Management for Professionals

Paul Baan *Editor*

Enterprise Information Management

When Information Becomes Inspiration

Springer

Management for Professionals

For further volumes:
http://www.springer.com/series/10101

Paul Baan
Editor

Enterprise Information Management

When Information Becomes Inspiration

Springer

Editor
Paul Baan
Incentro
De Meern, The Netherlands

ISSN 2192-8096 ISSN 2192-810X(electronic)
ISBN 978-1-4614-5235-5 ISBN 978-1-4614-5236-2 (eBook)
DOI 10.1007/978-1-4614-5236-2
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012951430

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Recently I attended a CIO gathering in Amsterdam, the Netherlands. Several remarks made by different keynote speakers caught my attention. First of all, most of the speakers referred to an ongoing discussion on the future of the CIO position. Some said there will always be a need for a CIO, others stated that the CIO is in danger of becoming extinct because of a diminishing importance to the business.

Without further exploring this existential question, the conference went on. Server availability, the cloud, dual core mobile devices and other technobabble took the upper hand. Online applications such as Yammer and DropBox were discussed, the standard security questions floated up.

> How will CIO's bring value to the I in CIO when all they focus on is the T in IT?

I was shocked. Somehow these people must feel an urge to prove their added value, following the discussion on the future of the CIO position. At the same time, they are acting as good old CTO's as if nothing has changed. How will you ever bring evidence for your added value if you copy the operational responsibility of another CxO? And maybe even more important, how will you bring value to the I in CIO when all you focus on is the T in IT?

The decision whether or not to host a business application internally or externally (cloud) has – as far as I know – never made any information worker smarter in his operations. Of course unless information then suddenly becomes available to this information worker.

The intelligence of an organization is determined by the sum of all the clever minds working in it. This is especially true for knowledge-driven economies such as the western economies. In order for an organization to operate smarter, it has to focus on making these minds effective, preferably in an efficient way.

When we look at recent developments in for instance neuroscience, we see that the IT industry so far has been giving all kinds of tools to users. But these tools are not necessarily well designed for the users of these tools. IT has been giving us hammers with two steels (more force!), screwdrivers the size of broom sticks (reaches every screw!) and a saw that is completely safe to use (and doesn't cut through wood either).

What I am trying to say is that we are learning that in order to make human minds smarter, we need to consider IT no more than a basic hygiene factor. And focus on how to offer information so that it is optimally digested. Any tool used should focus on the needs of the (mind of) the information worker. These tools should acknowledge the fact that the human mind is not a rational decision maker per se. That it is biased and filters out information that does not confluence with our own vision. That it is easily influenced by earlier information provided, or by specific wording. That human minds have a tendency towards overconfidence in their own capabilities.

IT has a history of approaching users as rational actors as if they were interfacing on a logical level only. That's why you will not hear any traditional IT service or solution provider talk about the impact of neuroscience. They have yet to find an answer to the growing importance of our understanding of the human mind.

Other important trends are more easily interpreted by traditional IT. Social media, new generations of information workers and growing pressure on transparency are trends that impact the way we make information productive. But they are manageable from within the old dogma's.

This book is not for CIO's who are in fact CTO's. This book is also not for people who think that the way we are currently operating in terms of information productivity is *well enough*. Of course, it is. But it can be further optimized. And as far as we can tell, that is the single most important responsibility of today's CIO.

At that same conference in Amsterdam, one of the keynote speakers mentioned that European CIO's are closer aligned to their businesses than US CIO's. I would not dare to make such a generalization. I feel that any CIO that is open to learn from trends that are impacting information productivity, can be a successful CIO, whether in the US or in Europe. This book is aimed to do just that. To keep you up to date with recent findings and the most up to date beliefs on how to get that expensive production factor *information* work for you and your organization.

I wish you pleasant reading and promise you that we at Incentro will be on the lookout for new developments in this area. And when there's news, we'll let you know. As sharing information is one of the most basic ways to get information productive.

De Meern, The Netherlands Paul Baan

Reading Guide

This book aims for readers to:

- Get an insight into what enterprise information management (EIM) is;
- Get an idea of the value of EIM for increasing the return on information;
- Be able to work with an approach to gradually increase the return on information;
- Be able to place and apply the necessary technologies in the framework of enterprise information. There must be a possibility to integrate the technologies from BPM, business intelligence, content management and enterprise search.
- Be able to translate EIM to their own situation.

This book is exploring the theories, features and characteristics of all the elements of the fields concerned with information management that, combined, led to the new phenomenon of enterprise information management. Both business processes, business intelligence and enterprise content management are strongly involved in this. Being able to retrieve stored information by using advanced search technologies (enterprise search and retrieval) offers new opportunities in the form of EIM.

Chapter 1 is in introduction to enterprise information management from a management (top-down) perspective. Increasing the return of information is central in this. We discuss what role EIM can play in this and what approach may be used to increase the return on investment. The EIM maturity model is also discussed in this chapter.

Chapter 2 focuses on the concept of enterprise information management and illustrates this with models and examples. It is an integrated interplay of several fields of study, including change management.

Chapter 3, 4, 5, 6 and 7 give content to the substantive model for EIM and the individual components from a technical perspective; enterprise information management (EIM;3), business process management (BPM; 4), enterprise content management (ECM; 5), business intelligence (BI; 6) and enterprise search and retrieval (ESR; 7).

Each chapter starts with a summary and a first paragraph containing an exploration of the field under the umbrella term Concept, followed by a paragraph about Vision. After that, a description of the content of the field of study is giving based on the "EIM information triangle".

All chapters then elaborate on the hurdles in the information process. Accessibility, availability, relevance and interpretability in relation to information, seen from the perspectives of the information worker, process and technique.

Acknowledgments

In 2010, at EIM2010 the first edition of this book was presented to Stef Lagomatis, founder of VLC, now Incentro. Little did we know that the book would be so well received. Over time, we learned that this book was even used at Dutch colleges to enlighten students on the different aspects of good information management.

When we were given the opportunity to publish our book via Springer, one of the worlds' leading scientific publishers, we didn't have to think twice. But instead of just a translated version, we wanted the book to be updated with the latest insights we had discovered in our consulting business. We looked at each other, and unanimously agreed this was too big an opportunity to let go. We would again undergo the experience of writing a book next to our day-jobs. And we knew what it meant.

We are all very grateful to Incentro, our employer for fully supporting us in this decision. They have given us the opportunity to get this book published. We are also very proud of the team effort. Anja van der Lans, Robbert Homburg, Peter van Til, John Septer, Paul Baan, Aukje Wielens, Maaike Bloem, Sander Verstoep and Adrian Munro all worked hard to get all our thinking in comprehensive writing, to share the latest insights with you, the reader. Special thanks go out to Anja, who has relentlessly been driving the team forward into delivering this book.

We're hoping we'll be able to inspire you to a multi-angular view on information management and information productivity. And when we did, we're hoping you'll let us know your new-gained insights.
eimbook@incentro.com

Contents

Chapter 1
Information Productivity: An Introduction to Enterprise Information Management

Paul Baan and Robbert Homburg

At first, it may seem like a rather straightforward question. How productive is the information within your *organization*? But over the years we learned that the answer to this question is far from obvious. The problems that arise when trying to answer this question are illustrative of the state of the current IT industry. IT is supposed to stand for Information Technology, but in fact is much more about technology then it is about information.

> In order to understand the impact of this industry-wide myopia, we need to understand the enormous potential of information that is productive in optima forma.

1.1 Historical Background

In economics, factors of production are used to describe the input of processes to create value. In its most classical sense, factors of production are land, labor and something that was called "capital stock", being human-made means of production used for producing other goods.

This definition of factors of production is most accurate when seen in the light of the historical context. Adam Smith, one of defining characters for the above-mentioned factors of production, lived in the eighteenth century. His best-known publication, The Wealth of Nations[1] was first published on 9 March 1776, during the British Agricultural Revolution. This revolution[2] describes a period of development

[1] http://en.wikipedia.org/wiki/The_Wealth_of_Nations

[2] http://en.wikipedia.org/wiki/British_Agricultural_Revolution

P. Baan (✉)
Incentro, De Meern, The Netherlands
e-mail: paul.baan@incentro.com

R. Homburg
Advisory Services, Incentro, De Meern, The Netherlands
e-mail: robbert.homburg@incentro.com

P. Baan (ed.), *Enterprise Information Management: When Information Becomes Inspiration*, Management for Professionals 2, DOI 10.1007/978-1-4614-5236-2_1, © Incentro 2013

in Britain between the seventeenth century and the end of the nineteenth century, which saw an epoch-making increase in agricultural productivity and net output. This in turn supported unprecedented population growth, freeing up a significant percentage of the workforce, and thereby helped drive the Industrial Revolution.

Starting in the later part of the eighteenth century, there began a transition in parts of Great Britain's previously manual labor and draft-animal–based economy towards machine-based manufacturing. It started with the mechanization of the textile industries, the development of iron-making techniques and the increased use of refined coal. Trade expansion was enabled by the introduction of canals, improved roads and railways. With the transition away from an agricultural-based economy and towards machine-based manufacturing came a great influx of population from the countryside and into the towns and cities, which swelled in population.[3]

The Industrial Revolution demonstrated what an active approach towards optimizing production factors could deliver. For many decades to follow, the world would benefit from the increase in wealth. At the same time, the insular focus on mechanization and increased efficiency caused other resources to suffer heavily. The human workforce was seen as supportive to the industrial machines that at that time were the most important production factors. Work conditions were very poor and unemployment rose.

Since the Industrial Revolution there has been a constant strive to develop machines that could automate actions that would previously be dominated by the human body and mind. Human capabilities by far outperformed what machines could do. At first, machines were invented to perform very simple actions more efficiently. Later, the actions that could be performed increased in complexity.

In 1997, Kasparov, at that time reigning World Chess Champion for 15 years, lost a chess tournament to a computer, named Deep Blue.[4] It was a second attempt by the Beep Blue team, a year earlier Kasparov had won. There's a lot of controversy on this game, and later games between computers and other grand masters leave room for interpretation on whether or not chess machines truly have outsmarted human beings. But for sure this demonstrates that where machines in the early days could only perform mind-numbing repetitive work that merely replaced human muscle, since the 1970s machines gradually became more sophisticated and started to replace not only muscle, but – in some places – also the human mind.

Alongside this rat race between humans en machines was a new kind of revolution, the Information Revolution. Producing goods increasingly had become the domain of machines and lower wage countries, whereas service delivery, inventing new products and services and developing knowledge on producing and servicing had become the domain of western economies.

[3] http://en.wikipedia.org/wiki/Industrial_Revolution

[4] http://en.wikipedia.org/wiki/Game_Over:_Kasparov_and_the_Machine

In the United States for example, from January 1972 to August 2010, the number of people employed in manufacturing jobs fell from 17,500,000 to 11,500,000 while manufacturing value rose 270%.[5] It initially appeared that job loss in the industrial sector might be partially offset by the rapid growth of jobs in the IT sector. However after the recession of March 2001, the number of jobs in the IT sector dropped sharply and continued to drop until 2003. Even the IT sector is not immune to this problem. This is easily explained by some key elements of how we define the IT industry. The core of the IT industry is still build around a large base of programmers. Programming capacity has been moved to low wage countries over the past decades, and more software is becoming configurable, thus making it adjustable by business users rather than programmers. Western economies should be the driving force behind this development, and therefore always need to be frontrunners in developing new ways of using IT and thinking on its' uses. Enterprise Information Management is such a development. It combines traditional IT with neuroscience and usability, all from a business perspective.

As industry is becoming more information-intensive and less labor- and capital-intensive, this trend also has important implications for the workforce; workers are becoming increasingly productive as the value of their labor decreases.[6] However, there are also important implications for capitalism itself; not only is the value of labor decreased, the value of capital is also diminished. In the classical model, investments in human capital and financial capital are important predictors of the performance of a new venture. However, as demonstrated by Mark Zuckerberg (founder of Facebook), it now seems possible for a group of relatively inexperienced people with limited capital to succeed on a large scale.

All the more reason to consider information a key production factor, especially in western economies. And in order to perform and compete successfully in such economies, information productivity should be a board room responsibility, preferably covered by the CIO as liaison between business needs and IT capabilities. The Dutch CIO 2011 (as elected by a professional jury of CIO Magazine) Aloys Kregting of DSM (a global science-based company known for its innovative character) states this liaison role as his ongoing key ambition. "I'll keep trying to explain that IT is not about the technology itself, it is about what you can do with it", as he states in CIO Magazine (issue 6, 2011).

1.1.1 Information

For a proper analysis of information as a production factor we should agree on a definition of information. This is where it gets tricky. Information is one of those terms that seem to lack a widely-shared definition, yet anyone seems to have a keen

[5] http://en.wikipedia.org/wiki/Information_Age

[6] http://en.wikipedia.org/wiki/Information_Age

understanding of what is meant by the term. From an information and knowledge management standpoint, this is a dangerous starting point. For all we know, it might be that the lack of a clear and common definition leaves room for interpretability and therefore noise in the communication.

> We are aiming at having information change the behavior of information workers

For the purpose of this book, we will lend the definition as found on Wikipedia[7]: *"Information in its most restricted technical sense is a message (utterance or expression) or collection of messages in an ordered sequence that consists of symbols, or it is the meaning that can be interpreted from such a message or collection of messages. Information can be recorded or transmitted. It can be recorded as signs, or conveyed as signals. Information is any kind of event that affects the state of a dynamic system. The concept has numerous other meanings in different contexts. Moreover, the concept of information is closely related to notions of constraint, communication, control, data, form, instruction, knowledge, meaning, mental stimulus, pattern, perception, representation, and especially entropy."*

The same Wikipedia page also states some contexts in which the term information can have a different meaning. Information can be used to describe sensory input, as a representation and complexity (aiming to overcome the definition of Shannon-Weaver when dealing with subjective information), as an influence which leads to a transformation, as a property in physics, as technologically mediated information and as records.

Especially the context as an influence which leads to a transformation is of interest of this book. As we will state later, the productivity of information is determined by the behavior of information workers. Information should therefore be used to make these information workers effective or efficient (depending on their context, see later). Thus, we are aiming at having information change the behavior of information workers.

1.1.2 Value of Information

It should not come as a surprise that the authors of this book consider information to be a valuable asset. However, there is one important aspect that needs further clarification. In order for information to be valuable, it needs to be alive. Or, as Peter Hinssen calls it in his book "The New Normal" (Hinssen 2010), information should be liquid. It shouldn't be stored somewhere deep down, it should be flowing freely within the organization, from one information worker to the other, while there adding to it, adjusting it and building on it.

> The better the information is stored away in IT systems, the least likely this information is to be flowing through the minds of information workers.

Traditional knowledge management was very much focused on capturing knowledge and storing it. These two activities (capture and store) can also be found in many models for (enterprise) content management. Of course, when the term information is used in

[7] http://en.wikipedia.org/wiki/Information

the context of records, this is important. For all kinds of legal, historical and other records-keeping activities this is a valuable part of information management. It is also frequently the easiest part, as it can be solved by plain old technology. However, one rule of thumb to keep in mind here is that the better the information is stored away in IT systems, the least likely it is to be flowing through the minds of information workers.

There's an easy explanation for all this. Historically, it could take ages before a certain piece of information reached all its potential audience and therefore impact. When Columbus discovered America[8] in the second half of the 1400s, it took many months before that news landed back in Spain, from where he started his voyages. When man first set foot on the moon, the footage could be seen with a delay of only a few seconds.[9] Nowadays, breaking news events can be seen live on television worldwide. Alongside the increased actuality of information there also was an increase in the quantity of information. As sharing information has become easier and more effective, human beings are exposed to more information every year. In fact, as I read somewhere, the information my grandfather received throughout his entire working life mounts up to the same quantity I get exposed to in merely 3 months. Imagine the vast amounts my children will have to deal with. So, there's much more information and it gets to us much faster after the actual event that triggered the information. In the old days newspapers could be useful for reading about events that took place in far away countries months before. Later on, in the late 90s, yesterday's newspaper became good for the cat box or to wrap fish and vegetables. Traditional newspapers have to repurpose themselves to providing in-depth background stories while maintaining a real-time online publication for the real news. For real-time information, we have moved to channels such as social media. When Michael Jackson died, no one could beat the speed of information sharing of Twitter.

> Writing down information professionally may actually decrease its value, as the process of professional writing takes up to much time.

For information to be valuable, we therefore need to focus on how information can contribute to smart (efficient of effective, depending on the context) behavior of information workers. Storing information in IT systems is just basic hygiene, but often not contributing to the real value of information.

1.1.3 Enterprise Information Management

Enterprise information management (EIM) is a field of study that aims to increase the return on information. This is done by using two approaches.

First of all, EIM works from the information worker's perspective. New methods find their ways into the fields of activity of decision makers, directors, managers and

[8] Of course, Columbus was not the first European explorer to reach the Americas (having been preceded by the Norse expedition led by Leif Ericson), Columbus' voyages led to the first lasting European contact with America. http://en.wikipedia.org/wiki/Christopher_Columbus

[9] http://www.parkes.atnf.csiro.au/news_events/apollo11/tv_from_moon.html

employees. They all have a need for information, which can be found through different channels, often by the individual himself.

Secondly, EIM reasons from the perspective of information as a means of production. Seen from that perspective, a managerial approach is necessary, and the return on information needs to be transparent. This approach also offers the possibility of leveling the 'Chinese Walls', traditionally built by IT, between different fields of study.

This managerial approach searches for the place in business processes where information can make the biggest difference. Seen from a commercial perspective, we are talking about the value of information for the (sustainable) competitive position. For the public sector this means that (public) resources are used as effectively as possible, to reach the goals that have been set.

The best way to illustrate this is to look at an extremely simplified version of reality. Take, for example, a company (EIM International BV), which is about to organize a big and exclusive (and therefore expensive) event for its clients. Due to the exclusive character and the costs, only the most valuable clients can be invited. A report which analyses the turnover of each client has been made to decide which clients are most valuable for EIM International BV. This shows that client A's turnover has significantly diminished in the last 3 years. Client B's turnover has remained stable over the last 3 years, and client C's turnover has significantly increased. Only one of these clients can be invited to the event.

Although the choice, based on the abovementioned facts, seems obvious, it is not. By looking at the underlying information (e-mail correspondence or meeting reports) you get a completely different picture. The reason client C's turnover has increased so significantly is because they have been working towards the end of a major project. Now the project is coming to an end, and client C has informed EIM International BV that they are going to give future assignments to different companies.

Client B is in its operational management largely dependent on EIM International BV. This client operates in a market which will in all likelihood remain stable in the years to come, which means that EIM International BV's turnover with client B will remain stable as well.

Client A's history shows that they have continued to grow over the years. Client A initially started out as a small company, and now they are working on the launch of a new product, based on collective investments. EIM International BV supplied materials and services during the first phase to get the project started. As this process continued, and the product became ready to be launched, client A focused on marketing and sales. This new focus is the main cause of the decreased turnover. However, once the product has been launched, EIM International will continue to supply the products and services they supplied during the product development phase. Client A's product appears to be a hit, and EIM International BV expects to cash in on client A's impending success.

Based on analyses of historical data (typically the domain of business intelligence) client C should receive an invitation to the event, while based on unstructured information (typically the domain of content management) client A should receive and invitation.

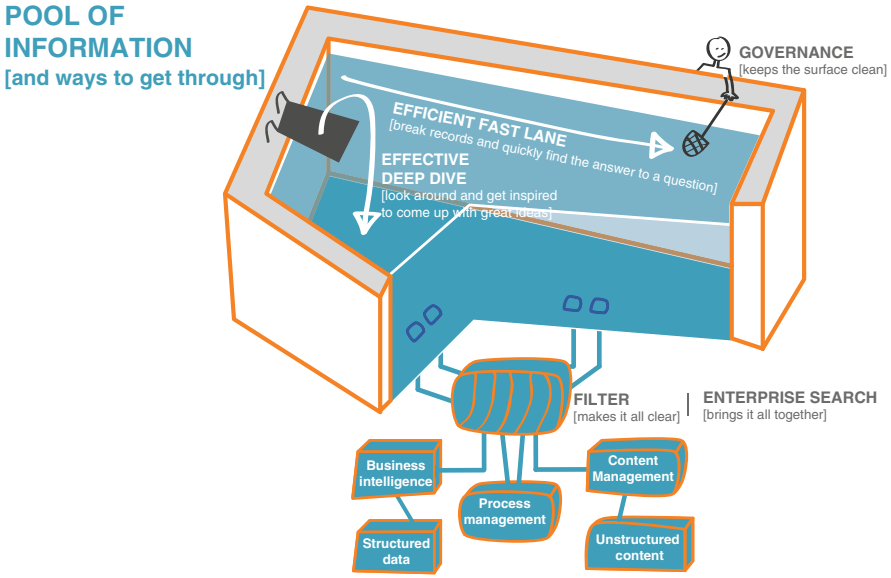


Fig. 1.1 The pool of information

This example reinforces the idea that decisions can only be made based on comprehensive information. EIM tries to level the traditional walls within IT (between BI and ECM) and focuses on the return on information present within the organization, to allow quick and adequate decision-making.

A managerial approach to a subject demands an interdisciplinary way of thinking about a topic. Thinking in terms of return will cause a more economical perspective, and the focus on the knowledge worker appeals to the social- psychological aspects of data processing. Legal aspects can also play a role in certain situations (Fig. 1.1).

Actually, all internal information sources can be considered the water in a pool. The water is maintained at a certain temperature and hygiene has to be monitored and managed. This is all done by the machinery in the pool house. That's where the hardware (IT) is. The pool can be used for deep-diving, as well as for racing. You need to know which part of the pool best supports what kind of activity. When you're not sure, a pool guard can give directions and tell you how to prepare for your nautical experience. The guy with the net is there to keep the pool's surface clear of leaves and bugs. A well managed pool is a joy to any swimmer, whether you just want to stroll under water or are in a hurry to increase your fitness.

