Knowledge Management and Information Management - back to their roots

Gabriel Anderbjörk
Stockholm, Sweden
For contact details, please see
http://gabriel.anderbjork.se

Published October 2012

Keywords: Knowledge Management, Information Management, Decision Support, Organization Behavior, Meta Data, Taxonomies.

***Abstract**: This paper addresses one of the current billion dollar challenges of large organizations – i.e. how to make optimal use of all information at hand. The bonanza of academic definitions and the never ending stream of "new" labels on old capabilities, typically offered by large consulting firms, do not help. It is necessary to get back to the roots of the needs and build from there. This paper suggests a definition triad; information - knowledge - algorithm and a usage dualism; decisions and enrichment as the only necessary terms to relate to for the theory building. Further, these five terms are perfectly sufficient in order to build complete models for both organizational architectures and computer aided tools for information management and all derivatives of that label.*

If all our design engineers knew what they ought to know, and if what they *do* know is accurate, we would reduce design and development costs by Euro 200 million per year.
CTO of major Northern European manufacturing company

Definitions, definitions…

Academia and others have for many years struggled with their definitions of various terms in the fields of information management and knowledge management. There is tacit knowledge and explicit knowledge. There is data, information and knowledge on a sort of value scale of... yes, what - …information?

One model suggests that "data" equals facts and thus can't be wrong whereas "information" (captured data and knowledge) can.

Harry Scarborough et al (1999, p.1) define Knowledge Management (KM) as "any process or practice of creating, acquiring, capturing, sharing and using knowledge, wherever it resides, to enhance learning and performance in organizations"

Hedlund (1994) suggests that KM addresses the generation, representation, storage, transfer, transformation, application, embedding, and protecting of organizational knowledge.

Davenport criticized technological approaches to KM in 1996. "The emphasis on codification in the KM literature probably reflects the dominance of the information systems view: many of the articles have focused on developing and implementing KM databases, tools (e.g. decision support tools) and techniques despite a now fairly wide acknowledgement that "the most dramatic improvements in KM capability in the next ten years will be human and managerial"

A popular definition is also that KM is a conscious strategy of getting

- the right knowledge,
- to the right people,
- at the right time.

The need

The purpose of this paper is not to cast a general opinion regarding the viability of any of the definitions exemplified above, nor any other definitions not referred to herein, but only to put them in perspective of a much defined need.

Whatever "it" may be called; data, information, knowledge or any other term, it is no exaggeration to claim that "it" by now is abundant in any context of business or society. It is so abundant that without the support of various kinds of software tools it is virtually impossible for any given individual to make use of whatever there is to make use of.

Hence, a set of definitions that can help bridging the complexity of the topic to the need for systems and architecture design is needed. It is unlikely that the argument that the above exemplified definitions does not do a good job, in that perspective, need to be further elaborated. Also, it can be argued that this is one reason for why it is still a rare treat to find a working "knowledge management system" in today's corporations.

Labels of software that, from different angles, claim an excellence in this filed today include "Enterprise Information Access", "Enterprise Search", Social Computing", "Content Management", "Collaboration Software" "Wickis" and of course "Knowledge Management Systems". Although each suite introduced to the market, under any label, very well may offer potent functionality to support the real need, the risk is to get lost in a situation resembling the tale of the Emperor's New Clothes when trying to define a systems need based on the labels available.

The bottom line need can, however, quite easily be defined by theoretically isolating one individual in a given context of business or society and ask a simple question; how to enable this individual to, by social interaction or by individual consumption of "information", form an opinion or take a decision that is most optimal for the context in which the individual currently finds itself?

It is this need that ought to be the driving force in any information management set-up or information architecture development.

A suggested definition triad

The number of quotes and papers regarding definitions in this field is, as already discussed, in abundance, but a few quotes proved helpful as a starting point. It is noteworthy that all of them are more than a decade old;

"I call my field knowledge management but you cannot manage knowledge, nobody can. What you do, what a company does, is manage the environment that optimizes knowledge." (Prusak 1997)

“Knowledge is a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information.” (Davenport/Prusak, Authors of “Working Knowledge” 1998)

“In a knowledge-driven society there are only two key assets to manage. The first is the information your company possesses. The second is the individuals who, for every minute that passes, increase the value of small pieces of information by adding experience, knowledge and associations in order to produce other, even more valuable pieces of information. This cycle continues endlessly and the companies and managers who succeed in making a given information cycle work to their advantage will be the winners of tomorrow!” (Anderbjörk, Intelligent Communities,1999).

