE and CL) in which (13) "the semantics are implicit;" those with deeper analysis where, for example syntactic structures are defined (KR, AI and DB) which they call "formal semantics." The authors go on to critique the various logics used, such as description logic and first order logic, which they deem to be inadequate to meet the demands of the semantic web. They are also uneasy about the ability of what they called formal semantics to meet increasingly complex demands to represent knowledge that is, in their words "imprecise, uncertain, partially true, and approximate." Referring to earlier research they call for a new approach combining the best of "implicit semantics" and "formal semantics" supported by appropriate logics. It is interesting to note that the authors return to the work of Zadeh who proposed "fuzzy set theory" and hence "fuzzy logic" as long ago as 1965 and applied by software engineers designing Bayesian retrieval systems.

Of particular interest to KO is linked data, the technique by which web pages of any purpose or content can be linked to provide browsing and searching irrespective of the schema initially used on each site. The entities on each site are allocated a URI (universal resource identifier) forming part of an RDF (Resource Description Framework) "triple" representing subject–predicate–object (for example scientific article, with its URI–written by (a defined relationship) with its URI–author, with its URI) . Because each website will use slightly different URIs with which to label things, there are facilities such as linked data search engines, (for example Falcons and SWSE) that can provide information concerning a specific URI and the way in which it is used and the context within which it is used. The arachnological analogy of a web is accurate as multitudes of linked data workers spin the links between resources. Note that the subject vocabulary for each resource, whether formal or uncontrolled is still independent. Formal vocabulary building and interoperability are still well within the province of traditional KO.

6.0 Models and systems

Throughout this paper the common words model and system have been used and they are basic to the work of KO. Models are created to gain a better understanding of complexity, to provide a means of manipulating that understanding and possibly to apply it to its particular environment. There are many different types of model employing different spatial forms. Some have a matrix format such as the previously mentioned Mendeleev periodic table of elements, which had predictive properties. The so-called Standard Model for particle physics also has a matrix format and, enumerating all the bosons, fermions and their component quarks, leptons etc., predicted the Higgs boson some years before it was identified in the Large Hadron Collider. Many models, from Aristotle onwards, are hierarchical and this includes the Darwinian evolutionary trees, which, as a common principle in science use what is called parsimony to choose the simplest scientific explanation. Similarly, KO moving on from lists created hierarchical classifications to reflect a logical order for the arrangement of books graduating to the flexibility afforded by faceted classification and so on to the more granular structuring of vocabularies explicitly hierarchical in arrangement but providing more cross references than traditional classification by use of associative relations. All of these are models of reality, and this includes the ontology, which, though incorporating hierarchy and association, is actually a network, a term that has become common with advances in scientific understanding and the power of information technology to map networks. For example, some Darwinian theorists are re-mapping evolutionary trees as networks following research in phylogenetics; and the most obvious example of networks is to be found in cyberspace with web clouds showing connections provided by linked data or the connections created by social media. It should be clearly stated here that networks are not replacing hierarchies; on the contrary, they may consist of hierarchical and associative relationships, though the definitions of these relationships may be complex. The word model has acquired, in some contexts, the meaning of perfection, a paragon to which humans should aspire, but the use of the word above is quite different in that, by their nature they are attempts at reflecting reality and hence are transitory by nature (as is, of course, the paragon example). Models must be tested, if necessary to destruction and in KO this means applying the organization of symbols of knowledge, with supporting rules, within specific environments to create a knowledge organization system (KOS). This may further require the matching of the KOS to organizational models such as in information architecture (itself part of an enterprise architecture) and to other models of, for example, user needs and behaviour. The application of the KOS should be (but often is not) evaluated and updated and amended as may be required. KO without this cyclical refreshment remains theoretical.

7.0 Conclusions

1. KO deals with abstract entities (information, knowledge and language).
2. KO creates and applies models that must be continuously updated.
3. KO applications are wider than traditional cataloguing and classification and computer-based information retrieval.

The history of KO is long and varied, and for centuries people working in what we now call knowledge organization have bravely succeeded in making a difference. Without the discipline and application of countless scribes, copyists, translators, librarians, scholars, bibliographers, lexicographers and encyclopaedists, information scientists and knowledge engineers the world would be a poorer, even barbaric place. Now, KO is alive and well in most areas, particularly in some of the large organizations such as the U.S National Library of Medicine, the Food and Agriculture Organization and the European Parliament with their large, multilingual thesaurus-based systems. There are also exciting advances in linked data projects in many areas. However, the situation in the business sector is not so cheerful as Foster (2014) reports in an account of a report from an industry analysis agency: "Gartner predicts in a recent study that by 2017, 33% of Fortune's 100 organizations will experience an information crisis due to their inability to effectively value, govern and trust their enterprise information." The report also found that unstructured information was particularly badly dealt with. In our technological age, suffering from communication overload we must not forget that there is still much to do and that we are still capable of making a difference.

References

Aitchison, Jean, Alan Gilchrist and David Bawden. 2000. *Thesaurus Construction and Use: A Practical Manual.* 4th ed. London: Aslib.

Belkin, Nicholas J. and Stephen E. Robertson. 1976. "Information Science and the Phenomenon of Information." *Journal of the American Society for Information Science* 27: 197-204.

Bush, Vannevar. 1945. "As We May Think." *Atlantic Monthly* 176, no. 1: 101-8.

Finkel, Irving. 2014. *The Ark before Noah*. London: Hodder & Stoughton.

Foster, Allan. 2014. "Initiatives." *Business Information Review* 31, no. 2: 122-39. http://www.gartner.com/newsroom/id/2672515

Frické, Martin. 2012. *Logic and the Organization of Information.* New York: Springer.

Gilchrist, Alan. 1997. *From Classification to "Knowledge Organization:" Dorking Revisited or "Past is Prelude."* FID Occasional Paper No. 14. The Hague: International Federation for Information and Documentation.

Hjørland, Birger. 2008. "What is Knowledge Organization (KO)?" *Knowledge Organization* 35: 86- 101.

Horrocks, Ian. 2012. Logical foundations for a Semantic Web. Unpublished presentation: http://www.iskouk.org/content/logical-foundations-semantic-web

Ingwersen, Peter. 1996. "Cognitive Perspectives of Information Retrieval Interaction: Elements of a Cognitive IR Theory." *Journal of Documentation* 52: 3-50.

Liddy, Elizabeth D. 1998. "Enhanced text retrieval using Natural Language Processing." *Bulletin of the American Society for Information Science and Technology* 24, no. 4: 14-6.

McArthur, Tom. 1986. *Worlds of Reference*. Cambridge: Cambridge University Press.

Myers, Gail E. and Michele Tolela Myers. 1958. *The Dynamics of Human Communication*. New York: McGraw Hill.

Nonaka, Ikujiro and Hirotaka Takeuchi. 1995. *The Knowledge Creating Company*. Oxford: Oxford University Press.

Remenyi, Dan. 2003. *How to Become a Successful IT Consultant*. Burlington: Butterworth-Heinemann.

Roszak, Theodore. 1986. *The Cult of Information*. New York: Pantheon Book.

Sheth, Amit, Cartic Ramakrishnan and Christopher Thomas. 2005. "Semantics for the Semantic Web: The Implicit, the Formal and the Powerful." *International Journal on Semantic Web and Information Systems* 1: 1-18.

Taube, Mortimer. 1953. *Coordinate Indexing*. Studies in coordinate indexing 2. Washington: Documentation Incorporated.

Zeng, Marcia Lei. 2008. "Knowledge Organization Systems (KOS)." *Knowledge Organization* 35: 160-82.