1.2 EIM from an Economical Perspective

Traditionally, return is measured in terms of financial parameters, whereby return on investment (ROI) is most commonly used. The return is related to the capital factor of production. This is relevant from a financial-economical point of view, but

from a competitive position there are different points of view that offer more insight. Capital is a material that can be easily obtained, and it is available for set and known risk premiums.

Information is a more elusive factor of production, and optimum usage can lead to a competitive advantage. This can lead to an increase of ROI. By focusing on the return on information instead of the derivative return on investment, new insights, which would otherwise remain hidden, can be found.

The elusiveness of information as a factor of production is evident from the efforts of professor Paul A. Strassman's (Strassman 1999). He has developed a model that makes it possible to talk about the return on information in terms of capital. He can derive the return on information from annual financial statements. This model shows that the percentage of return in mature industries (such as the mining industry) is more or less the same for all companies. This industry has apparently been industrialized to the point that there is no competitive advantage to be found through information. This is however not the case in younger industries, where significant differences can be found, showing that not all organizations use information as efficiently as possible.

The productivity of information is calculated on the basis of a company's Information Value-Added: the value added to the economical performances of a company through proper management of information, which equals the profit minus the costs of the invested capital by the shareholders, which has to be divided between the selling costs, general and administration costs (stock, general & administration), or the costs concerned with the management of information (Strassman 2006).

Strassman's model has some advantages and disadvantages. One of the advantages is the abovementioned benchmark effect, which allows organizations to compare their information productivity and draw their own conclusions as to their own efficiency in using information as a factor of production.

An important drawback is that the model does not offer any directions to increase information productivity. In the article "Five steps to improve your information productivity"[10] (Strassman 2006) Strassman gives a few recommendations as to how organizations can reach a higher position on the benchmark list.

The first three pointers deal with the reduction of costs, which has to be done rigorously, according to Strassman. Projects that do not lead to an increase of information productivity at the desired moment should be terminated immediately. Taking half measures does not lead to the desired reduction of costs. Outsourcing plays a role here too. Even when the total profit of a company increases the added value of an employee, seen from the perspective of information productivity, can be on the decline. In this case outsourcing can be an interesting alternative.

The two last recommendations are heavily influenced by EIM and the EIM maturity model, which will be introduced later. Strassman advises to focus on people and innovation. From a humanitarian perspective, the added value of many traditional roles in

[10] Baseline Magazine, October 2006, Paul A. Strassmann.

respect to low-wage countries is under a lot of pressure. Employees and managers that are able to see how information productivity can be increased, especially if this is based on solid knowledge of the organization and a quality perspective, are worth their weight in gold, and should never be replaced by an outsourced alternative.

Innovation is also an aspect of EIM. The EIM maturity model, which will be introduced later, can help organizations identify areas where information can lead to a competitive advantage. It also helps to identify those areas where this is not the case, and where future investments in information technology are only justified when their aim is to guarantee continuity.

A top 500 of organizations has been made based on Strassman's model. This list offers an insight in the importance of adequate integral information management for high-scoring companies and the role of de CIO.

1.3 EIM from a Social-Psychological Perspective

The way in which the knowledge worker uses information is important if EIM focuses on the return on information from the knowledge worker's perspective. The increase of return on information always has two phases. Phase 1 has to make sure the information is available and in phase 2 this information has to be offered in such a way that it leads to better and quicker decision making by the knowledge worker. This means that one needs to know in which way the knowledge worker likes to receive the information, but also how he or she processes the offered information.

The big challenge is to determine how the knowledge worker will process the information, since it is very difficult – if not impossible – to know just how exactly information is processed by knowledge workers. Moreover, people find it hard to offer any insight into their processes. There is actually a big discrepancy between the decision making process as it is described by the knowledge worker and how it actually took place in the knowledge worker's head. Often, many aspects of the process are rationalized in retrospective, even though the subconscious plays a crucial role in the actual process.

> The best decisions were made after the subconscious had had some time to process the information, without actually thinking heavily about it.

Many researchers have discussed this phenomenon, and in the Netherlands, professor Ap Dijksterhuis of the Radboud University in Nijmegen is a prominent researcher, who is also internationally acclaimed for his research in the role of the subconscious (Dijksterhuis 2007). Investigative journalist Malcolm Gladwell has also discussed a few interesting aspects (Gladwell 2005).

> You can't act on what you don't know.

Even though our book is not a social-psychological treatise that describes how to influence knowledge workers in their decision making, it is important to investigate several aspects of decision making.

It appears that the processing ability of our subconscious is 200,000 times as big as our conscious thinking facility. Professor Dijksterhuis examined the quality of decision making with regards to increasing complexity, for example by an increasing number of variables that had to be considered during the decision making process. He concluded that the best decisions were made after the subconscious had had some time to process the information, without actually thinking heavily about it (sleep on it). More importantly, decisions made after rational thought were often worse than decisions made based on what most people call 'intuition' (Dijksterhuis 2007).

In 2001, professors Nick Chater and Koen Lambert of Warwick university published an article that discussed how the worst decisions are always made at the most important moment. They argue that human decision making is always the weakest link, despite the fact that it is a crucial aspect in the decision making in company processes and therefore vital for the organizations' success (Chater et al. 2001).

It appears that experts always make worse decisions in their own field, even if they are competing to the most simple linear models that weigh the consequences of a decision. Experts apparently want to apply their knowledge when weighing factors that influence decisions. Another important factor that negatively influences decision making is the overestimating of one's own abilities. People feel that they are more often right than that they actually are. It has been shown that if 80% of the respondents think they scored really well, only 65% of them actually did. This overestimating effect is one of the most common effects in human decision making. The risk of making the wrong decision increases because of this.

A fool with a tool is still a fool.

As mentioned earlier, this is not a book about the social-psychological aspects of EIM, but Chater and Lambert fortunately offer some support for the improvement of decision making.

First of all the use of good, statistical decision support solutions, based on actual data, is more reliable than trust in the opinions of experts. We will demonstrate this in the next chapter when we discuss data-driven organizations. This is in favor of the business intelligence aspect of EIM, but the phrase "A fool with a tool is still a fool" is still relevant in this case. Everyone with a basic understanding of Microsoft Excel can pretend to be a statistician and offer his or her own version of reality.

Chater and Lamberts do believe that expert knowledge can indeed contribute to the decision making process. Experts should advise which information is relevant to be used for a management reporting solution. The management reporting tool itself is responsible for generating reports. Experts can also create methodologies for the measuring of scores beforehand, for the data compiled in the management reports.

In order to make the information compiled in this manner available and to maintain it for future usage, efforts have to be made to add it to the corporate memory. The corporate memory is often supported with a document management or enterprise content management solution, where different decision making processes and their results can be found (Choo 2006).

From the social-psychological perspective, there are many opinions and ideas that influence the productivity of information within and between organizations.

In this book we want to focus on the areas that traditionally belong in the comfort zone of the chief information officer (CIO) and the chief technical officer (CTO). It is however advisable to go on a company "excursion" once in a while, and look for information in adjacent fields, if only to become more inspired.

1.4 Business Value

All this talk about Enterprise Information Management could just be another hype invented by consultants to guarantee themselves work for the coming years. The question is therefore: is there any business value in considering and managing information as were it an important production factor that needs constant optimization?

> In a modern economy, information should be the prime asset—the raw material of new products and services, smarter decisions, competitive advantage for companies, and greater growth and productivity," writes Steve Lohr (reporter on technology for the New York Times). But, "is there any real evidence of a 'data payoff?' (Lohr 2011).

In his article titled "When There's No Such Thing as Too Much Information"[11] Steve Lohr refers to research performed by Erik Brynjolfsson and others at the Massachusetts Institute of Technology (MIT) (Brynjolfsson et al. 2011). They are the first to have adopted a quantitative approach to finding the truth in statements that businesses that actively use "real" information (instead of everyday gut-feeling-based decisions) in their decision-making are more successful.

Together, they studied 179 large companies. They found that those that adopted "data-driven decision making" achieved productivity that was 5–6% higher than could be explained by other factors, including how much the companies invested in technology.[12]

Also, they found that the improved performance could also be seen in other measures such as asset utilization, return on equity and market value.

The research was performed to support the growing volume of case evidence that there indeed exists a relationship between more precise and accurate information and decision making and thus firm performance. These cases are all set in specific situations, but still lacking was independent, large sample empirical evidence on the value or performance implications of adopting technologies to support "data-driven decision making".

There are two aspects of this research that need further exploration in a book on Enterprise Information Management. At first, the research seems entirely focused at data, which is the structured variant of information most commonly connoted to

[11] http://www.nytimes.com/2011/04/24/business/24unboxed.html?_r=1

[12] "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?", Erik Brynjolfsson (Massachusetts Institute of Technology (MIT) – Sloan School of Management; National Bureau of Economic Research (NBER)), Lorin M. Hitt (University of Pennsylvania – The Wharton School), Heekyung Hellen Kim (MIT – Sloan School of Management), April 22, 2011.

business intelligence, which is just one of the IT pillars under Enterprise Information Management. The research itself also suggest so by mentioning Business Intelligence specifically. However, when studying the variables used to determine the use of data for decision, the questions asked to senior HR managers and CIO's are not that specific on whether data is the key ingredient or maybe information in general (including content) is in fact being studied. However, it is safe to conclude that data alone represents a significant improvement in performance.

> By presenting large amounts of information in easy interpretable ways, infographics are trying to bring back the rationale in decision making.

The other interesting aspect is that of the rationality that is assumed in decision making processes. As we will discuss in this book, the question is whether most questions are indeed made on such a rational basis. In fact, we know the answer to that question. They are not. There is an ongoing debate in neuroscience on whether the growing insights into the functioning of the human mind should be considered an undeniable condition, meaning we should accept the irrational character of decision making and trying to influence decision making via neuropsychological effects such as priming. Others claim non-rational decision making is an human answer to information overload. With too much information at hand, we tend to overlook the information all together and decide upon gut feeling. Infographics are a trendy topic that try to tackle this. By presenting large amounts of information in easy interpretable ways, they are trying to bring back the rationale in decision making.

> The human mind has techniques that allow it to make good decisions in snap seconds.

Still others think we should not give in to the tendency of the human mind to decide for us. We should facilitate rational decision making by creating time and place for proper analysis of data. The research performed by Brynjolfsson et al suggest this approach does bring improved business performance. Nobel prize winner psychologist Daniel Kahneman[13] is a strong believer in this approach. He claims that our subconscious is dumb and can rarely be trusted, and he has plenty of examples to proof so (Kahneman 2012). He distinguishes between what he calls System 1 (intuitive decision making) and System 2 (rational or calculated decision making). Via all kinds of experiments he demonstrates that subconscious aspects such as risk aversion (fear of loss) and exaggerated self-confidence dangerously impact our decisions. Even in case of life or death decisions. More on the influence of neuroscience on information productivity later.

However, the most important obstacle to overcome might not be the collection and presentation of data, it might rather be the lack of discipline to carefully analyze this data and decide upon that. In that sense, again information productivity depends heavily on change management, as we will explain later in this book.

The human mind has techniques that allow it to make good decisions in snap seconds. And yet, more and more knowledge is collected on how the human mind is easily affected by external factors, causing it to make the wrong decisions. Relying

[13] Daniel Kahneman, "*Thinking, fast and slow*".

on gut feeling is a very human thing to do, and can save a lot of time. And when it comes to really complex decisions, just studying the data might just give you that edge over your competitors.[14]

1.5 Enterprise Information Management Origin

Analyst firm Gartner first mentioned the concept of Enterprise Information Management in 2005. Since then, Gartner has been publishing regularly on the subject, mainly driven by lead analyst Debra Logan.

In 2006, Forrester followed. They too embraced the idea that in order to deliver business value, traditional IT barriers needed to be taken down.

Both analysts seemed to be far ahead of the troops. Although they claimed to be advising large enterprises on the subject of EIM, the lack of cases to demonstrate the power of the concept frustrated them dearly. They seemed to start losing confidence in the survival of the concept. One day, one of the lead analysts told me – when confronted with the fact that the market didn't seem to pick up on the concept of information productivity – that EIM wasn't that much of a new thing. He suggested it was more like a new wrapper around the existing concept of making information work effectively for businesses.

In that same period (it must have been in the second half of 2007) we were researching the subject of Enterprise Information from an IT angle. Our main research question was what technical integrations were needed in order to bridge the gap between structured information (data) and unstructured information (content). These two types of information had created a Chinese wall between business intelligence (BI) experts and enterprise content management (ECM) experts. They could no longer understand each other due to the progressive effects of jargon. BI people seemed to be doing a whole lot different stuff then ECM people.

So our initial efforts were aimed at bridging the gap and bringing down the Chinese wall. We had BI and ECM people sit together and work on greenfield architectures that would help the end user, the information worker. We were looking for ways to keep the artificial difference between content and data away from this information worker. Information mattered; IT people who couldn't embrace concepts from outside of their small field of expertise invented data and content.

We found several ways of useful integrations of the data and content silo, shown in the diagram below (Fig. 1.2).

Surprisingly, we found no meaningful way of integrating data and content at the level of the sources itself. As always, context is a decisive factor in judging the value of information, be it data or content. As context was added in higher levels of the

[14] Techies have an appropriate saying for this. RTFM is a well-known expression they present to people trying to achieve things without proper analysis of information beforehand. RTFM stands for "Read The Manual!". You'll know what the F stands for.

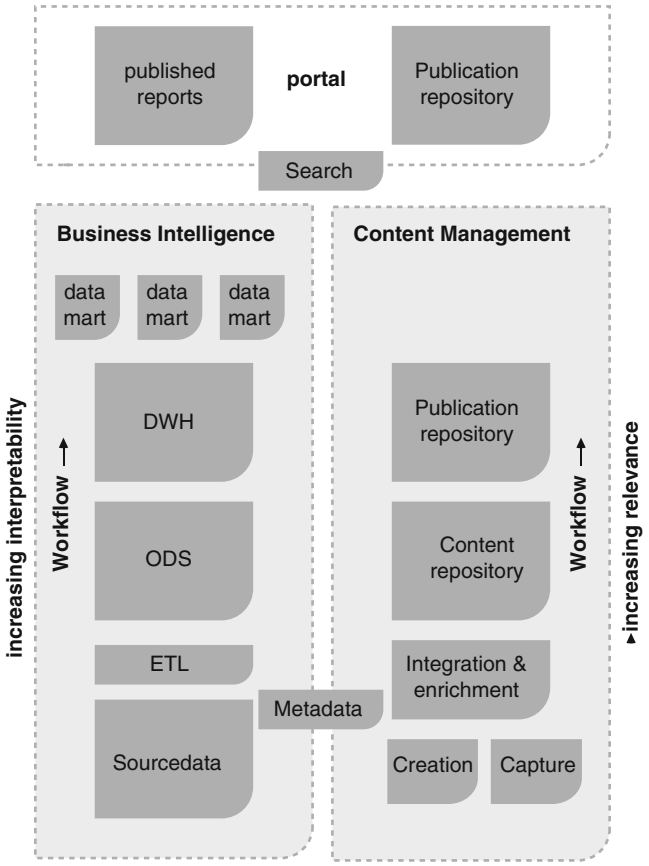


Fig. 1.2 Diagram initial EIM architecture

silo (where information was actually being presented to the end user), our focus moved upwards.

Up in the silo's we found that one could use the BI column as a leading column. Content (unstructured information) could be presented alongside reports generated by business intelligence solutions. Actually, reports themselves could be considered unstructured from the viewpoint of an ECM expert. That led to the belief that the ECM column could also be the leading one. When appointing the leading role of presenting information to the end user to the ECM column, you gain control over the publishing process. On the other hand, you could lose analytical functionality provided. That typically was a strength to the BI column, being able to not only present reports, but also allow the end user some flexibility in the way the information is presented, thus allowing the end user to analyze data using different viewpoints.

We concluded that in order to define which column should be leading, we needed to know the actual needs of the end user. That's where we struck a well-known obstacle in the world of business and IT working together. The real needs of end users in terms of information are hardly ever clear to IT. And if they are, there are

often big differences across business department, job roles and even individual information workers.

This lead to the third way of integrating; an enterprise search solution would allow end users to themselves stroll freely through all available information. This solution would not impact the existing columns, it would just provide end users with an integral view on what is available. This solution came with one huge benefit. Most search solutions bring monitoring functionality. So by studying the search queries information workers used to do their work, we could learn to understand their needs and thus in the long run decide upon a leading column.

Since then, we have seen an immense growth of Google-like implementations within enterprises. Google-like almost became a hard requirement, as Google is the one search tool everyone is familiar with. We ran into companies were information workers were complaining that it was a lot easier to find any information outside the company using Google than it was to find internal information.

1.5.1 The Power of Measuring

Alongside with the technical-oriented discussion on integrations we started looking at ways to measure information productivity. It soon turned out that we didn't have to dig deep. Paul Strassman[15] had – as mentioned earlier – worked for many years on a method that could do just that. In 2004, he published a book[16] in which he stated that "It is only a matter of time before corporate leadership will shift attention to information management as a resource of greater economic leverage than any other input." It's an interesting article that puts information management in an economic light. He uses a lot of data to demonstrate his point on the importance of information management for instance by looking at the declining importance of US manufacturing firms.

In his book, Strassmann presents a formula that can be used to calculate information productivity. His formula has a big advantage in the fact that it can be applied to widely available data on financial performance of organizations, as it is build on data that is presented on annual reports. That allowed Strassmann to benchmark whole industries. He was able to determine which companies in a certain industries were most effective in using information as a resource.

However, in our eyes there is a big disadvantage to his approach. The Strassmann-formula can be used to pinpoint your current whereabouts, but it doesn't tell you were to go.

His approach is rather academic and doesn't help to understand what to do in order to improve information productivity. It's like that balloon pilot that has lost his way in heavy fog and suddenly finds himself flying close to a skyscraper window and asks a person in the building to please tell him where he is. The answer "you're in a balloon" is 100% correct, but doesn't help much. The joke continues by

[15] http://www.strassmann.com/

[16] Strassmann, P.A., Defining and Measuring Information Productivity, 2004.

explaining that even though the answer was useless, still the balloon pilot was able to perfectly navigate from that point towards a safe haven. The explanation is that the balloon pilot immediately understood he was dealing with a typical consultant, so he knew what building he was flying so close to. In the consulting world, this is always the competitor's building, of course.

The Strassmann formula gives us a very adequate description of our current standings, and possibly even comparable to our competitors. But what we have to do to increase our information productivity remains unclear.

1.5.2 The Importance of Information Productivity

Before going into which factors have an effect on information productivity, we will define the term and provide argumentation for its importance.

Strassmann found that information technology played an increasingly significant role within companies, but was not regarded as important as other functions due to the difficulties associated with measuring its effect on financial performance. To address this problem, he came up with Information Productivity, which can "measure the value added to a corporation's profitability by information-centered tasks" (Strassmann 2004). It is a multifactor approach to evaluating the gains from a better use of information, which include a number of quantifiable measures such as R&D and software expenses. It has been empirically tested on a number of occasions and there is a high correlation between companies ranking high in Information Productivity and overall financial performance (Strassmann 2004). Several studies have likewise confirmed that focusing on information and knowledge sharing can be financially beneficial to a company (Ranjan 2008; Peslak 2003; Mehrjerdi 2010).

1.5.3 Information Productivity Factors

"What are the factors – that might be controlled or managed – contributing to client success in terms of information productivity, and what is their relative importance?"

The following paragraph seeks to identify and weigh factors that have an influence on a company's success in obtaining higher information productivity. Information productivity can be defined as the value-added from a more efficient use of information. Through an in-depth literature review students from the Rotterdam School of Management constructed a model (Larsen et al 2011) (Fig. 1.3).

Their theoretical findings suggest that the main focus area should be on the responsibilities of management, where especially commitment from top executives, a careful choice of management style and a high degree of organizational preparation are of importance. Secondly, employees are too often found to be lacking in proper training, and problems with nonuse or misuse of the implemented systems often arise. Information Productivity is also likely to be reduced in highly mechanistic and

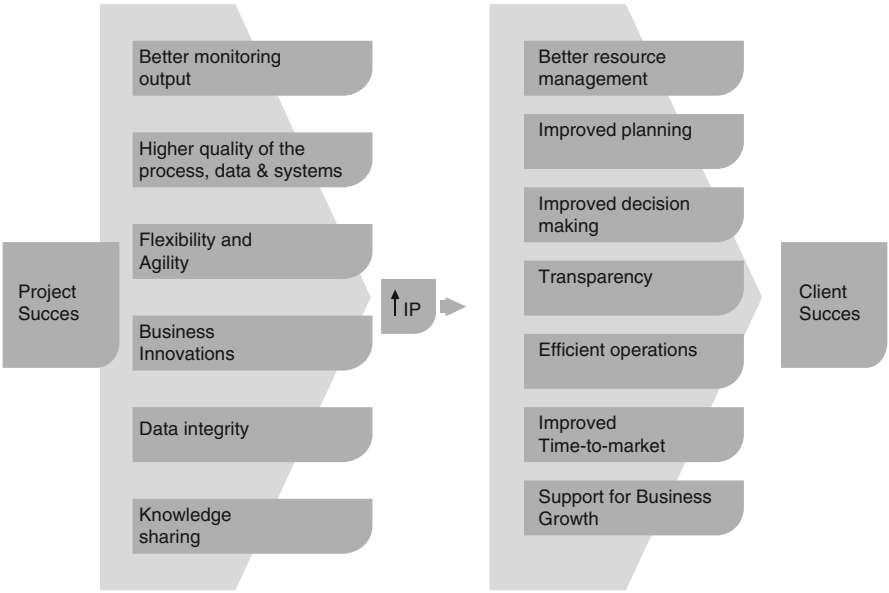


Fig. 1.3 Factors of information productivity

bureaucratic structures, since research has shown that these lead to less communication between different departments. In those types of companies a focus on creating interdepartmental knowledge-teams can be highly beneficial. Finally they found that many companies fail to change existing incentive systems when introducing new processes, resulting in mixed signals to the employees and suboptimal performance.