The essence of these quotes provides three very useful statements;

Knowledge can be seen as the processing capability of a human brain (based on formal learning, experiences etc.) to process information and provide insights, analysis and/or synthesis for further usage. Thus, knowledge cannot be stored. However, information about available knowledge (brains, or, in more daily terms, people) can.

If knowledge is defined as above, it is necessary to define a machine (computer) equivalent of the human knowledge. For that purpose the word **Algorithm** will be used.

Thus, by applying knowledge or algorithms to an element of information, another element of information can be crafted. One such information element could for instance be a documented decision.

If the two statements above are accepted, it can be derived that **Information** must mean any communicable element (graphics/video, audio or text) that can be transferred between two or many points of knowledge or algorithms, non-real time (and hence stored for future usage) or real time (may be stored but not as a requirement for the definition).

Hence the definition triad suggested: **information - knowledge - algorithm**.

A direct consequence of these definitions is that there are no longer any "levels" of information. Traditionally a minimum split has been between data, information and knowledge but it is very hard to get anyone to really provide a useful specification of their internal differences, other than that “information is more enriched and valuable than data” etc., but then what? Fact is; one or two digits in a number series can be seen as information (usually referred to as data and therefore traditionally of less value) and an entire PhD thesis can be seen as information (by some referred to as knowledge and hence immensely valuable).

Having stated the above, and in order to enable the reader to embrace this definition of information, it is necessary to also bring in the concept of information elements. An information element must be substantial enough to serve as a meaningful input to a knowledge based, or an algorithm based, process for which it is intended. On the other hand, in order to be optimally manageable it should typically not be much bigger than just that. Hence, a PhD thesis is most likely far too substantial to be considered as one information element. Rather, it is a large number of elements that jointly, in different combination, serves different processing needs. It is likely that the only single point in time at which anyone will seriously consider using an entire thesis is at the dissertation.

How then about the two digit information? Can two digits ever be valuable on their own? Consider a telephone network switch. The area code is usually two to four digits and the only value they provide in the connecting-the-call algorithm is to make sure the call is routed to the correct regional switch. It is a simple operation but literally invaluable for the completion of the process. Hence, the value of the information element is entirely defined by the context in which it is being utilized, not its implicit substance or source. (For a more extensive discussion on information valuation, see "The value of information – the quest for ROI in organization models and software investments", Anderbjörk 2009)

A suggested usage dualism

With the definition triad at hand, the next matter to discuss must be that of the "usage" of information. What is information used for? Theoretically it is convenient to split the usage into three major patterns; decisions, enrichment of the information or enrichment of capabilities (knowledge or algorithms).

Decisions would be when the information is used for triggering or altering an action (e.g. routing the phone call to the correct regional switch or setting the final price in a negotiation).

Enrichment of information is when knowledge or an algorithm is applied to one or many information elements only in order to make these even more useful in a later stage in the process. Such enrichment can be e.g. a Business Intelligence software bringing order and visibility to otherwise unstructured collections of accounting figures or a person crafting a presentation out of certain sets of information elements prior to a board meeting.

Finally, when information is used for **Enrichment of capabilities** we tend to refer to it as education (if enriching knowledge), disregarding distinctions such as whether the information is transferred through real time audio/video, e.g. a lecture, or if the "knowledge element" undertakes its own enrichment by simply reading a textbook. Experiences, i.e. immediate feedback on actions or observations in general, fall into the same category. A direct analogy is when AI systems use new information (feedback loops etc.) to improve its own algorithms. The only information usage pattern that seemingly falls out of this model would be when information is used by people for pure amusement and entertainment. Still, how often do we not refer to such information in dialogues and reference building in other situations in life? It is more than fair to group such usage in to the knowledge enriching category.

Hence the usage dualism suggested: **decisions - enrichment**.

Optimizing the use of information

The major question posed in the beginning of this paper was "how to make optimal use of all information at hand?" Now the only fair question would be; how may the theory outlined above help answer that primary question?

Firstly, it is necessary to deal with the word "optimize". In the theories of processes it is a lemma that one can only optimize a process for one target and so far that lemma has not been successfully challenged so it has to be accepted. Information, however, can contribute value to many different processes or activities, now or in the future, so the issue of optimizing *the use* or *the value contribution* of information is entirely contextual. Hence, the theory need to be deployed in a contextual perspective.