1.6 Differentiating Factors Influencing Success of Company

In identifying the factors influencing the success of a company, it is crucial to define what constitutes success for a company. Unfortunately, there is no single definition of success across industry and businesses. Inarguably, the definition of success is achieving goals but the true meaning is subjective (Lim and Mohamed (1999)), dynamic, complex, and varies accordingly to companies' goals and strategy. According to Kirby (2005), it is a challenge to define factors driving the success of company.

The differentiating factors influencing the success of company related to information productivity can be viewed from two continuous perspectives; the success of the project implemented which lead to increase in information productivity, and assumed to be beneficial to the company leading to economic growth and company success. According to Atkinson (1999), the project success is divided into three stages:

1. The delivery stage: the process: doing it right;
2. Post delivery stage: the system: getting it right
3. The post delivery stage: the benefits: getting them right.

Various researchers have identified factors necessary for the success of the project. Freeman and Beale (1992) advocated the scope and quality of the technical performance. Atkinson (1999), & Pinto et al (1988) consider that participants' satisfaction (known as 'soft factors') and satisfaction on the interpersonal relations with project team members are crucial measures while Wateridge (1998) supported the believe that project specifications and goal definitions are equally important.

The current state of global network economy has changed the environment in which there is a paradigm shift from data – information – knowledge – wisdom (Leong and Jarmoszko 2010). This global network economy exists due to the emergence of more advanced infrastructure and technology encouraging the economy and society to be up to date. In order for the firm to be sustainable in the industry, they should be lean and agile to the changing environment of the economy, innovative in global network economy and being self-entrepreneurial. Moreover, firms must implement strategies that are effective in responding to the opportunity or threat posed by the changing environment. These enterprise strategy factors will equip and enable companies to sustain competitiveness and capabilities to grow and be successful.

The factors that are discussed in the research paper (Larsen et al 2011) are within the four main scopes of management, employees, structure, and planning and control, after which a propose model was structured. However, it is important to note that external factors (e.g. financial, economy, political, and social) play a part to the overall success level of company, but are not considered in this research. According to Kearns et al. (2004), there is difficulty to link the benefit of IT investment and financial performance due to their complexity, timeline and indirect relationship, and difficulty in controlling variable that impact profitability. In order to gain competitive advantage, the use of IT infrastructures should represent IT applications and support for business initiatives.

1.7 Enterprise Information Maturity Model

The EIM maturity model was developed to help organizations increase information productivity. Even though Strassman's earlier mentioned method for the return on information offered relative insight (benchmarks), it is not the starting point for a new approach to improve the return on information. There is a large demand for such a starting point in practice. The return on information is still a rather abstract term, so a model that will make it tangible and usable is very welcome.

1.7.1 ROI Is Return on Information

The model is based on the idea that it is not possible to talk about return on information in absolute terms. The search for an absolute and measurable form of return on information is an academic one, which does not help organizations move forward. Developments in this field should be followed, but it is unwise to wait for the results.

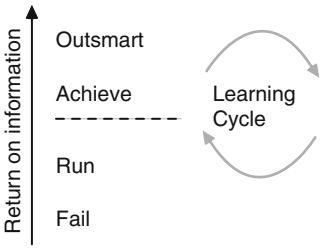


Fig. 1.4 Return of information

Economical developments, that originated during and after the economical crisis have made it interesting to start using the return on information as soon as possible.

During the 2009 crisis, several analysts explained how economical recovery would take place in three phases. The aim of the first phase is to reduce costs. The goal is to create and maintain a positive cash flow as soon as possible. The organization's continuity on a financial-economical level has to be maintained. In this phase, investments have a short-term return period.

The second phase deals with the organizing of efficiency. In this phase the business processes have to be established in such an efficient way, that the financial-economical continuity obtained in phase 1 can be maintained long term. From a competitive point of view it is also necessary to use factors of production efficiently, for example from the perspective of cost prices or time-to-market. Investments in this phase have a longer return period.

In the third phase, new ways of growth are found through innovation. Key word is the efficiency of processes, knowledge and human recourses. Investments in this phase have the longest return period, due to the sustainability of the possible competitive advantages. These advantages arise because the emphasis is on the smartness of an organization, not the efficiency. These competitive advantages are therefore also not easily copied; we can all work quickly and hard – or at least try to. Working efficiently is another matter.

The aim of the EIM maturity model is to help organizations to designate which processes or information domains can be used to increase the competitive advantage. This model allows organizations to see which investments need to be made in which information domains.

The model identifies five stages of return on information (Fig. 1.4). The lowest level is fail, which means that there is not sufficient information available to run certain processes. In the next level, a specific domain is able to run properly since enough information is available. This level is known as run. Comply[17] means that it is possible to report the abovementioned return on information adequately to the relevant agencies.

These first three levels of return on information hold very little surprises. They are most commonly used for processes and information domains where there is little competitive advantage to be gained. We can also view them as dissatisfiers.

[17] Comply is not shown in diagram above. This level is not always relevant. Basically, it covers the "run" level including the ability to report to internal and external agencies that require to do so.

Something only goes wrong if the processes or information domains are unable to reach even the minimal level of return. It is often useless to try and increase the return on information within these processes or domains, since that only costs money and does not lead to any significant top line results.

The two top levels are more interesting for two important reasons. First of all, these are the levels where competitive advantages can be made (short and long term), and secondly, in order to meet the current criteria the environment has to be monitored constantly to be able to operate on this level. It is necessary to implement a learning cycle to maintain operation at this level. This learning cycle is meant to map the changing demands of the market and organization.

The first maturity level of the learning cycle is achieve, which is aimed at efficiency. The main goal is to handle processes as quick and flawless as possible, while at the same time using the available information as efficiently as possible. Business process management (BPM) and operational business intelligence (OBI) are the terms that play a role in this maturity level. It is important to make quick decisions at this level. From a social-psychological perspective this deals with less complex decisions (fewer parameters) that can be supported with statistical analyses or other data-driven sources. Background information can be crucial to the context, so knowledge of files is also necessary.

The highest level of the maturity level is known as outsmart. At this level, the main goal is to make better decisions than the competition, though based on the same information. Organizations that manage to operate several processes with respect to return on information at outsmart level have a strong competitive position. Not only does the right information at the right time need to be available at this level, but it also needs to influence the knowledge or information workers in a positive manner. This means that there is also room to find other ways than the traditional ones to use information, and come to smarter opinions by doing so.

> Something one isn't looking for, can be of greater value than the thing one was looking for.

Just an example of a phenomenon that can happen when the maturity level outsmart is reached is serendipity. Serendipity means that one finds something one was not looking for, which ultimately turns out to be of greater value than the thing one was looking for in the first place. Famous examples of discoveries through serendipity are penicillin, Post-it memo notes and Viagra. Serendipity is also a special process; it means that the knowledge worker in question is able to run thinking-processes on top of his operational work that can generate more information in the background, which can lead to meta-insights. This also means that there is room for pondering, and that the knowledge worker can be fed information that may not be directly applicable to the task at hand.

This creates a clear distinction between the achieve level and the outsmart level. In the achieve level the only goal is to act quickly and adequately. In the outsmart level however, smart solutions have to be created. An environment where the emphasis is on the process, and where industrialization has to lead to an efficiency profit, does not often generate huge innovations. This will most likely happen on the level of process (can we do it more quickly?), but not on the meta-level (is there a better market outside this one?)

This emphasizes the crucial importance of estimating the desired level of ambition per domain of information or process, in order to have a competitive advantage. The simplest way to avoid capital mistakes is to always reason from a business perspective.

A good example can be found in an enterprise information management scan carried out in close collaboration with an IT manager. An analysis, based on a clear definition of the business of the employer, was made of the information processes that can play a role in the improvement of the competitive position. Subsequently, an estimate, based mostly on simple common sense, was made for the levels of maturity that should account for the information process. During the discussion of the information processes, which according to the model should operate on the achieve or outsmart level, the IT manager said that those processes were also the topics the management always held him accountable for.

The model gave the IT manager some guidelines in the translation of the organization's goals into the IT policy and the priorities within that policy. He also saw possibilities, again based on the model, to clarify why certain IT investments should not be made, which could be supported by the organization's goals.

Apart from the five levels there are two points of interest which apply to the entire model; information management and information value.

Aspect	Meaning
Information management	Information management, at management level, is concerned with the ways in which an organization can provide information effectively and efficiently
Information value	What is the added value of information for the organization

These two terms are always the assessment criteria for a certain level in the model. The following meanings are given for the assessment criteria per level.

Phase	Information management	
Fail	Collecting	The organization only focuses on the day to day business processes without looking at them in connection with the rest of the organization. There is no information management. All decisions are made at local level and the required information to justify decisions is generated manually
Run	Organizing	The organization has made a 'map' of the sources of information and has knowledge of duplicate processes (writing down of customer data more than once) and knows how the information is used and formed during the process
Comply	Reporting	Process and data managers have been appointed so that data is only written down once, and is then used by the rest of the organization. The definitions of the data elements are known and are used by the organization
Achieve	Processing	The quality of data is an important issue in this phase since it can improve the effectiveness and efficiency of the processes. Data stewards are appointed to oversee the improvement of quality
Outsmart	Sensing	Information is an asset and new ways to use this are researched. New opportunities are used to anticipate tomorrow's needs and are used to draw more information from the source registration

Phase	Information value	
Fail	Integrity	Information is a burden to the organization as it uses the existing system for the daily processes, without worrying about the consequences of the work-arounds
Run	Formality	Information now has a formal meaning within the organization. This will lead to less discussion about the meaning of the terms used, among other things. Reporting has become more clear and certain elements have added value. The term KPI comes into being
Comply	Control/transparancy	Information is realized in a clear manner and company rules are known. Everybody has the same information, and makes decisions based on that information. Information is used to justify decisions
Achieve	Sharing	Data is used by the entire organization and leads to clear and efficient processes. Data is only written down once and there is no discussion about the meaning of certain terms, as the metadata is known and accepted. Information is shared, allowing benchmarking (competitive element) between business units. This increases the efficiency
Outsmart	Proactiveness	The data is used to improve the services offered to the clients. People know what the client wants, what they consider important. Services and products are adjusted, based on the available information, to strengthen the ties between the organization and the client

1.8 Method Enterprise Information Management Scan (EIM Scan)

The EIM maturity model is especially valuable in combination with a specific method, which has been translated into a EIM scan. The operation and phasing of the scan will be explained in more details in Chap. 2 in relation to EIM Strategy.

The next move is to decide which stereotypical statements are true for each information domain and maturity level. Stereotypical statements are the statements you can expect within an information domain if a certain level of return on information is reached. The goal of making such stereotypical statements is that they help to determine the current maturity level of the organization for the relevant information domain.

Take for example an organization that wants to know how information about their business opportunities are used in the sale process. We could call this domain market awareness. The following summary gives some stereotypical statements for each level of return for this domain.

Outsmart	We use our industry knowledge to generate demand for our product
Achieve	We already know what the client wants before they have even asked for it
Comply	We are actively looking for new opportunities within and outside our current network and know our own hit rate[18]
Run	We are ready as soon as the client makes its demands known
Fail	We are reactive and wait for the client to approach us

A possible standard reaction for commercial organizations is that this information domain should always function at the outsmart level. However, there are situations that call for a focus on different information domains. A more or less unique product or service can make sure that certain market demands end up at the relevant organization. In this case it would be wise to focus solely on the information domain that this unique aspect of the product or service can guarantee in the long run (provided it has potency), than to spend a lot of time, attention and money on a process that runs properly anyway. There are many possible scenario's that can influence the levels of ambition for each information domain for each organization to such an extent.

The discussions in this phase offer the possibility to formulate several research hypotheses on which recommendation for future return on information can be based.

The relevant information domains and hypotheses are extensively tested during step 2 of the scan. Extensive knowledge of the organization and its goals is an important aspect of the EIM scan, so comprehensive testing is necessary. This discussion also usually reveals the right people for the interview phase. These are the people that have distinct opinions about the way the organization operates in certain information domains, and therefore also about the return of information within those domains.

The actual scan is done during step 3 (Execution EIM scan). By means of interviews with several people the current return on information in the different information domains becomes clear. This also reveals the level of ambition that belongs to each information domain, though this is usually already known during step 1, the identification of the information domains. It is important to check with the client if the ideas revealed by the scan and discussions are in fact true, to avoid the risk of continuing based on wrong assumptions about the organization's market, which is usually caused by insufficient knowledge.

Eventually it has become clear what the return on information is for each information domain, as well as the desired level of ambition. Any differences revealed at this point should be the basis for further execution of the scan.

The findings of the scan are reported during step 4 Presentation of results, as well as the tested hypotheses. These hypotheses can then be linked to the findings concerning the desired return on information and the actual return on information.

[18] Not all versions of the EIM scan mention this level. Basically, it covers the same maturity level as the run level, but for branches where compliance and reporting are of extreme importance, the comply level can be relevant.

A properly executed EIM scan usually leads to recommendations for improvements, according to priority based on the expected impact of the sustainable competitive position. Current budgets and the organization's capacity to change can be taken into consideration as well.

The EIM scan provides organizations with a guideline on how to fill out the EIM maturity model. In that sense, it adds extra value to the Maturity model. When filled out appropriately, the maturity model will pinpoint areas that need to be prioritized in order to significantly improve information productivity. It separates the necessary from the desired. It still does not come up with a comparable figure as the Strassmann formula does, but instead it brings insight into how to improve the competitive position of an organization in terms of information productivity.

1.8.1 Pitfalls of the EIM Scan

There are some typical pitfalls associated with the execution of the EIM scan. Experience has taught us that participants easily trust in return on information based on rational perspectives, especially since it concerns the axis "business – IT". This means that people are looking for ways to improve the supply of information, yet fail to investigate how this influences the information workers' behavior. This can potentially lead to costly advices and projects that fail to meet the desired outcome. Even if information is available, that does not necessarily entail that this information is used efficiently and effectively. The executors of the EIM scan should continually challenge the participants to think about the effects of the solutions for information workers instead of the solutions themselves.

An unclear focus is the same as no focus at all

Inability to specify the exclusive goals of the organization is another pitfall of the EIM scan. This happens when participants from different departments and hierarchical levels are asked what the important themes for the next year should be, since a great diversity in answers is most likely. It is very hard to decide what the priorities concerning information management should be if the participants are not on the same page regarding the organization's most important goals. The cause of this problem is often twofold. There can be a problem with the way the senior management communicates its goals for the organization, which leads to different interpretations of those goals at different departments, leading to significantly different opinions. Another cause is that an organization, according to the participants of the scan, simultaneously focuses on growth and on consolidation; working on attracting new clients as well as keeping the old clients. This can be summarized very effectively as "an unclear focus is the same as no focus at all".

Improving the communication about the organization's goals is one way to solve this problem, another way is to only have the senior management participate in the scan. Higher up in an organization's hierarchy one can expect more consensus towards the most important goals.

The last pitfall is concerned with the definition of information domains. Undoubtedly there are sharp definitions for information domains from an enterprise architecture. A definition that is highly usable when it comes to the EIM scan distinguishes an information domain as a part of the supply of information belonging to business processes. Information domains usually track the business processes. This leaves enough room for interpretation within the EIM scan, which is why it is wise to allot some time for creating good definitions of the several information domains. By leaving some room for the definition of the information domains the participants are able to add details to those areas that are relevant for that specific organization.

To conclude, the three pitfalls of the EIM scan are:

- Too strong a focus on the rational aspects of provision of information;
- Unclear organizational goals;
- Discussion of the definition of information domains.

1.9 When Information Is not Productive (Enough)

Following the EIM scan it may turn out that there is a gap between desired information productivity and actual information productivity. Since we wanted to steer away from strictly academically analysis, we also wanted to define a structured way to fight such gaps.

We closely examined all kinds of gaps that can occur in real life. Ultimately, we found four factors that have an impact on information productivity. These factors are blocking factors. When they occur, information productivity is limited.

The first blocking factor is the most obvious one. When information is not available, it cannot be productive. When this blocking factor occurs, for sure the current level of information productivity will be characterized as fail.

The second blocking factor has strong ties with the IT domain. Information that is available must be accessible for it to be effective. Accessibility can be restricted because of ill-defined authorization schemes, complex metadata structures that an end-user must understand to access information, or the lack of an easy accessible search solution. The earlier mentioned wave of internal search implementations can solve this blocking factor in many cases. Sometimes information is available within certain silo's that are not accessible for the entire organization. Then too, search tools may prove to be a solution.

The third factor has to do with popular beliefs into how individuals collect information. Studies have demonstrated that people tend to look for information that confirms already existing beliefs. This is quite contradictive to former belief that people will start to act differently once they understand the problem that is caused by or related to their behavior.

The large environmental campaign by former US presidential candidate Al Gore was aimed at just that (Roberts 2010). Convincing people that there is a problem. Millions were spend, and behavioral effects were minimal. So the information was available, it was accessible, but no one started to behave "smarter".

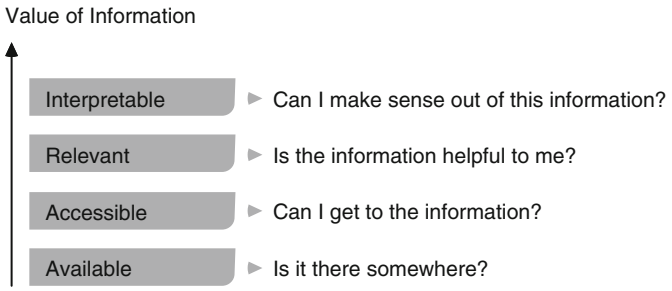


Fig. 1.5 Blocking factors

In the end, it was found that information seeking behavior is driven by one's identity. Our identity determines what information we absorb, and what we ignore. Changing this identity is difficult. The focus should not be on providing information on problems that occur, but on enhancing relevancy by bringing the desired behavior to the people. The manipulative effects of one neighbor cycling to his work every day are much greater than a weekly leaflet on the state of CO_2 emissions worldwide.

Therefore, in order for information to be effective, it also has to be relevant. It has to touch upon the needs and beliefs of the information worker who's behavior is to be optimized towards truly smart and swift decisions. Relevancy is the third blocking factor.

> Most people consider information workers rational actors in an information-rich environment.

The final and fourth blocking factor has its origin in the working of the human mind. In order for our behavior to become smarter, our minds must be able to actually understand what is presented to us and act consequently. No sane person will argue with this statement. However, we see that most people – and certainly the IT industry – consider information workers and other end users as rational actors in an information-rich environment. The fact is, they are not.

Our brain has a mind of its own. Neurological effects such as priming and mirroring have an effect on a person's behavior, even without the person knowing. Furthermore, the immense overload of information has led to scanning behavior. Students nowadays are explained that it takes more than just finding a quote on a website to do proper fact-finding. The rise of the infographic is also one of the industry answers to this phenomenon. Infographics spread much easier than the extensive reports that may sit behind them hem.

Maybe even more important, the artist creating the infographic can decide which data to present and which not, in that sense influencing the image that will remain in the memory of the reader. So, interpretability is the fourth and final blocking factor when it comes to information productivity (Fig. 1.5).

With these blocking factors in mind, any gap between desired and actual maturity level of Information productivity can be further analyzed. The bottom two hurdles can usually easily be taken away, often by means of IT. However, they are not

likely to get you in the two top maturity levels. For that, you will have to find out what you can do in order to improve the relevancy and interpretability of the information you provide to information workers.

1.10 Trends of Enterprise Information Management

We focus on trends in information, technology, processes and organizations. These four combine our view of EIM. Technology, processes and organizations are shaping the world of information. You can change one aspect but changes will have greater effect if you change all. Because technology is catalyst there is a meta-trend that we need to address, the speed of change.

Ray Kurzweil describes in his book "The Singularity is Near" (Kurweil 2006) that we often find not just simple but double exponential growth, meaning that the rate of exponential growth (that is, the exponent) is itself growing exponentially. Looking back at the past and translating progress to the future discards this insight. We are making progress and making progress faster.

Just take a look at the prediction of growth of information and you can see this phenomenon in action. We are creating more and more information every year. Every year a research company publishes a new prediction, it predicts more growth than the last researcher did. Another example is the adoption rate of the channels by which content is published. The famous 'Did You Know' series by The Economist calculated the time it took a medium to reach a market reach of 50 million (Fig. 1.6)

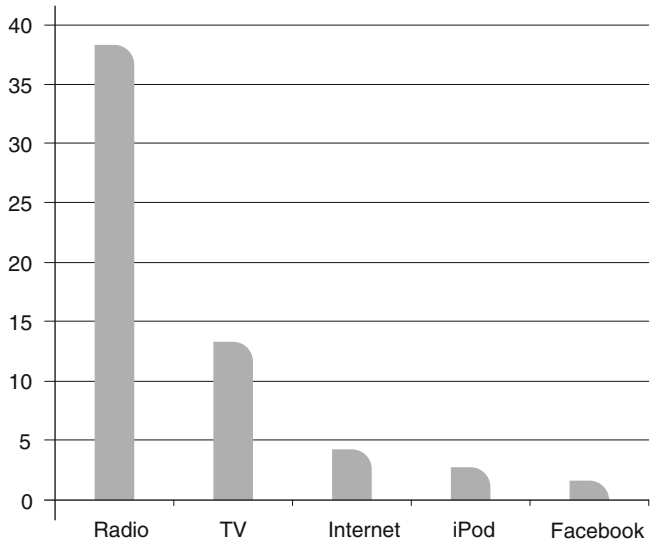


Fig. 1.6 Market reach of 50 million in years

One of the hottest startups in 2012, Pinterest, reached an audience of 11 million in about 1 year.[19] The growth rate of Pinterest is faster than Facebook, which took from March 2004 until December 2006 to reach 12 million users. These are just some examples that the world is changing faster and the pace of change is getting faster as well.

1.11 Information Movements

We see three important movements that shape information, the way we use information and our perceived value of information: abundance, transparency and serendipity. Information is important and will gain importance.

1.11.1 Abundance

We live in a world where the volume of information increases and for the last years we under estimate the increase every time! We upload more video to YouTube every month than US TV stations have aired in their complete existence. Twitter gives people a way to express themselves and share what they are doing. Some people have tweeted complete books. Actually the people of earth tweet 230 million times each day.[20] This means we post more than 10,000 500-page books on twitter every day of the week. In 2010 we sent about 294 billion emails a day.[21] Ninety percent of these are spam but that is still an impressive 29.4 billion emails every single day! For mind-blowing view of information (and a number of other statistics) on the move http://www.worldometers.info.

The abundance of information can lead to information overload if no filtering and other strategies are applied. The information overload challenge has been around since the first book press. In those times a person was already unable to read all the information in the world. Today that still is not possible and the percentage we can read is getting smaller every minute. Choosing the right information filtering is different for each individual.

Information will increase and the rate in which it increases will increase as well. This has impact the way we use information, our reliance on effective filtering we get bigger and bigger. The future will bring us ultra effective filtering but also an epidemic of information addiction if our filtering strategies fail.

[19] http://www.vabsite.com/2012/02/pinterest-users-usage-trends-statistics.html

[20] http://blogs.wsj.com/digits/2011/09/08/twitter-shares-active-user-numbers/

[21] Radicati Group, http://email.about.com/od/emailtrivia/f/emails_per_day.htm

Two specific trends add a lot of information to this ever increasing number. We produce a lot of information actively but the most data is produced without our knowledge. This is called exhaust data. The other big trend that will add information is the Internet of things, earlier this was called Web 3.0. It literally means the Internet and data creating capabilities of things, like cars and buildings.

1.11.1.1 Exhaust Data

In a report on Big Data McKinsey launched the term Digital Exhaust Data[22] (Manyika et al. 2007). In our everyday lives we create some data and information but the exhaust data we leave behind is much, much bigger. If you do one Google search, you leave at least 57 data points.[23] If you visit a webpage on a news site you generate a lot of data points as well. Facebook operates clusters of servers that can store 30 petabytes of data. This is much, much more than the content we add. The data we implicitly leave behind gives a lot more and accurate information on actual behavior. I can fill out a review of a site that says it is nice and easy to navigate. The actual click trails that I leave behind, the time spent on each page and the conversion rate tells a lot more. The popularity of Foursquare and Facebook check-in tracks your whereabouts and movements, but this can also be done with traffic camera's, wifi and GPS. Taking a picture with your iPhone or another device also stores your GPS coordinates with the picture. Apple had a little nightmare with the iPhone when people noticed it was tracking this actively and they could give you a view of your travels on a map, without consent.

The Quantified Self is a movement that lets people generate data about themselves. People are wearing devices that record their heart rate and other vital signs continuously. The applications for eHealth are numerous, for example the data a person can be used to send out alerts if your cholesterol is going to exceed a certain threshold.

What if all these sources of personal data are combined in a complete timeline. Image the Facebook and LinkedIn timelines enhanced with travel, purchase and health data. Imagine that buildings or parking spaces can have timeliness as well, who entered or exited the building, headcount, temperature, energy consumption, etcetera. Is this a dream or reality in 2 years or maybe tomorrow?

1.11.1.2 Internet of Things

Why should only humans add information? Non-humans can do the same. Buildings, sensors, parking lots, every object in public and private space can be internet enabled

[22] http://www.mckinsey.de/downloads/publikation/mck_on_bt/2008/mck_on_bt_13_trends_to_watch.pdf

[23] Eli Parser on http://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles.html

and add information. Arduino[24] is an open-source electronics platform to create internet enabled objects, you can do this today and internet enable your office buildings or homes. People are measuring air pollution, radiation, electricity consumption and sharing these on platforms like Pachube.[25] With the introduction of Internet Protocol version 6 (IPv6) we have taken away a possible hurdle of the availability of IP addresses. Now it is possible to have 100 IP addresses for each atom on earth.

The internet of things could give us information to solve parking problems. Streetline is doing this in San Francisco and a number of other cities in the US., predict nearby traffic jams real-time, the lines at the McDonalds and do groceries based on the contents of the fridge. An application that has been launched recently is 'e-thermostaat' by Essent, a Dutch power company. It enables people to adjust the heating at home with a mobile app. In The Netherlands everybody has a smart energy-meter, which makes it possible to have a near real-time insight into your energy consumption. This also has the potential to actually tell what you are doing based on your energy consumption. The OV chipcard, a pay as you go card in Dutch public transportation, is tracking actual usage of public transport. This should enable the providers to predict higher and lower levels of usage.

The internet of things is very relevant because it opens up a lot more data that can be used by organizations to generate value. This data is objective and provides insight into behavior and status. The power of combination and integration can enable companies to outsmart each other.

1.11.2 Transparency

I like to think there are three forms of transparency. This first and most well-known form of transparency is forced. Society or legislators force organizations to be open and transparent about their work. The second from is active transparency where organizations create transparency to directly create value through information shared with suppliers or customers. The other form is transparency is passive. It creates value indirectly.

1.11.2.1 Forced

This is the type of transparency legislation and society want from organizations. Organizations are forced to open about what they are doing in order to prevent malpractice. Enron showed us that transparency is needed. WikiLeaks showed us that governments need to be open as well.

[24] http://arduino.cc/

[25] https://pachube.com/

1.11.2.2 Active

Transparency is needed to intertwine processes between organizations. Active transparency creates a direct effect on value. Information about progress is crucial to monitor and prepare. Customers want to have insight into the processes that lead to their product. Today you can look into a production line of a car manufacturer and modify the color or certain features up to specific points in the process. Customers of transport agencies can literally see their products on the road at UPS. Transparency means sharing information that for a few years ago was not even available. Today the information is available and organizations share it with their suppliers and customers.

1.11.2.3 Passive

This type of transparency is about organizations. This type of transparency creates a indirect effect on value. This is the transparency others create about an organization. Today we can get a quality rating of a Medical Doctor or a product. In the past it was hard to get an indication about quality of products and services but today we can get ratings about almost anything. People rate quality and others make assessments based on these ratings. Organizations do not have an active role in this type of transparency. They just need to deliver the products and services they say they do. Organizations can build a platform where customers can assess quality if no platform exists. Positive feedback can be used to promote the product, negative feedback can be used to improve the product.

1.11.3 Serendipity

Serendipity is the coincidence of finding something good or useful without looking for it.[26] Think of the next situation, you are looking into solar energy and find a way to improve wastewater treatment. This is serendipity in action. I think priming (I will discuss this in depth later) has a big impact on serendipity. Priming is an implicit memory effect in which exposure to a stimulus influences a response to a later stimulus. If I read about a specific type of car I will 'see' that car on the highway far more often than before. Of course this is not true but I am primed to see the car. This means the definition of serendipity should be 'the coincidence of finding something good or useful without *consciously* looking for it'.

[26] http://en.wikipedia.org/wiki/Serendipity

A force working against serendipity is personalization. For example Google personalizes our search results with 57 data points.[27] Which actually means no search result is the same. The consequence of this personalization is that everyone lives in his or her own information bubble. We must look for a balance where people get results that are relevant but where serendipity has a chance. Relevancy is important for information to matter but surprise and new insights from other perspectives also have value and relevancy.

Personalization of the future will provide us relevancy and serendipity. In the future we need the personalization-filtering strategy to prevent information overload. New filtering mechanisms and algorithms will arrive that allow for serendipity as well.

1.12 Technology

Moore's Law (Moore 1965) prescribes that every 18–24 months chip performance doubles. Most technologies ride Moore's Law and conventional products that get a technology incorporated will catch the ride.

One of the greatest trends here is ubiquitous computing. We will all be able to access information everywhere we are. To make this happen there are two prerequisites that need to be filled in. The first prerequisite is mobile hardware and hardware everywhere. Mobile was a trend in the 2000–2010 era but today it is mainstream. In The Netherlands there are more mobile phones than people. Components are getting smaller and smaller up to nanoscales. This means they can be put in other objects as well. The difference between a wall and a computer screen are obvious today but these will disappear soon. Touch and gesture technology enable these walls to function as computers, like tablets do today. In 2009 the Inamo restaurant used technologies like this to display extra information and take orders.[28]

The second prerequisite is connectivity. In 2011 43% of the Dutch population accessed the internet via mobile broadband. Internet access through Wifi or 4G will be as normal as water flowing from the tap.

Hardware and connectivity evolve and at the same time so do the applications that enable the ubiquitous access to information. Innovative apps like Layar and great inventions like SixthSense of the MIT Media Lab[29] add information from the web to the physical world. They create an augmented reality where the virtual world will enrich the physical world. Layar does so via a mobile phone, the SixthSense uses a small projector to do the same and adds gestures like the Microsoft Kinect. Google will launch Internet enable glasses by the end of 2012 making it possible to be online and be in an augmented reality full time. Next step will be internet enabled contact lenses that we can control with the gestures we see or with the movement of our eyes. Why not?

[27] Eli Parser on http://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles.html

[28] http://www.wired.com/gadgetlab/2009/03/inamo-restauran/

[29] http://www.ted.com/talks/pattie_maes_demos_the_sixth_sense.html

1.13 Processes

Tom Davenport created a simple way to classify processes and work in organizations (Davenport 2005). The first axis is complexity of work. Are the tasks of the work easy or hard and what is the level of education needed to fulfill the tasks. The second axis is level of interdependence. Can a person do a task alone or does he need others? For many people their daily work consists of different tasks that can be mapped in different parts of the map. For work typologies can be deducted from this model:

Transaction (low complexity, low interdependence),
Integration (low complexity, high interdependence),
Expert (high complexity, low interdependence),
Collaboration (high complexity, high interdependence).

In our western world work is moving from transaction work to collaboration work. This movement started years ago and will continue to last.

1.14 Knowledge Work Still Growing

Information and knowledge are closely related. In the past knowledge 'systems' tried to capture knowledge into information and store it for future use. We now know this is senseless. The transaction cost of information is close to zero and will get closer to zero every day. The transaction cost of knowledge is very high. The reason for this is the difference between information and knowledge. The difference is made by experience, competences and attitude (Weggeman 2000). These are a lot harder to pass on from one person onto the other.

The share of knowledge work in the western economies is getting bigger and bigger. In the Netherlands the creating industries have all been moved to countries with low wages like China or Eastern European countries. These industries have been replaced by knowledge work. There will always be some non-knowledge work in some countries but the shift is not finished yet. In The Netherlands the last car manufacturer will most likely move production elsewhere in 2013.

1.15 Organization

The way organizations are organized has impact on the value and use of information. Today we see a shift away from the command and control culture to a culture where people are empowered, trust is a catalyst to share and collaborate. The size of organizations will go down due to new ways generations want to work.

1.15.1 Empowerment and Trust

The winning organizations of today and tomorrow are based on empowerment and trust. This means a big shift from command and control and has a big impact on how information is used. Command and control is a power game. And information, or to be precise, the not sharing of information, enables power. But in the successful organizations of the future information is shared and is used in collaboration. Information becomes productive because it flows and generates value.

In the future of management Gary Hamel describes companies that spread information on performance down to the work floor to enable people to make better decisions on marketing and purchasing of products (Hamel 2007). This company is far more successful than its competitors. Part of empowerment is also knowing what goals to achieve and how to be effective. To do this people need information on performance to assess and adjust.

1.15.2 Organizations of the Future

These will not be the big companies. Smaller companies and companies formed by other companies are the future. The number of creative startups and self employed people are growing every year. Combine this trend with the fact that labor relationships get shorter every year. Our fathers worked for one of two companies in their entire lifetime. We switch every few years but our children will have no jobs but gigs. They will combine forces with peers and competitors and split up just as easily. On boarding of new personnel and consultants has to be about tapping into the right information streams and also sharing information with the organization as fast as possible. For an old school organization this would seem like a big risk and an instant loss. For the organization of the future it will be the best thing to do to get value fast.

This has big impact on the organization of information. If the half-life of information is getting shorter, than people in new jobs need to find and get access to the right information faster. We need to enable people to generate value faster because they are working at organizations shorter. This means getting them relevant information faster. Old command and control ways of withholding information will fail miserably. The new way of sharing and collaborating is the new normal.

1.16 The Last Frontier: The Brain

In the end there is one place where information processing takes place: the brain. This is a blessing and a big risk. At the moment the brain can do more and smarter work than computers or smart phones, so there is your blessing. So if I feed information to the brain smart decisions and ideas will emerge? Maybe.

1.16.1 The Way the Brain Works

This field of research is gaining momentum now we have better technologies to study the brain and the way it works. We are getting a clearer idea of the way the brain processes information and make choices. Economic and decision theory is based on the assumption that a person makes rational choices. For example if the price of a product goes down than sales will go up, as indicated by supply and demand theory. But that simply is not true, because a person only makes some rational choices and makes a lot of unconscious choices. According to Daniel Kahnemann a brain has two systems at work (Kahneman 2012). This roughly is concurred by others like Camerer, Loewenstein and Prelec (Camerer et al. 2005). The first system is the automated or unconscious system. This system is fast, parallel and effortless and you do not have insight into this system. The second system is the rational system which is serial, effortful and you can do have insight in how this system makes decisions. The first system calculates $2+2$ and the second system calculates 175×849. Economic theory is based on the assumption that the second system makes all decisions but research has show that this second system is very lazy. Your unconscious makes a lot more decisions that you are unaware off.

1.16.2 Preparing your Brain

One of the most important concepts is priming. Priming is an implicit memory effect in which exposure to a stimulus influences a response to a later stimulus. This means the way a person processes information can be influenced to make a decision based on that information. A very well know test is a word test. Two groups of students had to make sentences out of words. All students were of the same age and physical fitness. The first group was given words that were about old age like gray, golf, Florida etcetera. The second group had random words. After the test they were asked to pick up the results down the hall. They had to walk 200 m. The test showed that the first group walked slower to the next room than the second group. From an information perspective choice of words can prime the reader.

Anchoring is a specific form of priming. In a group of people everybody was asked to state the last number of their social security number and then estimate the pricing of a piece of electronics. The people that had the lower numbers estimated lower pricing than people with higher numbers. The low number set an anchor for the next number. What if the decision you make actually says more about the way that you were primed than your rationale? What if your manager was about to think about your income for next year and saw a progress report with a lot of zero's? Anchoring can be used to choose the order to present information to the reader. The piece of information (i.e. a word or a paragraph) sets the anchor and will have a bigger impact. Think about this comparison of two persons made by Daniel Kahneman (Kahneman 2012):

Alan is intelligent – industrious – impulsive – critical – stubborn – envious
Ben is envious – stubborn – critical – impulsive – industrious – intelligent

The first word anchors our opinion and the rest does not matter as much as the first description. Although the descriptions are equal we think Alan is intelligent and Ben is envious.

1.16.3 Information Interpretation

Another thrilling concept is asymmetric dominance. Dan Ariely did some research and stumbled upon an offer by the Economist (Ariely 2009). He tested this on a group of smart people (200 Harvard Students) and asked which subscription they would want.

Economist online subscription	$59
Economist print subscription	$125
Economist online and print subscription	$125

Eighty-four percent of the people chose the online & print subscription. Nobody choose the print subscription. But when he asked again and deleted that option only 32% choose the online and print option. So the option nobody wanted anyway boosted sales by more than 50 points. This shows us that the way options are presented can help or manipulate our choices.

The context in which we gather and interpret information has great impact on our decisions. Dan Ariely did a series of tests in which the subject got himself in a state of physical arousal. This had great impact on the way they answered questions about sex. This is the same when you do your groceries when you are hungry or not. In the first case you will buy more and less healthy groceries. Your mental state influences your choices and makes them less rational.

Another powerful concept is association. Our unconscious does all the work here and with the speed of light we associate one item with the other. The exact process is unique for all of us, because the associations are formed by our past experiences. Our brain consists of millions and millions of brain cells that are connected and form a network, similar to the Internet for example. This network is setup so a connection can be lost but the network keeps functioning the way it is supposed to be. These connections enable us to reach other ideas quickly independent of the starting point. In our brain the fast and furious system it quickly navigates this network based upon the first impression. It helps us understand the impression fast and experience based. This concept is used to remember large numbers and it is called a metal walk. For instance, Sherlock Holmes stored pieces of evidence in a ‘mind palace’ to remember everything.

1.16.4 Speed

Our brain is connected to a series of input ‘devices’ like our eyes, ears, mouth, nose and fingers. According to Zimmermann our sight is the fastest by far (Zimmermann

1989). It processes input at a 10 mbit/s, while hearing travels at 100,000 bit/s. This means we can 'read' a 100 times more than we can hear in the same time. But there is a big difference in speed between normal reading and watching an image. During normal reading people sub-vocalize the text, meaning they are actually reading the text out loud and hearing the content. This means normal reading is on a speeding limit of 100,000 bit/s because it is actually hearing! This is why speed-reading claims to be faster by shutting down the sub-vocalization of the text. Viewing and interpreting images is travelling at sight speeds of 10 mbit/s. The recent trends of infographics and data visualization are trying to leverage the speed of sight (and association). Data artists are making beautiful and meaningful visualizations that enable people to view big stories in a blink of the eye. These data artists understand how to tell a story by evoking sight speeds combined with association. The infographic trend is struggling to use the speed of sight. More and more infographics are using too much text and combining more and more images to elaborate multiple issues.

The downside of speed in normal reading is the loss of comprehension. The faster a person reads the less he will recall. For information processing and decision-making this is an issue because if a person does not comprehend the text and makes the wrong decision the text has failed. Using visualization can ensure speed and comprehension, because association can amplify comprehension.

Concluding the brain can help us to process information and make decisions if we are aware. If we are not aware our brains will make its own irrational decisions. If we are aware we can use the power of the brain to produce smarter information and enable effective and efficient processing and decision-making.

1.17 Enterprise Information Management Bottom-Up

This chapter gives a bottom-up overview of aspects of Enterprise Information Management. This chapter is intended for those that are familiar with specific domains within information management.

EIM offers a point of view for the managing and generating of decision supporting information at corporate (company-wide) level (Fig. 1.7).

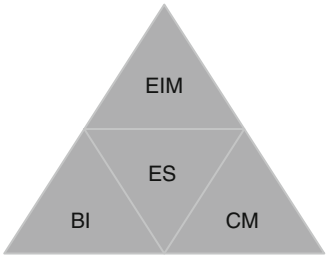


Fig. 1.7 EIM field

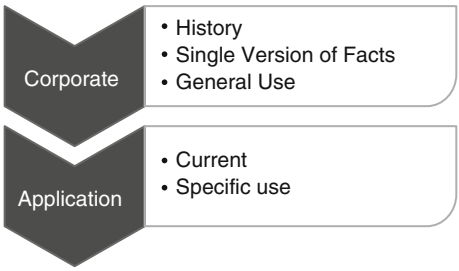


Fig. 1.8 Integral information

Decision supporting in this case means that the information is used to make strategic, tactic and/or operational decisions. The fact that it is company-wide entails that the meaning of information has to be the same for everybody within the entire organization.

EIM supplies synergy between structured and unstructured information, between point of views and implementations, between known and potential information needs. EIM's implementation is based on business intelligence (BI) and enterprise content management (ECM). Enterprise search and retrieval (ESR) is the unifying factor of both fields. Critical (business) information (corporate information), in other words, information that is vital for the organization's survival, is integrally available, and if not, it should be. This information can be used to create a clear picture of the subject under investigation. Corporate information is created by the refining of application data, refining in this case meaning that the application data is fitted to be integrated and combined with data from other applications.

The operational process is the source of all information within an organization. Application data originates as the product of the operational process on the one hand, and on the other as the product of the description of products and services. The running of an operational process can be compared to the running of a transaction. This transaction is registered and the necessary data is recorded in the information system. This is known as structured information (Fig. 1.8).

The products and services a company provides are described in documents such as emails, pictures, presentations, manuals, flyers, articles, press-releases, brochures, websites etc. These services are realized without transactions, and do not have a pre-defined form or content. This is unstructured information.

Transaction processing systems (online transaction processing systems or OLTP-systems for short) are used to record these different types of information, as well as enterprise content management systems (Fig. 1.9).

We summarized the recording of content (documents or text) as online content processing (OLCP) as a counterpart for the transaction processing (OLTP). OLTP-systems are the basis for structured data. The OLCP-systems are the basis for the (semi) unstructured data.

All the sections mentioned contribute to the possibilities of enterprise information management (EIM) and to possible synergy of EIM-related systems within an organization. EIM uses insight, information and knowledge about the management of information and data to support the organization in reaching its goals. These goals are derived from the organization's vision and strategies.

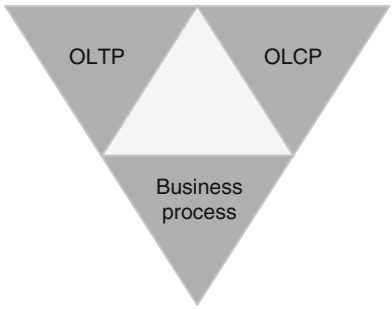


Fig. 1.9 Information domains: business process

1.17.1 EIM Is Insight, Information and Knowledge

Scorecards are used to depict an organization's goals. The four factors that determine the success of an organization are divided into four quadrants: client, process, innovation, finances. The so-called key performance indicators (KPI's) are used to measure the factors of success. These KPI generate information requests for the entire organization. Information components, such as reports, dashboards, OLAP and data mining applications are created to determine these information requests or information needs, and so that they can provide information. This provision of information based on structured information is enabled by OLTP. The goals of an organization also increase the demand for documents and web content. Communication about an organization's products and services has become a necessity, just as it is necessary to know how a client responds to certain events. This is known as content requirement, which can be translated into content elements that supply the content provision. Content intelligence is formed based on this content provision, which is knowledge of the most important entities in the organization and its expressions directed at the environment (Fig. 1.10).