The real key to the answer lies in the fact that there is now only one type of entity called information, with no levels or variants thereof. “All” that is needed is a model to enable knowledge or algorithms to make optimal efficiency in the usage of each information element. Thanks to history such models already exist, but they are sadly underutilized in business. This underutilization, however, very often seem to be directly related to exceedingly complicated and non-transparent models to tackle the definition issues discussed above. Put in very simple terms, companies tend to never come to this point in the development of information management as they get trapped in “the bonanza of academic definition” referred to in the abstract. Hence, let’s leave that behind and move on to a very straight forward, and actually quite simple, model for the management of information; that of meta data and taxonomies.

Meta data and taxonomies

The label "meta data" has a certain unfortunate and irrelevant restriction attached to it. Meta data is usually referred to as facts such as document author, date of creation, document type, potential process relations etc. However, the key to the next generation of information management is to extend the concept of meta data, or "the information about the information", substantially. In fact, all thinkable information management control can be implemented by the use of meta data;

- business/operations usability
- information quality
- security and access requirements
- geo/positioning
- time stamps
- etc.

From a usage perspective (knowledge or algorithm) meta data is represented by multifaceted "maps" of topics, so called taxonomies. The amount of available literature on taxonomies is vast; more than sufficient. Hence, this paper will only briefly illustrate the essence of taxonomy principles to support the content herein. A taxonomy represents or visualizes its described reality through topic structures. As an example, companies almost without exception have business taxonomies with at least the following so called first level topics; companies, products and markets. Under each first level topic a list of further descriptive topics are available. Under markets you will most likely find all markets in which the company operates listed and so forth. A very good rule of thumb is that any topic on any level should, by its name/label, intuitively inform the user of what to expect when viewing the next level. The same line of thinking applies to all dimensions of a taxonomy. First level topics in a security taxonomy would most likely include confidentiality levels, access rights, etc. The critical aspect of a taxonomy, though, is that it should represent the operation it is designed to support. There are no taxonomies "out of the box", although it is quite feasible to start with a standard taxonomy for e.g. an industry and then evolve this into the own organization’s term, jargons and, most important, own understanding of the operations environment in which it is to succeed.

Now, if taxonomies are the tools of choice to both enable and control usage of information, it is of course essential that each and every information element is correctly tagged with topics that properly and sufficiently describes its content, context, quality, security level etc.

With tens of thousands of information elements being added to an average corporate environment on a daily basis, it is not even to consider to apply knowledge for such enrichment. Algorithms are the only possible way forward. By ensuring that each topic is programmable in terms of different search or classification languages, computers can do the tedious job of reading through immense volumes of information element and applying the correct meta data /topics to each element.

Based on the topic maps, users can automate information access (notifications etc.) based on their current needs. In an ideal world, human users (knowledge) trust that algorithms (topic rules) work so well for them that the concept of searching for information is reduced to an odd out activity when scouting for out-of-the-box insights.

Letting the algorithms take over

It is time to close the loop and return to the challenge outlined in the abstract. For a given organization, how to make optimal use of the vast amount of information at hand? To make use of it, organizations have two options only; either apply knowledge or apply algorithms to the information. To apply knowledge is tremendously costly and, in comparison to algorithms, exceedingly slow. Knowledge should thus be applied only when there are no algorithms available to do the job. As a consequence, any knowledge that can be transferred into an algorithm should be exactly that. Yet, a frighteningly large majority of big organizations still find their employees spending hours and hours every week on one of the simplest information tasks to make algorithms for, namely finding what they need. That is certainly not to make optimal use of available information.

Conclusion

For quite a few years now, words such as codifying and taxonomies have been harshly bashed by many scholars within the fields of Knowledge Management. This paper agrees when it comes to the actual deployment of *knowledge* but when it comes to the management of the indescribably vast amount of information, needed to support knowledge, some kind of structuring tools are imperative. So far nothing better than meta data and taxonomies have been presented, it is just a question of how they are deployed.

So, is this to say that all knowledge should be replaced by algorithms? Not at all! However, a very sensible strategy for information operations is to direct as much knowledge capacity as possible to the task of algorithm enrichment in order to continuously free up more knowledge time to the information usage we still have to believe is only science fiction for algorithms - actually understanding the complex environment in which the individual or the organization is operating and in the light of "perfect information" from algorithms enable individuals and organizations to do the "right things".

A CIO of a major Australian bank was asked ”Why do you happily pour millions of dollars into the management of structured information and only fractions into management of unstructured information (that after all accounts for 70-90% of all your information)?” The answer has a bearing on the content of this paper; ”Because, with regards to structured information we ***believe*** we know what we are doing….”