The combination of business intelligence and content intelligence offers new information possibilities. Not only does it produce the combined documents and articles, but also a joint index of both intelligences, based on the topics found. Unstructured information now has the structure of the structured data collection at its disposal, and can be turned into semi-structured information. All data, so transaction data as well as content and meta-data (data that describe content as title, author, keyword etc.) are recorded in the enterprise information environment (EIE), which forms the heart of the EIM integration model (Fig. 1.11).

Document and content management: content demand, content supply and content intelligence.

The EIM integration model consists of three layers; a business layer (outer ring), an application layer (middle ring) and a data layer (inner ring). The business creates source data by using the transaction supporting systems in the daily business process. These systems also supply information that can be used during the execution of transactions. Apart from the transaction data, the organization also produces and processes content in the content management environment. This environment

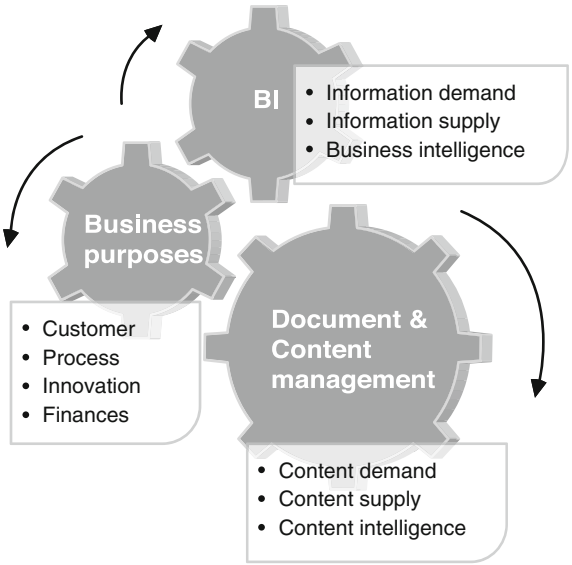


Fig. 1.10 EIM sections

Fig. 1.11 EIM integration model

supplies the information for the daily content management activities. It is necessary to combine the two types of data (transaction data and content) to make integral information more available.

The enterprise information environment (EIE) uses both environments as a source, and functions as a joint environment for the information demand in the business intelligence environment and the content management environment. The EIE creates history with regards to both types of data. New information becomes available through the joint data storage, and its goal is to gather new information and increase innovation or the competitive position. Structured information from the BI-environment can be combined with the content from the ECM-environment and vice versa, by means of storage in an EIE. Apart from this combination, it is also possible to look for unknown information by means of the enterprise search. In the search results of a search, business intelligence reports are combined and integrated with elements from the content management environment.

Business intelligence, content management and enterprise search are the three pillars of enterprise information management. We use the following definitions for the three pillars in this book:

Pillar	Definition
Business intelligence	Business intelligence is the actual process, with corresponding facilities to gather structured data, to analyze the data and to use the resulting information
Enterprise content management	Enterprise content management deals with the company-wide management of unstructured information, and comprises of document management, digital asset management, records management, web content management, collaboration and imaging.
Enterprise search	Enterprise search offers the possibility to use electronic tools to retrieve data or information from many different source systems, such as databases, file-systems, legacy systems, document management systems etc. within one organization, via one standard, central search interface and one or more indexes

Each pillar consists of the general aspects of the fields, such as:

- Information architecture,
- Data development,
- Database management,
- Security management,
- Reference & master data management,
- DW & BI management,
- Document & content management,
- Metadata management,
- Data quality management.

The EIM-field deals with the integration of structured and unstructured data and information. Business intelligence is the field for the structured data; content management deals with the unstructured data. The types of data are brought together by enterprise search (Fig. 1.12).

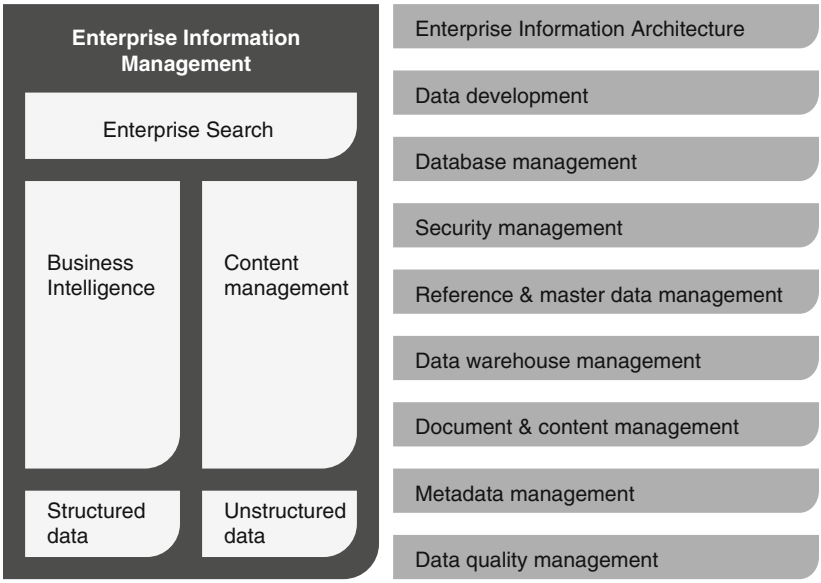


Fig. 1.12 EIM field (structured data/unstructured data)

The design of the previous book (Baan et al. 2010) was based on the more or less fixed sequence of activities within EIM that is displayed with terms like Collect, Connect, Compose and Consume. This sequence is released for the purpose of this book. This book is primary pushing on the blocking factors (access, available, Relevance and interpretability). The activities are found in Chap. 2 in the form of information (value) chain.

Both knowledge and ability, and the willingness to use these new concepts influence the end result. To manage this properly an enterprise information competence center (EICC) can be used. This is either a physical or virtual part of the organization that on the one hand arouses awareness of the new possibilities within the organization, and on the other hand provides norms and guidelines in order to be able to work as efficiently as possible. This is explained in greater detail in Chap. 6.

Chapter 2
The Importance of an Enterprise Information Management Strategy

John Septer

All men can see these tactics whereby I conquer, but what none can see is the strategy out of which victory is evolved.

Sun Tzu (Chinese general, military strategist, c. sixth century BC).

In today's world it is difficult for organizations to keep up with the continuously changing market demands. The consumer is an individual and wants a personalized service. Customer loyalty is much harder to keep. This creates a tremendous challenge for organizations to keep existing customers and attract new ones. Developments in social media, the exponential growth of data, ever-changing laws and regulations, and the enormous amount of touch points (channels) are examples of the continuously changing environment for organizations. Organizations must make choices and not jump on every new development without first mapping the consequences. This requires leadership and the ability to make choices and to underpin these choices. These choices and the underpinning are part of an organization's strategy (Fig. 2.1).

The funnel model gives insight into this complex environment. Each organization must deal with this every day, month and year. As mentioned previously, the world is changing and changing fast, making it increasingly harder to stay ahead of the competition. Organizations need to focus on added value for the customer. With this in mind they should orchestrate every activity, asset, value, or resource within the organization to keep up. This is very hard to do. The challenge is to create a stable strategy to which every aspect within the organization can relate.

In addition to all the external threats such as innovations, new or changing technologies, competition, and so on, there also are a number of internal forces at work. These forces, like money, politics, resources, and time, compel organizations having to make choices. This is also related to the fact that changes have a cost.

J. Septer (✉)
Advisory Services, Incentro, De Meern, The Netherlands
e-mail: john.septer@incentro.com

P. Baan (ed.), *Enterprise Information Management: When Information Becomes Inspiration*, Management for Professionals 2, DOI 10.1007/978-1-4614-5236-2_2, © Incentro 2013

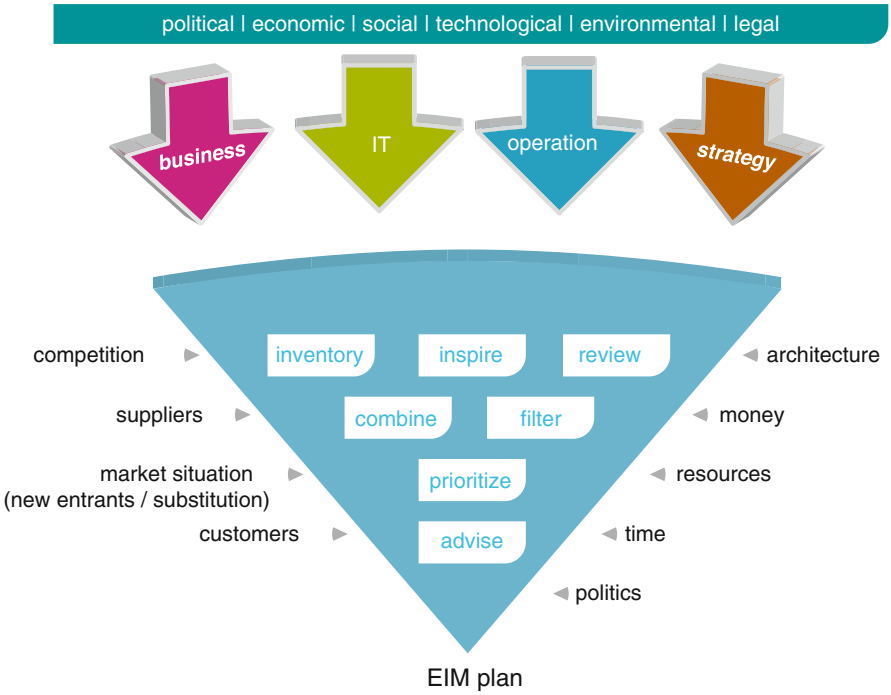
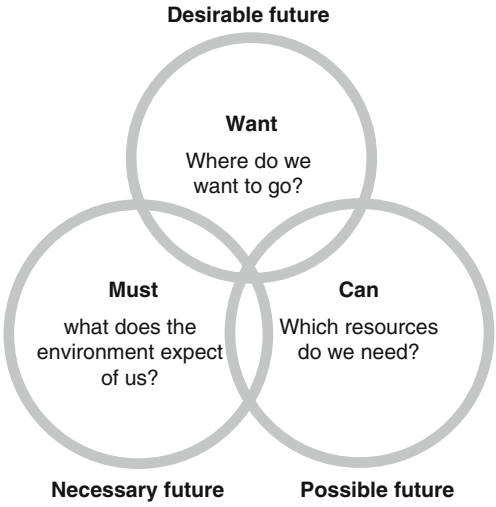


Fig. 2.1 Funnel model

Fig. 2.2 Want–must–can model



Some changes are necessary, because of regulations, and some are needed to keep up with the current competition or to create a unique position. Striking a balance between what an organization wants, what it must do, and what it can do is essential for success (Fig. 2.2).

To make the right choices, it is necessary to weigh all factors (both external and internal). By doing this, you are able to get a clear understanding of what is needed – and when, because you cannot do everything at the same time as an organization. The main objective is to fulfill the greater good (strategy), and on that basis a distinctive competitive advantage can be achieved. The different activities that must be undertaken to make the right choices for an organization are identify, judge, combine, filter, prioritize, and advise. The choices made are captured in a plan, the enterprise information management (EIM) plan.

Identify	Perform an analysis of the organization according to the need, want, can model. What does the organization want? What are the ambitions (needs and requirements) of the business? What should the organization do, from the standpoint of policy or external laws and regulations? And finally, what are the organization's capabilities (can)? The *can* represents the physical conditions (technology), resources, time, and finances
Judge	Based on the analysis, the state of the organization can be assessed. The need, want, can model gives insight into the degree of success of the organization. If one or two of the three elements are mostly lacking, it makes little sense to continue. It is then important to find out what is needed to make the weaker elements stronger. All this relates to the maturity level of the organization
	Furthermore, the analysis gives insight into the interdependencies of the various stakeholders. It can be used to estimate the extent to which a stakeholder needs, wants, and can. The need, want, can analysis may indicate that critical stakeholders can meet and strengthen each other
Combine	Try to combine the different needs, wants, wishes, and necessary changes as much as possible to serve multiple stakeholders. This provides more support within the organization and prevents things from being done twice. Essential within this process is communication. Combining the needs, wants, wishes and changes will lead to compromise. These compromises will not always represent all the desired results of a particular stakeholder (in the short term) but will provide added value to the organization and the stakeholders in the long run
Filter	Filtering the actual need to get to the next level of maturity in information management. Define the various projects/initiatives in the different areas (people, process, technology, and information) of the organization to get to the next level. Initiatives necessary to improve a weak element will be given priority. This approach is necessary if the organization wishes to eventually achieve its higher objectives
Prioritize	Then prioritize according to business needs. This will allow information management to target the most urgent business needs and issues. These in turn are derived from the overall business strategy and direction for the organization as a whole. Always keep in mind the need to deliver tangible and visible benefits
Advise	There is no single application or project that will address and resolve all the information management problems of an organization. The answer is to let go of the desire for a perfectly planned approach. Instead, the journey is more important than the destination. This approach recognizes that hundreds of small changes are often needed to improve the information management practices across an organization. These changes will and can often be implemented in parallel or can only be implemented sequentially. With a consideration of all the different forces, both internal and external, and the corporate strategy the right choices can be made in planning the different initiatives and projects over time. What is adding value for the here and now? What is adding value for the future? What is needed as a foundation to achieve added value? These are all questions that will be addressed during the advice activity

2.1 Challenges

Besides the industry dynamics in which an organization finds itself, it is crucial for organizations to get a grip on the management of internal information . The ability to apply and use correct and timely information is a major issue for many organizations. Examples of problems currently facing organizations include the following:

The amount of information generated by organizations is growing exponentially;
Current resources (applications, systems, infrastructure) are inadequate for managing the tremendous growth in information;
It is becoming increasingly difficult for organizations to determine the relevance of information, despite enormous investments in information technology;
The outside world is changing and requires that organizations, regardless of which channel whatsoever, provide the same information at the time the outside world demands;
Current ways of working require increasing levels of transparency and reinforce the need to share information between departments, suppliers, customers, and partners;
Incompletely documented end-to-end processes within organizations lead to inefficiencies and unnecessary delays, which ultimately have a negative impact on the time to market of an organization;
The lack of clear governance policies and guidelines result in a proliferation of applications and inconsistencies in information models;
There is no "single version of the truth; it is impossible to know if the "right" version of the data is being used. This issue is compounded by the potential for duplication and inconsistencies;
Knowledge workers within organizations waste a lot of time either looking for or getting access to information or implementing workarounds of the information they need.

Given these issues, it is essential that organizations understand the importance of information management. Information management refers to the storage, management, and provision of data and information within an organization. When an organization understands its information, it can work to improve information management capabilities. This gives the organization control over the current situation and over further improvements in its information management policies. In order means providing the right information, to the right people (internal/external), at the right time, at the right cost. Only then can an organization achieve its "higher" objectives.

2.2 What Is Strategy?

A simple question with many different possible answers. A good definition with regard to strategy is given in the book by Kenneth R. Andrews *The Concept of Corporate Strategy* (1971):

> "Determining the goals of the organization and the set of coherent choices regarding the allocation of resources and activities to realize the goals."

Porter (1996) gives the following definition of strategy:

"Strategy is creating fit among an organization's activities"

Richard Rumelt[1] states that a strategy is in essence simple. Business strategy is about the focus on resources, and a good strategy focuses multiple resources on a single objective.

The development of a strategy consists in the integration of activities, functions, and resources across an organization. Many people confuse strategy with operational effectiveness. Operational effectiveness is about achieving excellence in individual activities, functions, or resources; strategy is about combining these (Porter 1996).

Let us focus on resources. A resource can be seen as an asset or capability within anorganization. Every organization is built up on a very different set of tangible and intangible assets and capabilities. No two organizations are the same, because no two organizations have have the same internal set of capabilities or assets. The combination and integration of the different resources in support of the overall strategy determine an organization's success (Collis and Montgomery 1995).

A resource can be either tangible or intangible. Thus we can say that information is an intangible asset for an organization that represents a certain but unidentified value.

It is clearly very important than an organization think about what it wants to achieve and how these goals are to be achieved based on the current set of resources and subactivities. A strategy provides insight and answers to the following questions:

What are the vision and ambition of the organization for the long term?
What is its unique position within the market, branch, or industry?
Which products and services does the organization want to bring to market?

- Which markets (segments) does the organization target and how?
- What current and future developments in the market influence the functioning of the organization?

You could say that strategy is calculated behavior in a changing environment. Calculated behavior indicates that an organization has thought about its ability (capabilities) to act, combined with ambition. The complexity in defining the right strategy derives from the fact that an organization is made up of different parts. Consider the finance department, the marketing department, Information & Communication Technologies (ICT) department, human resources, and production. All these departments exist and make their contribution to the whole, and all have their own strategy. The difficulty for the organization is to align all these different strategies within the organization in order to achieve the higher defined business goals, which form part of the corporate strategy.

Having a strategy is the first thing. What needs to happen to fulfill the defined strategy is the second. *Begin small, think big*. An organization cannot do everything at the same time. The fine-tuning of the different initiatives is of great importance. Precisely on this point there is often friction within an organization. As mentioned, departments have their own strategies and find that the defined plans and activities

[1]Richard Rumelt is Harry and Elsa Kunin Chair in Business and Society at the University of California-Los Angeles.

should have the highest priority. Alignment and risk reduction are needed here. When executing the strategy it is good to break up the planned implementation of the plans and initiatives into short-, medium-, and long-term time boxes.

Essential in having a strategy is being unique in an industry. Nowadays we see that many companies have operational excellence[2] as a strategy. This is not a sustainable strategy because operational excellence is becoming the norm, rather than the exception, in industry. For the other two value strategies (customer intimacy[3] and product leadership[4]), from Treacy and Wiersma (1995), the same thing applies. However, delivering customer value is very important and should be the main driver for an organization in defining the strategy.

Making choices in defining a strategy is very important. Focus and choices determine the structure (business model) of the organization. The business model of the organization and the strategy should be consistent with each other. The overall business strategy is to steer the other strategies within the organization.

Without a strategy an organization is rudderless and will not survive long in today's changing and demanding world. *Without a strategy, no sustainable entrepreneurship is possible.*

2.3 EIM Strategy: What Is It?

An EIM strategy can be defined as an organization's unified blueprint for *capturing*, *integrating*, *processing*, *delivering*, and *presenting* information in a clean, consistent, and timely manner. An information strategy supports the goals and strategies of the business, as part of the overall business strategy, with a clear focus on information as asset.

2.4 Information Management

Information management deals on the strategic and tactical levels with the provisioning of information within the organization.

[2] Operational excellence: Strategy characterized by a high degree of standardization and tight, uniform control of processes. The starting point is a quick, unambiguous delivery of products and services to customers at the lowest cost and least time. Focus on efficiency, cost, and time savings.

[3] Customer intimacy: The wishes of the customer come first. Focus on loyalty and customization.

[4] Product leadership: Flexibility, adaptability, initiative, and individual thinking are central to this strategy. The focus is both on innovation and on market conquest.

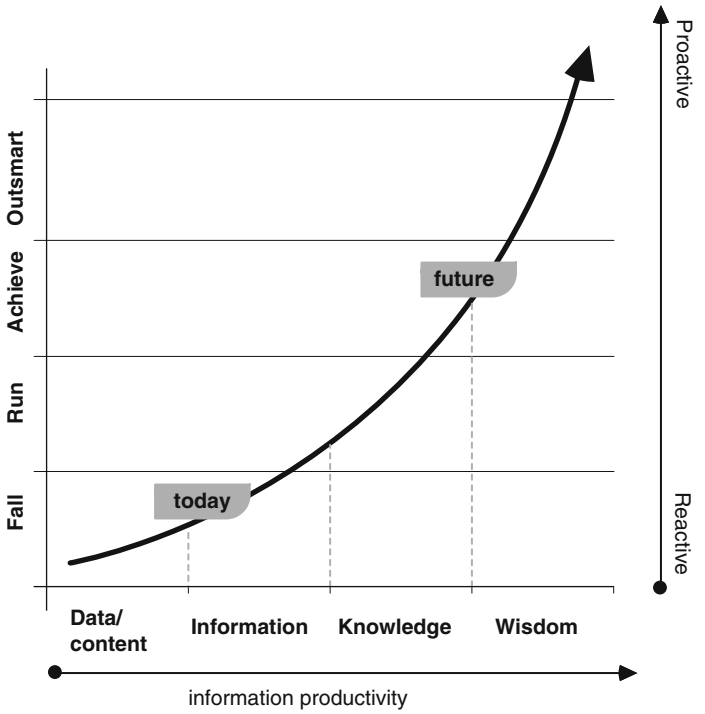


Fig. 2.3 Necessary journey for organizations

The driving force in information management is to provide information. By provisioning of information we mean the set of people, resources, and actions of an organization, where one focuses specifically on the information needs of the organization. The information needs of an organization can be divided into three categories.

Operational information:
All the information needed to perform the day-to-day work within an organization

Control information:
Specific information needed for managing an organization

Accountability information:
Information on how operation and control proceed within an organization

Progressive companies are those companies that are able to put information at the heart of how they do business. Organizations that do not adopt the new way of thinking will find themselves becoming increasingly less competitive. For most organizations this will necessitate a complex journey (Fig. 2.3).

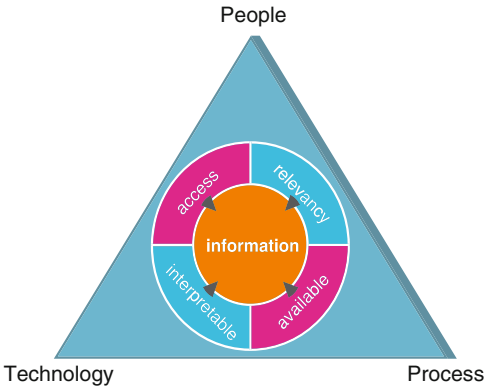


Fig. 2.4 EIM triangle

This complex journey results in the need for change within the organization. The change is complex because it can only be achieved by a combination of actions. Each organization needs to plan its own personal journey in detail to ensure that it is moving forward in the right direction.

But how will you know as an organization that you have been successful? There are several signs that indicate success for organizations:

Information-based decision making becomes a core competency of the organization.
A strong focus is developed on driving a step change in business outcomes through effective exploitation of information, often in conjunction with customers, partners, and suppliers, who are able to act proactively on changes in the outside world.
The right, relevant information is provided on time through different channels to support the customer journey.
Information is delivered consistently across the organization, i.e., asking for the same information in different divisions yields the same result, and users and applications do not have to wait long to get their requested information.

The starting point for developing and implementing an information strategy for a specific organization is composed of a number of primary focus areas, dimensions that will give insight and answers on how to move forward (journey) (Fig. 2.4). These are:

Information
Human, Culture and organization
Process
Technology

2.5 Motives for an EIM Strategy

Several reasons exist for implementing an information strategy within organizations. Peter Hinssen (2010) conducted a small study in his book *The New Normal: The Revolution Has Begun* to identify the reasons for implementing an information

strategy. Based on this study Hinssen distinguishes six motives for implementing an information strategy. These six motives are then divided into two groups: defensive and offensive.

Defensive	Offensive
Compliance	*Speed*
Because of rules and regulations, certain documents must be archived, and there are strict rules about who gets access to what information and who can modify the information	Quick access and distribution of information throughout the organization are key
Control	*Share*
The quality of information is essential; the whole cycle of creation, approval, publication, and archiving is closely monitored	No silos in which information is locked, but creating value by working together to create and share information
Archiving	*Intelligence*
The emphasis lies on "lose nothing"; all documents are stored for easy retrieval	Accumulated knowledge and cocreation are the keywords

It is clear that the defensive motives have a strong internal character. They are aimed at solving current problems within the organization in the field of information management.

The offensive motives are much more progressive in nature. Of course, it is necessary to solve current problems. But matching the information strategy with the business strategy also requires having a vision. It requires more than solving current problems. Addressing the defensive motives can be a sign that something is wrong in the internal information provisioning. The defensive motives block the progressive motives within an organization. You can see the defensive motives as part of the foundation. The foundation is the precondition for excelling as an organization.

Most of the time the motive to start formulating and implementing an information strategy is process optimization. Based on the current processes within an organization you can see where there are problems regarding the information provisioning. From these areas you can investigate what the impact is on the overall processes and what the improvement could be if the information provisioning were optimized.

Another reason an information strategy is needed in an organization is the fact that nowadays there is too much information to manage loosely. We can speak of an information overload. It is clear that an organization needs a plan, a strategy on how to acquire, capture, store, govern, and deliver all information to internal and external users of this information.

The overall goal for developing information is to be consistent with the corporate strategy. In the end every aspect of the EIM strategy should support the vision and ambitions of the organization.

2.6 Defining an EIM Strategy Is Complex

Establish an information strategy based on an evaluation of the business strategy.

Defining an information management strategy is complex because most organizations do not know what their current maturity level on information management is. In addition to that, where do you start as an organization? Organizations are very complex environments. There are many challenges that need to be overcome (both external and internal). Successful information management is underpinned by strong leadership that defines a clear direction.

To understand the importance of an information strategy, imagine the following "what if" scenarios:

What if you knew where all of your corporate information resided?
What if everyone in your organization understood how their efforts contributed to meeting organization goals and objectives?
What if your organization used its management information to actively steer the organization forward, not just for historical reporting?
What if you could access the right information at any time to support decision making?
What if you could predict the outcome or leverage opportunities based on trends in your business?
What if you could achieve a marked improvement in collaborative working with partners, suppliers, and customers?
What if you knew that you were fully compliant with all regulatory requirements?
What if you could trust that all your information was accurate, complete, relevant, and up to date so that you could rely on it when you made business decisions?

Based on these questions we can see that setting up an information strategy can be very complex, but to be able to manage information in the right way, it is essential that an organization develop a strategy. It starts with understanding where the organization is today regarding information and where it wants or needs to be in the near future. This all relates to the maturity level of supporting and managing information within the organization.

It is very difficult for organizations to determine their end goals and ambitions ("To Be" vision) and to prioritize between the different initiatives related to information. Do you start as an organization with data quality, content management, or information governance? It is impossible to do everything at the same time. Where do you start?

2.7 Enterprise Information Maturity Model

The EIM maturity model was created to help organizations increase the return on information (ROI). The model does not assume that you can talk in absolute terms about ROI.

The EIM maturity model aims to assist organizations in identifying areas of processes or information domains where the difference with the competition can be made.

The model allows organizations to identify which kind of investments in which information domains need to be done.

The model recognizes four different stages or maturity levels of ROI. The lowest level of maturity is *fail*, which simply means there is insufficient information available for certain processes to run smoothly.

The next higher level means that the information for a specific area is just enough to get the associated process to be executed satisfactorily. This level is called *run*. Some versions of the model also cover a state called *comply*. This state is mainly interesting for organizations with strong pressure on complying with internal or external rules and regulations.

The next level of maturity is called *achieve*. The focus of this level of maturity is efficiency. This level is about executing primary and important processes as quickly and accurately as possible, making optimum use of available information. Business process management (BPM) and operational business intelligence (OBI) are common terms used in connection with this level of maturity. Quick decision making is important here.

The highest maturity level is called *outsmart*. This level is about using the same information as your competitor and, perhaps, making smarter decisions. Organizations that have primary processes on the outsmart level establish a strong competitive position. This level is not only about having access to the right information at the right time, but it also reflects that the behavior of knowledge or information workers is positively influenced. This means that there is room to include other than purely rational ways to deal with information in order to achieve better understanding.

The first two levels of maturity are not very surprising. Here the focus for organizations lies purely on survival, and organizations are simply incapable of showing distinguishing behavior over their peers. However, it gives much insight into areas an organization needs to work on before it is able to act proactively.

The top two levels are for two reasons much more interesting. First, these are the levels at which an organization can stand out from the competition (short and long term). Second, operating on these levels requires continuous monitoring of the environment to adapt to the ever-changing situation. To continue to operate at these levels, it is necessary to implement a learning cycle. This learning cycle is aimed at continuously identifying changes in market requirements and in the organization and being able to act on these changes.

Determine as an organization your maturity level with respect to information management, and define your goals and ambitions. Then take into account the different factors and influences in today's world. Based on that you can define what is needed for you as an organization to grow to the next level of maturity and achieve a competitive advantage. The easiest way to prevent capital failures is to always reason from the business objectives.

Outside the four levels of maturity are two aspects of concern applicable to the entire model. These are information management and information value.

Aspect	Meaning
Information management	Information management is concerned at the management level with the question of how an organization can effectively and efficiently satisfy the need for information on an ongoing basis
Information value	The added value of information for the organization

These two aspects are the assessment criteria (benchmark) for a particular maturity level in the model.

Maturity level	Information management	Information value
Outsmart	Information is an asset and organizations are constantly looking for opportunities to do more with this asset. New opportunities are used to better meet the information needs of tomorrow. New opportunities are applied to retrieve more information from existing information sources	Data and content are used to improve customer service. Workers know what customers want, what they find important. Services and products are based on the available information adapted to customer needs to bind the customer more closely to the organization
Achieve	Data quality is an important issue at this stage because this influences the effectiveness and efficiency of processes. Data stewards are appointed to oversee the improvement of quality	Data and content are used throughout the organization and deal with uniformity and efficient processes. After all, data and content are recorded only once. There is no discussion about the meaning of information. The metadata model is well known and accepted. Information is shared so that benchmarking is possible between business units. This increases efficiency
Run	The organization has mapped the information flows. It has a view on duplicate processes (for example, multiple captures of customer data) and how the information is used and shaped during the process	Information has been given a formal meaning within the organization. This example leads to less discussion about the meaning of terms that are used. The reports are ambiguous. The term *key performance indicator* (KPI) makes its careful entrance
Fail	The organization focuses only on daily business processes without seeing these processes in conjunction with the rest of the organization There is no question of information management. Everything is arranged in a decentralized way, and the necessary information to justify the works is managed manually	The organization considers information as a burden and is busy with cramming the daily processes into existing systems without worrying about the consequences of workarounds

2.8 Key Elements for Success

The following elements are key to successfully developing and implementing an information strategy within an organization.

2.8.1 Communication

Make clear that what needs to be done is in line with the overall business strategy. Communication is extremely important. Communicate, communicate, and communicate. Departments within enterprises can be seen as so-called islands, in which the people are very committed to the interests of the department. Misunderstanding often emerges when the organization gives priority to other plans. It is therefore important to keep communicating. In addition, it should be made clear that the execution of a strategy is a *marathon rather than a sprint*. It is a long-term vision, an ambition that the organization pursues. To confirm the success of strategic implementation, it is very important that KPIs be defined and actually measured and evaluated.

2.8.2 Leadership

Showing leadership within the organization is key. Upper management should have a vision (regarding information as an organizational asset), be able to inspire others within the organization, and take accountability for that vision. Strong leadership is all about making choices. Porter (1996) indicates that leadership is more than the stewardship of individual functions. It is all about communicating the organization's unique position, making tradeoffs, and forging synergy among activities.

2.8.3 Commitment

Commitment is very important in building an information strategy. The information strategy is owned by the entire organization, not the happy few. It is very important to define benefits for all the different stakeholders throughout the organization. Benefits differ among stakeholders, and it is very important to be aware of that. Everyone should understand "What's in it for me." Only then will there be commitment throughout the entire organization.

2.8.4 Execution Discipline

It is important that a strategy be executed. Execution discipline is the key success factor for optimal implementation or execution of strategy. In addition, one of the most important preconditions is commitment. Everyone must agree with the path and the principles set. The mapping of the stakeholders, both inside and outside the organization, provides insight into the forces that influence the optimal execution of the strategy. Matching the various plans/strategies is therefore one of the biggest challenges for companies.

2.8.5 Governance

Governance is all about encouragement of the desired behavior in relation to information; tt is not about coercion. Influencing the behavior of people to do the right thing for the organization is key to achieving the goals of the organization. If people want to act differently, that will affect the success of information management. People should feel and see the added value of acting according to policies and guidelines and they should realize that this acting does not always have a direct benefit.

2.8.6 Change Management

All ingredients that are preconditions for success are implicitly part of change management. The involvement of the organization – explaining why, clarifying the course, and making the message clear –results in greater understanding by workers.

2.8.7 Added Value of an Information Management Strategy

The added value of performing an information management strategy is significant. It gives great insight into where you are as an organization, and it determines what path you must take as an organization and what you need to do to remain successful. Following are some examples of added value for your organization:

- Your business processes will better perform with improved information provisioning.
- The return on ICT investment will increase.
- Risk of failure of ICT projects is minimized.
- Relevant and right information is delivered at the right time to improve decision making.
- Possible future outcomes will be more predictable and how to act on them will become more clear.

- Information provisioning responds more quickly to market changes and changing laws and regulations.
- There will be a stronger focus on what is good for the organization instead of the best solution for select individuals or departments.
- The primary focus will be on the information problem area and its impact on the business and possible solutions; technology will become the secondary focus.
- Information will be used in a consistent manner within the organization and the overall framework of organization objectives.

2.9 What Is Information?

But what is information? There is no single comprehensive definition of information. A number of similar concepts surround the term *information*, such as message or data. Everyone has an intuitive notion of the term. Information is an abstract concept, subject to interpretation, consistency, an understanding of the situation, and context (Fig. 2.5).

Information is compiled by the interpretation of two raw materials: data and content. Data are facts or things such as text, numbers, and images without meaning or context. (The context is the overall environment in which anything acquires meaning.)

Content refers to unstructured information. Information is obtained through the interpretation of data and content. The reason for interpretation is to give meaning or value to data and content. The integration of information into existing business processes leads to knowledge, and knowledge is the use of information. Weggeman

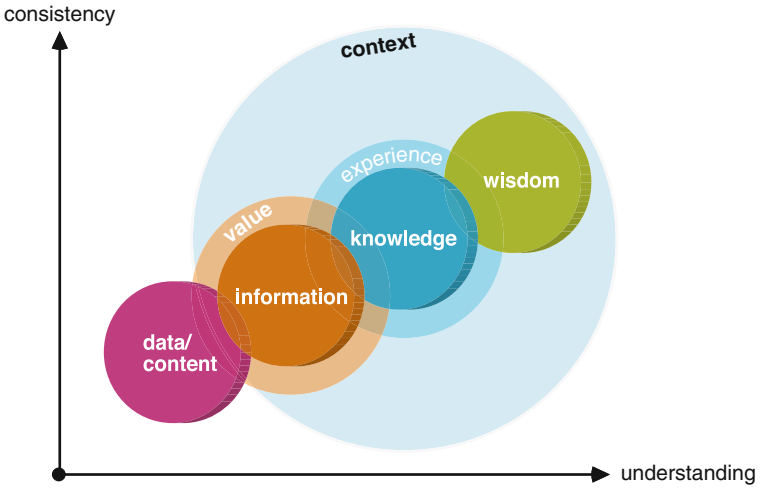


Fig. 2.5 Characteristics of information

(2000) describes knowledge as "partly unconscious – the capacity that enables someone to perform a certain task. A power that is a metaphorical function of the Information, Experience, Skills and Attitude that someone has at some point in time: K = f (I.ESA)."

If employees apply knowledge within their day-to-day work, this will lead to (positive) change in behavior. Ultimately, applying knowledge leads to wisdom (intelligence).

> We are drowning in information but starving for Knowledge (John Nesbitt 1990).

Wisdom is more than just knowledge; it has to do with insights about life, moral responsibility (which is good and bad) about our self-knowledge, and mastery of our passions and pitfalls (as individuals and as a species). Wisdom also has to do with living, perhaps coming to a collective consciousness, harmony with others, and with the environment (nature) around us. For that we need not only our heads but also our hearts and stomachs. Much of our wisdom lies in the structure of our society, government (democracy, respect for minority opinions, care for the weak), the law (justice rather than retribution), and scripture and other sacred works.

In his book Het informatieparadijs, Guus Pijpers (2011) indicates that information is all around us. Sometimes it is sensible (a wall, a pen, someone's arm) via the senses through which we acquire information: smell, sound, vision (what you see), touch, taste. Not everything around us is intended to provide you with information; it depends on the person and the situation. Most people see or associate information with something tangible, such as paper, a CD, or photographs. These are essentially the media and not information. Information is part of what these carriers convey.

An apt analogy is the comparison of information within an organization with water. Nowadays we are used to the fact that we have access to clear, clean water whenever we need it. We use water for many different purposes within our households: water for showering, cooking, drinking, washing the car, or spraying plants. It has become a common commodity. We believe that we have access to clean water at a moment's notice. Therefore, the water department is responsible for providing us with water.

The same goes for information. We want to have access to right (clean), relevant information at the right time. Therefore, we need a clear strategy in relation to information that will allow for acquiring, creating, cleansing, enriching, and integrating information. Only in that way can we trust in the fact that we will have access to the right information at the right time.

2.10 Characteristics of Information

Information has a number of characteristics. It represents value, is volatile, can be shared, is time dependent, and can be enriched.

2.10.1 Information Has Value

Information *can* have value within a given context. It will have greater value when it is shared, combined, and analyzed. The value of information depends on its use or purpose. The value of information changes over time and depends on the context and experience level of the receiver. You might say that information can be perishable, because it diffuses rapidly or becomes outdated (Porter 2008). In addition, the value of information depends on other pieces of information. Within a given context, information can typically be quantified.

2.10.2 Information Is Volatile

Information has a volatile character. People forget information relatively quickly (consciously or unconsciously) and interpret information personally (depending on background, situation, experience). Information is subjective. By capturing information (record), based on context, it is possible to make this information independent of time and place for the purpose of sharing. In this way, information is not forgotten. You can speak of forgotten information when such information is available, but it is not known what role or importance that information has played.

The recording of information can be facilitated on the basis of technological and procedural measures. This is based on making choices. Exponential growth is an enormous threat for organizations that wish to capture information based on a fixed set of principles. What information should be recorded and in what format? For how long and with what technology? In addition, the context in which information is to be recorded should be considered together with the information to make the information accessible over time.

2.10.3 Information Can be Shared

Information, or pieces of information, can be shared. Information can be shared during the execution of business processes or as a result of meeting legal obligations to provide certain information to third parties (such as tax authorities). The fact that information can be shared is important for organizations. Encouraging reuse and sharing of information (knowledge) leads to efficiency and effectiveness and consistency in information provisioning. The culture of the organization or the behavior of people can influence, counteract, or facilitate this. By nature, people tend not to reuse information. They rather recreate the information because they often believe that that is quicker than searching for the information.

2.10.4 The Value of Information Is Time Dependent

The value of information is time dependent. Some information is worthless after some time has passed. For a knowledge worker, it is very important that the appropriate, relevant information be provided as quickly as possible. Information represents a great value if this information is delivered in time. If information is provided too late, then the same information (content) no longer represents the same value.

2.10.5 Information Can be Enriched

To make information retrievable, information can be provided with descriptive metadata. Metadata are data about data. They enrich and add context to information. This enrichment improves the eventual retrievability and relevance of information.

2.11 Information Is an Organizational Resource

Decision making within organizations is based on information. To make decisions, it is essential that information be relevant, accurate, timely, and available. Information can therefore be seen as the most important business asset or resource within organizations because it is key for decision making.

Collis and Montgomery (1995) developed the concept of a resource-based view (RBV). In this model, which can be applied to all kinds of resources within an organization (tangible and intangible), the central question is: What makes a resource valuable? Value is determined on the basis of three criteria: scarcity, appropriateness, and demand. If we look at intangible resource information, we can say that information is everywhere. One can speak of an overload of information. But if we look at the need for right, relevant information at the right time, then we can say that such information is *scarce*.

The extent to which resource information can be *appropriated* determines positive action and decision making, and there is high *demand* for relevant information. All three criteria can be influenced by having the right strategy and combining the various activities (Fig. 2.6).

It is important for organizations to realize that information is an organizational resource and as such needs to be managed. As with any other resource, it has value. The value of information lies in its accessibility and accuracy. Therefore, the development of an information strategy is critical for business success and key to sustaining a competitive edge over time.

The ability to use information as a business resource ultimately has to do with three key elements: the quality, accessibility, and transparency of information.

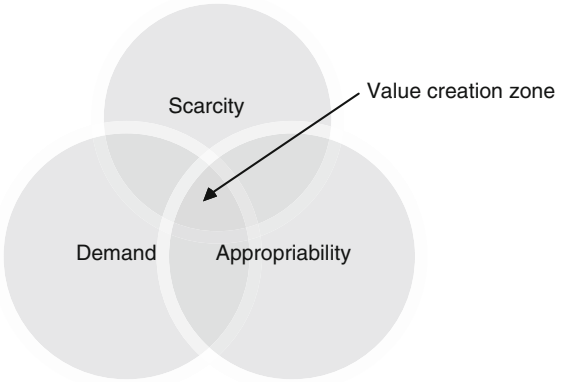


Fig. 2.6 Resource-based view

2.11.1 Quality

Quality of information refers to the contents, metadata, and application or use of information. If an organization is able to guarantee the quality of information, it will be able to make the best use of available information within the organization.

2.11.2 Accessibility

Information within the organization should be retrievable and accessible, either by internal staff or the external customer. Accessibility is all about information integrity and governance.

2.11.3 Transparency

Eventually good information management leads to transparency. Transparent information provides insight into customer information needs and the organization itself, ensures risk management, and provides insight into the information landscape and responsibilities. Ultimately, effective management and control can be facilitated if information is transparent.

On the other hand, the risks increase in relation to resource information when information is not managed in an orderly and consistent way. The risks related to information are as follows:

Noncapture of critical information
Loss of information
Lack of context or descriptive information (relevance)
Damage to information

Unauthorized access to information (control)
Unavailability of information
Loss or misplacement of information
Inaccessibility of information
Lack of ownership of information

2.12 Competitive Advantage

Competitive advantage, whatever its source, can ultimately be attributed to the ownership of a valuable resource that enables organizations to perform activities better or more cheaply than the competition (Collis and Montgomery 1995).

If organizations can excel on the resource information front, they create a competitive advantage over their competitors. Such organizations are able to proactively respond to changing customer needs and market movements. Today's customer is very demanding and not very loyal. The customer wants to communicate through various channels with an organization. For an organization it is important that the information in the various channels be up to date, relevant, and valuable to a particular customer.

In addition, information should be accessible to customers at any time depending on the type of information, whether direct or indirect access to the information is needed, or whether or not an employee needs to intervene. Only then will the customer feel like he or she is receiving individualized treatment. Proactive can also mean that, based on available information from the customer (or customer group), certain future behavior is encouraged. This may entail sending a shopping list, including suggestions, based on historical buying patterns. An innovative organization acts and focuses its services to match actual customer demand. Only then will it stand out from the competition.

Most organizations are only able to respond reactively, simply because the level of information management does not allow for a proactive response and the necessary knowledge based on available information is lacking. When an organization can respond or operate proactively, then information is used to its optimal value and power. Information is then a strategic business resource and used as such.

This also shows the tension between willing (want), need (must), and ability (can). Creating and putting in place the conditions necessary for satisfying customer demands will ultimately ensure that the organization will be able meet the want and the need.

2.13 Information Value Chain

Every organization has an information value chain in which raw data and content are systematically acquired and then transformed and enriched through various stages that add value to that information. The information value chain concept is based on the value chain of Porter (1985).

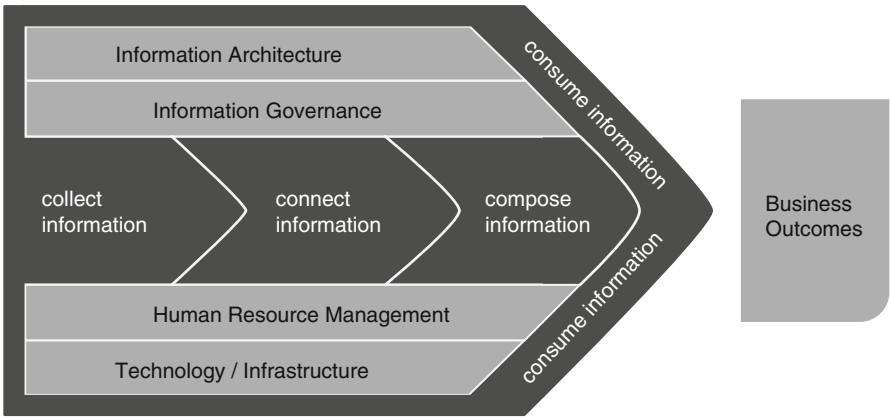


Fig. 2.7 Information value chain

The information value chain identifies the value activities required in the use of information. It represents the cycle of a common information asset, with the activities most likely undertaken by one entity (Fig. 2.7).

The activities can be broken down into two components within the value chain.

- Primary value activities relate to aspects of the chain that represent the core information product. They are collect, connect, compose, and consume.

Collect	This category includes all activities related to the acquisition, creation, enrichment, storage, and management of information. The collect activity is very important for the quality, accessibility, and life cycle management of information. It is responsible for creating the strong foundation on which all demands and requirements rest
Connect	This includes all activities related to making information appropriate by combining, integrating, and linking information from the various sources and systems across the organization. This means making information relevant by adding context
Compose	This group comprises all activities related to assembling, analyzing, interpreting, and navigating information. This activity is responsible for delivering right and relevant information, regardless of its source, in the right format and presentation form
Consume	This includes all activities related to the use of information by knowledge workers within the organization or people outside of it. The focus lies in sharing information, making decisions based on the delivered information, and informing others about it

Supporting value activities relate to aspects of the chain that assist the core information product. They are information architecture, information governance, human resources, and technology/infrastructure

Information architecture	Information architecture provides the framework against which change can be assessed. Information architecture concerns business capabilities and the translation into logical technology components
	It includes information structure, information integration, rules, and information modeling regarding all areas of the information value chain
	Information structure: definition of information entities across the organization; concerns what kind of information entities exist and the specifications of those entities
	Information integration: concerns how information will be exchanged between various systems
	Rules: definition of sources of information; has to do with which application is responsible for the management of what kind of information and holds the single source of truth
	Information modeling: definition of information structure models, metadata model, and taxonomies; these definitions are applicable across the entire organization
	The results of information architecture are as follows:
	Alignment of business and technology via the common language of information
	Definition of an end-to-end approach to information management
	Handling of risk management and mitigation
Information governance	Information governance has to do with encouraging desired behavior in relation to information, behavior in terms of actions like valuation, acquiring, creation, storage, use, archiving, and deletion of information; it includes the full set of policies, procedures, standards, and metrics (which are actively maintained) for all aspects of the management and provisioning of information to ensure effective and efficient use of information set down in the information architecture; they are actively maintained
	The following factors are subject to information governance:
	Accountability: clear definition of roles and responsibilities across the organization in relation to information; accountability concerns ownership.
	Governance organization: organizational structure that is responsible for monitoring governance policies
	Standards and guidelines: standards and guidelines for creating, processing, delivering, and using information within the organization
	Metrics: methods and instruments for measuring the effectiveness of a policy
	The results of information governance are as follows:
	Minimization of risk of failure through structured governance and oversight across all EIM initiatives
	Challenges for user adoption, growing data, content, and management controls are addressed through best practices, standards and guidelines, and a governance model
	Clear identification of responsibility and accountability for all EIM initiatives
	Continuous improvement in environment

(continued)

Human resource management	Human resource activities include recruitment and selection, compensation, training (education) and development, and career management. The focus lies on the continuous monitoring of required capabilities for people to act optimally within the organization when it comes to information
	The results of human resource management are as follows:
	Selection of the right people with the right capabilities
	Guided and controlled growth path through the right training at the right time
	Reduction in mistakes in day-to-day operations
Technology and Infrastructure	*Technology*
	Concerns the equipment, hardware, software, procedures, and technical knowledge brought to bear in the firm's transformation of inputs into outputs
	Infrastructure
	Serves the organization's needs for transporting (all physical and technical resource) information across the organization and ties its various parts together
	The results of technology and infrastructure are as follows:
	Alignment between various technologies and infrastructural needs
	A solid and scalable foundation for future development
	Consistent platform in relation to information management

2.14 Role of CIO/Information Manager

Today's CIO is judged on the efficiency of information

The responsibility of developing and implementing an information strategy usually lies with the CIO (chief information officer) or someone with a similar function. He or she is preeminently capable of bringing the supply and demand of information into balance.

Matching the different strategies with afocus on information resources is very complex. This requires many skills on the part of the CIO or information manager. There is a strong similarity between the roles of the CIO (or information manager) and the role of a movie producer. The film director (regisseur) coordinates the efforts of various people involved in making a film and wears many hats. The CIO serves an analogous function within his or her organization. The CIO's role is in practice characterized by the fact that it does not fulfil just one single function. CIOs play a multitude of roles and carry out a variety of tasks. This is seen as the chief capability of this role: being able to perform a wide range of roles and tasks.

The directing (control) refers to the interplay between demand and supply – knowing what the demand side wants in order to be able to optimally translate the question to the supply side. The demand side is usually the business, which has certain needs and requirements. The art of a good translation (see below on translation step of information strategy) consists in listening; recognize the question behind the question and manage expectations. Based on the translation of the needs and

requirements the CIO starts talking with the supply side, usually the ICT department. Together they will discuss ways of answering the question from the supply side based as much as possible on existing resources. To do this right, the CIO must have different competencies.

Several competencies can be applied to both the CIO (or information manager) and the producer. Coach, director, conductor, performer, designer, artist, and manager – all are applicable and needed. The multitude of roles and competencies is needed in a variety of situations. Both the CIO and the director need to adapt to the continuously changing environment. CIOs (director) who are capable of inspiring others through their own strength, vision, and behavior will generate the best results within a given situation.

The metaphor of director (regisseur) is almost completely applicable to the role of the CIO. Another good metaphor is that of game leader conducting a game. The CIO or information manager can be seen as a game leader within a complex environment.

Game leader of a special game. There are no losers in the game

A CIO must possess the following competencies to execute his or her work:

Competency	Activity
Overview of situation	Stakeholder questions on all relevant information, needs and requirements to pbtaom a good and comprehensive overview of current situation
Companywide accountability	Accountability of actions and outcomes of all stakeholders
	Showing commitment to the whole organization
Set out shared policies	Encouraging the vision behind the policy direction for the whole organization
	The theming of problems as a result of conflicting goals or interests of stakeholders; articulating and promulgating policies
	Monitoring the progress of processes and, if necessary, arranging for timely adjustments
Organizing cooperation	Mobilizing stakeholders, stimulating enthusiasm, and inspiring others to contribute to the whole
	Connecting stakeholders (for example by combining goals)
	Monitoring the commitment and input from stakeholders as well as the results of cooperation by the whole organization

Source: Propper (2004a)
Directing is both a career and a competence.

2.15 Plan of Approach to Information Strategy

The approach for executing the information strategy is composed of three steps (Fig. 2.8):

Preparation
Analysis of design
Translation

Step 1: Preparation

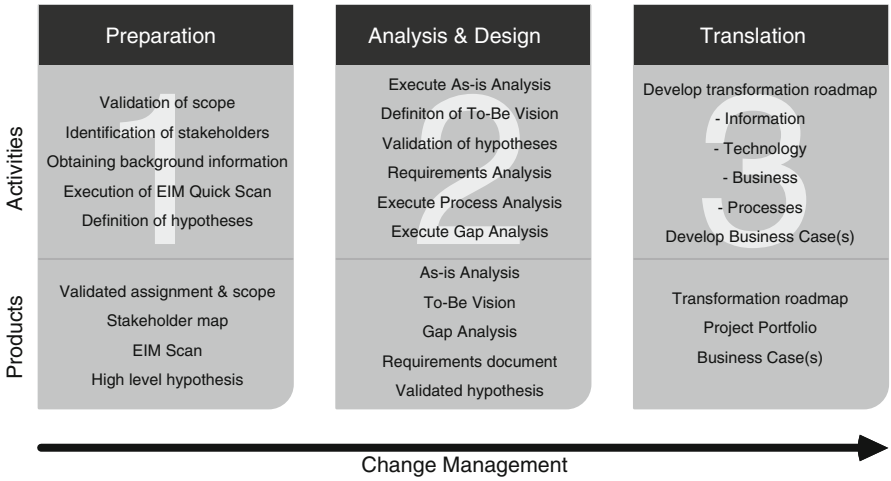


Fig. 2.8 Information strategy plan with respect to approach

The preparation step is crucial for the successful implementation of the roadmap. This step has its primary focus on defining and validating the scope, based on the results of the EIM quick scan and defined hypotheses. In addition, the identification of the stakeholders is of great importance. If one understands the forces within an organization, then the right tools and methods can be implemented to give meaning to the following steps.

Step 2: Analysis and design
The analysis and design step continues with the deepening of the defined hypotheses from the preparation step. At this stage it becomes clear whether a hypothesis or assumption is correct and what the impact is on the organization. In addition, the analysis step explores where the organization would like to be ("To Be" vision), independent of the current problems. Based on the As-Is (starting point) and To Be (destination) and the needs of the business, a gap analysis is performed. The gap analysis provides insights into the differences between the current situation (departure point) and the desired situation (destination).

Step 3: Translation
The translation step outlines the journey. The transformation roadmap provides a clear translation of the various initiatives required and relationships between these initiatives in the field of information, business, IT, and processes. Where necessary, a business case will be developed to add value and justification for an initiative to substantiate. Of interest in the translation step is that all the factors (either internally or externally) will be taken into account and weighed.

> The approach accelerates the time to value through prebuilt diagnostics, analysis, and design and transformation methodologies

What distinguishes this approach from other methods and approaches?

The journey is key. The destination may change
It is a journey together
Strong focus on preparing the organization's transformation
Strong focus on what is best for the organization instead of for individuals or departments
Primary focus on (information) problem area, impact on the business, and (possible) solutions; secondary focus on technique.
Start small, think big

2.16 Preparation

2.16.1 Validate Scope

In the preparation step, it is important to agree on the scope of the assignment, as it was in principle decided upon previously. The great danger in carrying out such assignments is having to deal with so-called scope creep. In the case of scope creep, the size of the problem will increase undetected or uncontrolled. At the moment a potential problem or problem areas are being investigated, on the basis of the defined hypothesis, different areas are encountered. These problem areas are all important and related to the original problem area. It is therefore important in advance that both contractors and clients establish clear guidelines about how to deal with new insights.

2.16.2 Stakeholder Analysis

Another very important activity during this step is the mapping of the various stakeholders. A stakeholder is someone who has an interest in the strategy to be developed. A stakeholder can be positive, negative, or neutral regarding the initiative. It is important to ensure that all stakeholders are identified within (and outside) the organization and that the picture or insight regarding the stakeholders and their attitude or interest is continually updated. The interests or concerns of stakeholders may change. It is therefore important to develop a plan that will be able to influence the different stakeholders. Ultimately they can affect the success of the rollout of the strategy.

2.16.3 Acquire Background Information

Also in this step, the necessary background information will be studied to define a wider reference context. This is necessary for exposing the problem area to different viewpoints and get a grip on what lives within and outside the organization.

2.16.4 EIM Scan

The EIM scan is used for the definition and preparation of hypotheses. The goal of this approach is to gain insight into the problem areas on the level of information within the organization. On the basis of this scan, a number of hypotheses can be defined that will be further investigated in subsequent stages of the information strategy's development. The defined hypothesis should include the following points:

The key business issues with a business rationale for the transformation program,
A vision of what the business change and technology enabler could be,
A high-level transformation design outlining how the change could occur.

The EIM dcan has the following steps:

1. Identify the information domains
   - Identification of information domains,
   - Formulation of core properties per domain and per degree of maturity,
   - Formulation of hypotheses in relation to the maturity model.
2. Verify information domains
   - Validation of information domains with client,
   - Identification of stakeholders' information domains,
   - Scheduling of appointments with stakeholders.
3. Execute EIM scan.
   - Stakeholder interviews,
   - Validation of hypotheses,
   - *Scoring of maturity model*.
4. Present results.
   - Presentation of the scan's important conclusions,
   - Maturity model.

In step 1 the client's organization will be extensively discussed together with the client, along with the market in which it operates, developments in that market, and the goals the organization has set itself. In collaboration with the client, the places in the organization where information plays a crucial role will be defined.

The information domains are defined based on the analysis in the first step. An information domain can be defined as a set of subjects belonging together on which information is present or should be present. Then, for each maturity level of each information domain, stereotypical statements are defined. These are statements that one could expect within an information domain at a certain level of maturity. The purpose of predefining stereotypes is that these statements help in determining the current level of maturity of the organization in the information domain under consideration.

In step 2 there is a thorough analysis of the information domains and the defined hypotheses. Knowledge of the organization and its goals is an important aspect of the EIM scan, which is why there is a lot of attention paid to the analysis during the review. Moreover, by discussing the model itself, it is often the case that the right people for the interview come forward. These are people with strong opinions about the functioning of the organization in certain information domains and the related ROI.

In step 3 the actual scan is executed. By means of interviews with different persons within the organization, an image is created of the current level of ROI in the different information domains. Moreover, this may create an image of the level of ambition that comes with the information domain. It is often the case that this image will have been created during the first step. If this happens, it is important to validate the image as soon as possible with the client because the risk will exist of making the wrong assumptions caused by a lack of knowledge regarding the organization's market.

At the end of step 3 it is then known what the ROI is within the different information domains and the level of ambition. Any differences that come to light here will serve as the basis for the further development of the scan.

In step 4 the findings of the EIM scan are reported, including the validated hypotheses. These hypotheses can be assigned to the findings regarding the differences between the actual and desired levels of ROI. A well-executed EIM scan leads to recommendations on improvements, including a prioritization of the expected of the recommendations on the organization's sustainable competitive advantage position. The actual current budgets and the capacity to make changes in the organization can be taken into account.

2.17 Analysis and Design

In the analysis and design phase the focus is on the "as is" and "to be." Additionally, in this step the hypotheses will be further investigated with respect to correctness. This will take place based on focus interviews, workshops, and various analyses.

It is very important to map the as-is situation within the scope of the problem. The as-is analysis provides insight as to where the organization is at this moment and how things are done. It also provides insight into the causes of the problems. The symptom remedy will generally not solve the deeper underlying problems. An organization should know where it stands if it wishes to define concrete goals. A different perspective and experience often underly the execution of an as-is analysis. The performance of an as-is analysis is often faced with a fair amount of trepidation. One has the impression that the focus will be on finding errors and inefficiencies in daily operations and that this understanding can lead to staff reductions (redundancies) or even dismissals.

However, the true aim of the as-is analysis is to bring to light the various resources and how these resources are deployed within the organization. The focus is on optimizing the various resources and increasing consistency. This is to ensure that

employees are inspired to be open to new and better solutions instead of limited thinking. The ultimate goal is optimization based on the best (can) use of corporate information (resource), with the higher purpose of supporting the business objectives.

Executing the as-is analysis provides an ideal opportunity to socialize changes. The as-is analysis can serve as the basis for the change process.

2.17.1 To-Be Analysis

In addition to the as-is analysis, the execution of the to-be analysis is equally important. The objective of the to-be analysis is to inspire stakeholders within the organization to define a healthy ambition level (vision) for the organization, without taking barriers into account. The purpose of the to-be vision is to set the bar so high that the vision seems utopian. This will drive the organization to try to fulfill the ambition. At a feasible level, people will be much less likely to set their sights very high. In the end the organization will be at a lower level because of the lower level of ambition. By challenging itself, the organization will achieve loftier goals. This leads to other behavior, entrepreneurship, and thinking in terms of solutions and not of constraints, with the final result being a better optimized implementation of the stated business objectives.

2.17.2 Validation of Hypotheses

The hypotheses from the preparation step are further investigated in the analysis and design phase. They are first based on various analyses (as-is process, requirements) and then refined through focus interviews and workshops. It is important to understand the hypotheses and to identify the causes of problems explicitly. Problems rarely, if ever, exist in isolation. They are almost always the result of a combination of deficiencies in resources. In addition, it may turn out that the hypotheses are refuted after they have been examined. This could lead to a realization that assumptions in relation to a hypothese are incorrect or that a reformulation of the hypothesis is necessary.

2.17.3 Gap Analysis

The gap analysis is performed based on the results of the as-is and to-be analyses. How big is the gap and what is needed, in which dimensions (information, people, process, and technology), to fill the gap? In practice it is usually a combination of activities within different dimensions that ultimately results in a solution to the actual problem (within the given problem area).

2.18 Translation

2.18.1 Transformation Roadmap

The final translation and underpinning of what is needed to solve the different problems within the organization and to get to the next level of maturity take place in the final step. Based on the gap analysis, a transformation roadmap is created that identifies various activities and initiatives to be implemented over time. Also part of setting up the transformation roadmap is defining the interdependence of the various activities and initiatives. By establishing the underlying relationships between the different activities and initiatives it is possible to create the right preconditions at the lowest cost (in time, money, and resources) to eventually be able to fulfill the higher objectives of the organization. Also, this step looks very carefully at so-called quick wins. Quick wins can achieve fast results at a relatively low cost without hindering other activities and initiatives. Quick wins will also contribute to positive acceptance by users. As outlined previously, this entails a journey for the organization to fulfill its higher objectives (Fig. 2.9).

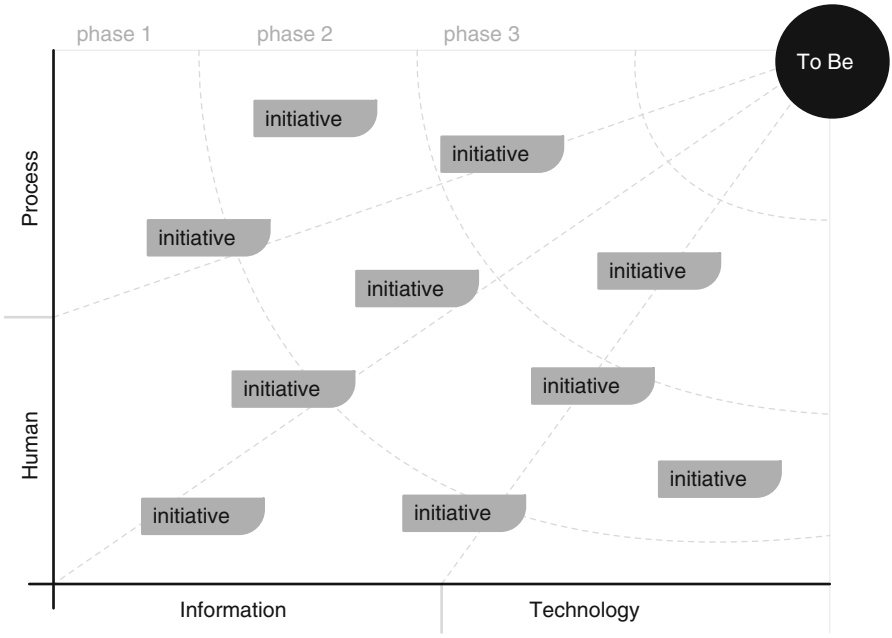


Fig. 2.9 EIM transformation roadmap

2.19 Business Case

The business case provides the rationale for the proposed change activities. The business case is very important from the point of view of supporting the actual decision making of the different stakeholders. Sometimes projects are necessary, but it is very important for an organization to see whether the investments pay off or not. The business case usually has the following main elements:

Provides confidence that action is affordable and benefits are achievable;
Is built logically and based on business benefits.

2.20 Critical Success Factors

To successfully develop an information strategy, the following factors are crucial:

Strong stakeholder management;
Key stakeholders on board early on; this will smoothen the data gathering process (interviews, workshops, background information) and keep the stakeholders there;
Focus on critical-information-related business issues (business outcomes);
Define goals together and confirm deliverables early on in the proces, then keep focused;
Get deliverables right first time so that sign-off process is quicker and smoother;
Focus on issues and solutions first, technology second;
Setting up of a responsibility assignment (RACI) matrix to get a clear picture of who is responsible or accountable for what.

2.21 Cases

Mortgage Application
Within a financial institution the release of a mortgage/taking on a mortgage was an especially long and error-prone process. The error sensitivity arose from the fact that data and content were stored and managed in different systems, independently of one another. Another reason was the privilege level of the different knowledge workers responsible for carrying out this process. This resulted in many errors such as a wrong name on mortgage documents, so-called pieces that were lost during the process of release, and long lead times. These errors affected mainly mortgage release efficiency and the quality of data and ultimately had a negative impact on customer satisfaction.
Through the development of an information strategy for the organization it was made clear that the actual problem was a combination of various causes in information management and information provisioning of the organization and not an isolated problem.

(continued)

2.21 Cases (continued)

Based on this information strategy a number of projects were initiated (parallel and sequential) that ultimately resulted in the greater efficiency of mortgage releases, which became quick and error free (high quality). In addition, integration between the various systems was considerably promoted, meaning that data quality was optimized. This resulted in a high-quality process on mortgages, a drastic reduction in errors, improved customer satisfaction, and a higher ROI.

Increasing Importance of the Internet

Times are changing. Large groups of consumers use the Internet to orientate the selection and purchase of products and services. The growing importance of the Internet requires major adjustments in the fields of business, ICT, and information.

Consumers wish to conduct increasing amounts of business through the Internet. The move to the Internet places high demands on the information resource. Information must not only be of high quality but also available and delivered almost in real time. This includes not only transactions and financial data but also surfing and clicking by the customer.

The Internet is increasingly becoming a sales environment, where having and presenting the right information about customers/users is essential; for example, people need to be properly addressed and their information must match exactly.

To be able to cope with these changes, organizations must think in advance. Reactive response to changing needs leads to isolated initiatives. These initiatives may have their usefulness in the short term, but at the organizational level, they come with a long-term cost.

It is therefore important to develop a strategic plan to cope with the continuous changes in the field of information. Based on the strategy it develops, organizations can be flexible, scalable, and adaptable. Coordination of the various initiatives in the short, medium, and long terms will lead to improved competitiveness and a higher ROI.

EIM at a Pharmaceutical Organization

Within a large pharmaceutical organization compliance is one of the main drivers of optimization of information management. To be compliant, it is necessary to create a sound foundation within the organization with respect to data and content management.

The validity of information is the key to success. It is therefore essential to be in control as an organization and at all times be able to deliver valid information. To do so, the organization must control all sources of information and optimize provisioning of internal information. Analyzing the processes within the organization gives clear insight into information needs. It gives also insight into the different transfer moments between the different departments. In particular, the transfer of information between different departments generally causes problems. Optimization of the processes, governance, ownership, availability, and reuse of information is key to maintaining control.

2.22 Role of Change Management Within EIM

The EIM approach focuses on the integration of information from the perspective of humans, processes, and techniques.

So far, the process to achieve results – including the order of steps one needs to take to achieve the best possible results – has been the starting point – the process to achieve results. The role of technique in this process will be extensively discussed in subsequent chapters (Chaps. 3–7).

These things are important, but equally important is the role of human beings in the process. Human beings, as employees, a group of employees, or the management of an organization, determine to a large extent the success of integration.

Integration means change, and making changes is an unruly activity. In other words, change can be planned beforehand and conditions can be set for the change process, but in reality change always happens differently than expected. This has to do with the human component in the process.

People do not like change. Therefore, change should not only be well thought out, planned, and implemented, but above all the process to be followed in making changes should be well communicated. And that is often the crux of the problem.

Change goes through three phases: initiation, design, and implementation.

The initiation phase is at the organizational level; it is the phase in which employees must acknowledge the necessity for change, mentally let go of the old, and must be given a justification for the change. Letting go of the old and establishing an improved situation are central (people must understand why change is needed).

The design phase focuses on creating an alternative to the familiar, and it must offer real perspectives, show personal advantages, and demonstrate that the change will have positive effects. Everything focuses on getting the organization and the people moving in the direction of a solution. During this phase, the future is outlined (SOLL, to be; people must understand how the change will be made).

In the implementation phase, employees have time to get used to the new situation; they can work with changed conditions and the situation becomes stable, calms down. Anchoring the change in behavior is an important part of this phase.

The course of such a process at the organizational level is characterized by a group of supporters that quickly accept the innovation, a majority that is neutral and feel little resistance to the change, and opponents who fiercely resist the change (Fig. 2.10).

Whether or not an organization should pay attention to the opponents depends on their role in the organization (management, key users, or less important players). Sometimes some people will need to be convinced individually to participate. This may lead to pressure.

Individual handling of the change is a different story. The employee goes through a grieving process, saying goodbye to the old and getting used to the new (Fig. 2.11).

This could be accompanied by considerable levels of stress; it may lead to burnout and should be managed. This cycle can be processed in a time span ranging from a couple of minutes to a few months or even years. Obviously, the impact

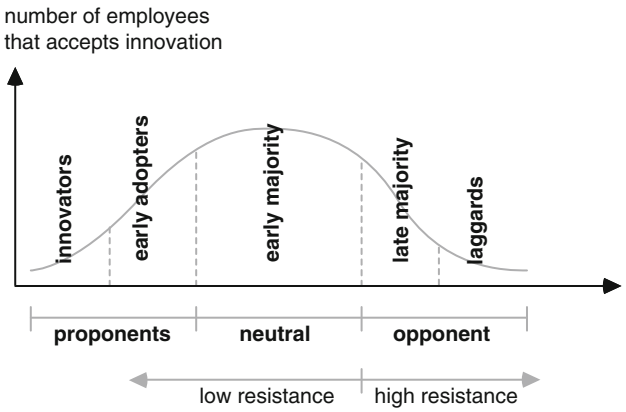


Fig. 2.10 Number of employees accepting innovation

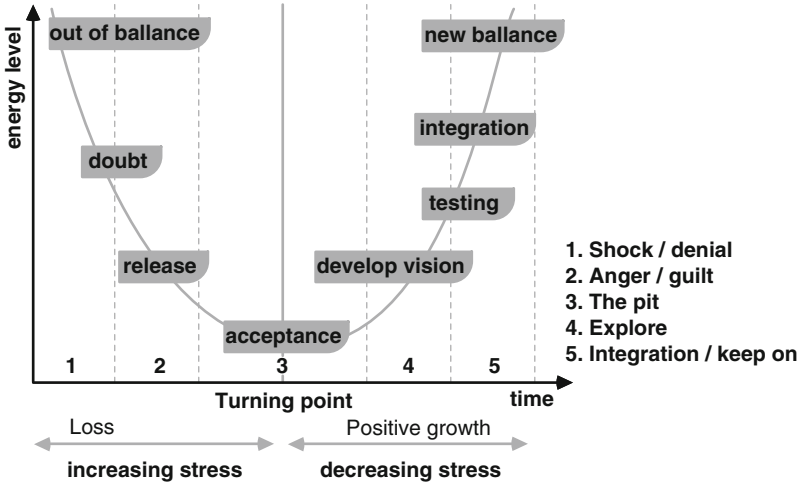


Fig. 2.11 Personal processing

depends on the resilience of the employee and how much the circumstances have changed. The overall resilience of all employees is called the adaptability or innovativeness of an organization.

So change is not just about the things that need to change, but also about changing the behavior of people individually and across the entire organization; it also includes the change strategy that is chosen to implement the change as well as the intervention method that is used when things do not go as planned.

Change management methods and techniques come in many shapes and sizes. There is not one specific method or technique that will be best for managing changes within the spectrum of EIM projects. In practice, we mainly use John

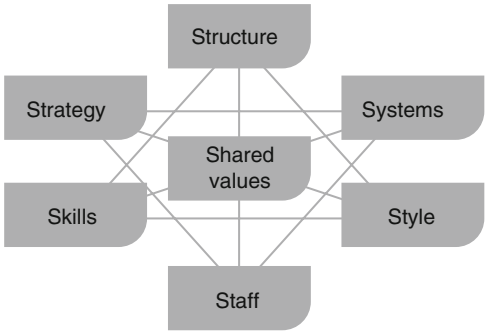


Fig. 2.12 Shared values model

Kotter's eight-step change model (Kotter 1995). This model says that a process of change may be erratic, but one can always discover a number of steps that are taken in a more or less fixed order.

The first step is to establish a sense of urgency (if the purpose of the change is unclear, the necessity for change is not felt and everything will stay the same). The second step is creating a guiding coalition, in order words,some leaders should stand up and say this is what they want and this is what the organization needs. When a coalition has been created, it is time to develop a change vision and strategy: How are we going to do this?

Up until now, relatively few employees have been involved in the change. The organization had been pondering a plan for change. These steps are important, but the next step is essential: communicating the change vision with anyone who is interested. The tone has been set, and after that all employees that want to can be put into action. This is the actual implementation of the change.

Of course, not everything will go right the first time, but here and there the first successes of the change become apparent. This also needs to be communicated to obtain the support of the laggards (late majority and sluggards).

The experiences that the organization now has with the change will be a reason for improvements and for acclimating the employees to the new circumstances. The organization will have to remain alert that employees do not relapse and also monitor and anchor the change in the organizational structure or culture.

Overall, change presents a challenge for many people. Within BMP change is mainly about making processes run smoother (BPM), making decisions based on the available information (BI), or obtaining swifter access to better results (ESR) or a better survey of the (internal) knowledge (ECM). These are already complex issues, but combining or integrating them only increases the complexity.

McKinsey's 7S model (Pascale 1981) shows that change within EIM mainly concerns the systems (how work is processed), strategy (why we are doing this), and skills (employee competencies), and then partly the structure (organizational setup), style (how everything is managed), and staff (execution of staffing process). Finally, this interaction will affect the shared values (the common values and principles in the business culture) (Fig. 2.12).

Change actions within one field will always affect the other six.

In other words, even though it seems that, in many cases, an EIM project only involves the choice and implementation of a new application or redesigning the process, one must be aware of the effects the project has on all other aspects of an organization and the changes it entails for the employees, various groups, and the organization as a whole.

Chapter 3
Enterprise Information Management (EIM)

Anja van der Lans and Peter van Til

3.1 Synergy and the Details

This chapter discusses the possibilities of obtaining new insights based on the same data and information using business intelligence, enterprise content management, and enterprise search.

Section 3.2 starts by clarifying some concepts within EIM.

Sections 3.3–3.6 tie in with the EIM triangle (Fig. 3.1) .

Section 3.2, *Availability*, outlines the conditions for bringing together structured and unstructured data, in other words, availability seen from a technical point of view. How is the storage of information arranged? How are the rights over the information arranged in the system? What does availability of information mean in an integrated environment of EIM applications?

Accessibility, covered in Sect. 3.3, focuses on indexing data and how this is achieved when structured and unstructured data need to be processed simultaneously.

Relevance, discussed in Sect. 3.4, means a different way of presenting results and compiling the results in a form custom designed for the end user. The human processing of data and information is central. When is information relevant and how is this determined? The human brain separates useful information from useless information.

A. van der Lans (✉)
Advisory Services, Incentro, De Meern, The Netherlands
e-mail: Anja.vanderlans@incentro.com

P. van Til
Business Intelligence, Incentro, Nootdorp, The Netherlands
e-mail: peter.vantil@incentro.com

P. Baan (ed.), *Enterprise Information Management: When Information Becomes Inspiration*, Management for Professionals 2, DOI 10.1007/978-1-4614-5236-2_3, © Incentro 2013

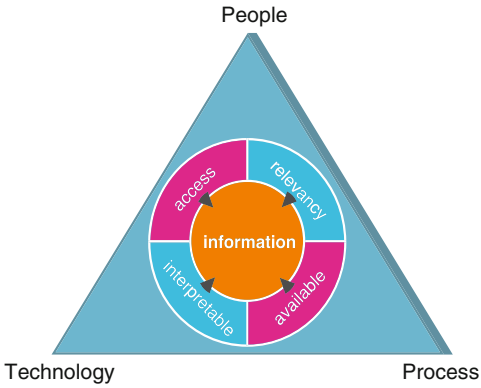


Fig. 3.1 EIM triangle

Lastly, in Sect. 3.5, *Interpretability*, the results of the EIM process is broken down into information and knowledge that one can keep to oneself but can also be shared with others; it comes closest to processing information. Interpreting is having access to all the available information in a form that fits the end user's processing capability. Interpretation is needed for making the best decision, initiating an activity, or managing a situation (monitoring, verifying, and so on).

3.2 Concepts

3.2.1 What Is Enterprise Information Management?

Just when some people in the field understand the consequences of merging the fields of business intelligence (BI) and enterprise content management (ECM) and making it available through a sound overall search, research firm Gartner decided to adjust the definition of EIM as follows:

EIM is an integrative discipline for structuring, describing and governing information assets across organizational and technological boundaries to improve efficiency, promote transparency, support agility and enable business insight. EIM is operationalized as a program with a defined charter, budget and resource plan. An EIM program implements the principles, models and requirements expressed within the organization's enterprise information architecture (Logan et al. 2009).

The good news in this statement is that almost everything can fall under EIM. The emphasis is on architecture and no longer mainly on business management aspects such as efficiency and understanding policy. A side effect is that this term is likely to be replaced by a new, more specific term.

Brunello Bonnati of IBM Italy (Bonnati 2009) is a pioneer in coming up with a changed insight into the definition of EIM. He noted that the combination of BI and CM in definitions so far was old-fashioned. The description given by Gartner in 2009 is very close to the combined EIM architecture that Bonnati had in mind.

Just like all organizations and products, EIM must go through its own performance curve, and it has made huge progress in its own maturity process. In our first book on this subject in Dutch, *Enterprise Information Management: de fusie tussen Business Intelligence, Content Management en Enterprise Search*" (Baan et al. 2010), we stated that EIM was still in its infancy and needed case studies, maturity models, and frameworks.

The most important model that we currently use is the EIM triangle. The incentro (center of a triangle) consists of information, and the sides of the triangle include process, people, and technology. Information has four important aspects: availability, accessibility, relevance, and interpretability. These aspects form the requirements that the information must meet, and in practice, they also often lead to obstacles. The different approaches used to avoid or clear obstacles are in particular discussed in Sects. 3.3–3.6.

Availability is characterized by the facts that are registered (and thereby made available) by the organization; accessibility means that everyone knows what facts have been registered, that there is knowledge within the organization.

Relevance invites you to examine the facts, and interpretability means that, based on the facts, science, and research, knowledge is created that helps the organization to make the right decisions and to distinguish itself in the market. Such a distinction is necessary to survive and, if possible, to surpass the competition.

From a technical point of view, facts are registered in the operational processes within a certain system. This is called system data. These system data are spread over multiple systems and become more valuable when these systems are integrated. This is called integrated data. Relevance means that the information must add value to the company and that research on deviations from normal patterns is possible, based on the data. This situation is known as support data. Interpretability means that the data are used for decision making, also known as decision data. The added value of data increases from system to decision.

From a human point of view, the four aspects lead to four types of users of the information. System data are mainly used by employees. These employees register facts, which are processed through his or her process, and then use these data in the same or a related process. Integrated data are used by the team leader and represent information at the departmental level (not yet aggregated into guiding information for the entire organization). Usually, this is fairly detailed information that only relates to the team that needs it. So it is quite easy to define what the data should be about. Support data are used by business or information analysts. Based on these data, problem areas and possible deviations become clear; however, the causes are still unknown. The analyst searches for the causes and comes up with measures to make full use of the deviations. The manager (depending on the organization, this will be senior management, directors, or the board) has access to the decision data. These data were used to feed all kinds of key performance indicators (KPIs). As a result, the manager has an overall knowledge of the organization and can determine or adjust the course based on this information.

The final dimension of the information is the process. Information follows a fixed order that can be captured in a process flow diagram. When the process flow

Table 3.1 Concept overview

	Availability	Accessibility	Relevance	Interpretability
Technology	System data	Integrated data	Support data	Decision data
People	Employees	Team lead	Analyst	Manager
Process	Register	Integrate	Analyze	Decide

diagram is defined, it becomes clear whether or not the information follows the most efficient route and what improvements could be made. The numbers, times, frequencies, and other aspects of each step in the process can be registered. This makes the data available within the organization. In most cases, this provides overviews of many different systems.

To obtain a more comprehensive overview and address contradictory information from different systems, connections can be made between systems. Data from all source systems are integrated, making the data accessible to the information worker. Relevance is added during the analytics process to supply the right support on questions within the organization. Finally, the right decision can be made during the *decide* process (Table 3.1).

3.2.2 EIM Vision

Now that we have looked at technology, people, and the process within the model, it is time to look at the characteristics of information, the criteria that information must meet in order to fit in the EIM vision.

EIM can be summarized by a number of key characterizations. These are listed below.

- Information is content AND data;
- Knowledge about knowledge workers' information needs is crucial;
- Maximizing return on information requires an integrated vision of ECM, BI, and enterprise search;
- Metadata and taxonomies play a crucial part in both BI and content management, but supply insufficient guidance to integrate both fields of study;
- The EIM maturity model helps focus on areas where increased return on information can make the difference;
- EIM in itself is an umbrella term;
- A maturity model gives insight into the steps that need to be taken when the situation is mapped out;
- A search solution helps to identify information needs and is therefore often a useful first step.

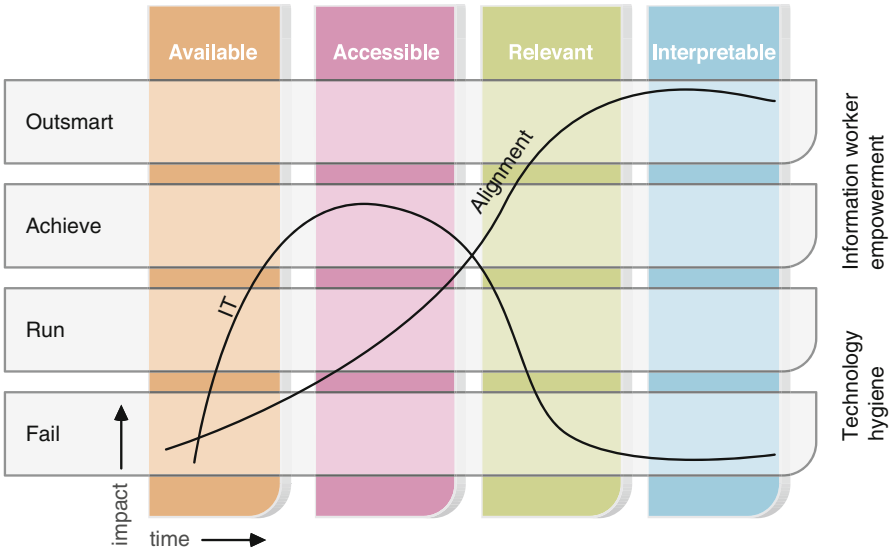


Fig. 3.2 EIM maturity model and constraints

EIM means supplying information from integrated systems in an efficient manner so that the organization can make full use of the available information. In other words, the organization sees a return on information (Fig. 3.2).

With this principle in mind, the phases of the maturity model within EIM are characterized as follows:

3.2.3 Fail

Structured information is seen as a byproduct of the production process. There is no insight into the significance of the information outside the process. Only information that is primarily of interest to the proprietary process is important. Unstructured information is a matter for the individual. There are no guidelines for the metadata to meet or about how to handle information.

3.2.4 Run

Unstructured information remains a matter for the individual. There are official documents such as product and service descriptions, but there is no process that

guarantees that the documents are up to date. There is no version management and no modification procedure. Many documents are laid down in the personal environments of the employees. There is no overview of availability. The organization relies on employees' expertise for making documents available. Structured information is still seen as a byproduct, but the whole information chain is also given consideration. However, the situation is not stable yet. Information is available internally. A change in method does not automatically affect the steps preceding or following the modified process. One is often overtaken by events. Introduction processes last long due to a lack of coordination.

3.2.5 Achieve

In certain organizations such as the financial sector and the government, the requirements about what is daily practice are determined by the need to comply with the laws and regulations. External pressure forces the organization to disclose information, making the organization transparent for the outside world.

Within many other organizations, there has been a growing awareness on this matter. Organizations understand there is added value in information. Programs are established to stimulate a companywide application of information. The focus is on data quality. The coherence between structured and unstructured information is considered, and both are involved in the information chain within the organization processes. Information is made available internally and externally to improve communication and to achieve a more efficient business. By publishing information on the Web, the organization hopes for fewer questions to the helpdesk because customers can visit the organization's Web site and find there the information they seek. Using and analyzing information generated from recent experiences, the organization hopes to gain a better understanding of the process. Better knowledge leads to better insights and better management. Based on these experiences, the company tries to make predictions and, thus, anticipate what is coming.

3.2.6 Outsmart

Unstructured information and structured information advance together to do battle with the competition. The only goal pursued is competitive advantage. Based on the available information (both internal and external), an appeal is made to the creativity of the employees to use information in a way that creates a distinction between themselves and the competition. In this phase, the return on investment is visible because the organization stands out in the market. This distinction naturally has consequences for management, so that the outsmart phase changes back to achieve. Then, how to achieve the outsmart phase again is examined.

3.3 Availability

3.3.1 How Do We Make Integrated Information Available?

This paragraph outlines the conditions that can be set for bringing together structured and unstructured data. In other words, we look at availability from a technical point of view and especially with the following questions in mind: How does it affect the way we find combined information for search purposes in an index? How do users store information? How are the rights over the information arranged in the system? And what does availability of information mean in an integrated environment of EIM applications?

To make integrated information available from both structured and unstructured information, systems need to be linked. In a structured environment, this can be done through the use of a data warehouse. Within an unstructured environment, no connections are made, but an overview page is created to show what the systems are (portals).

To connect the systems, the content of both systems must be included in an integrated overview. The process of adding structured and unstructured (meta-)data to a unified index is called indexing.

This allows organizations to gain insight into what they are doing (unstructured: policy on paper) and what they say they are doing (structured: facts from processes).

What will actually happen when EIM is applied? Previously, it was emphasized that the integration of BI and ECM takes place through enterprise search and retrieval (ESR). Data and content are combined in a portal using a search engine (Fig. 3.3).

In its most basic form of search, the occurrence of a search term in an index is enough to generate a result (hit). However, combining data solely based on the occurrence of a search term in an index can result in strange information that will not always be considered relevant by the end user. It can also show the index in which the hits occur. Search solutions become more advanced once they can be sorted with a predefined subdivision (Fig. 3.4). Information was sought about a colleague named “Weegink” in the structured database Webfocus and the ECM system Documentum. The results are considered relevant and efficient by the user because all available information from the two systems is shown in one screen.

Do database records and unstructured documents use similar terms? In other words, do you get relevant information from both streams when searching for a certain word or term? Sometimes, but usually you do not. This means that the focus should be on quality indexes in which synonyms and related terms and concepts are present. A database consists of tables, columns, and cells. The content of the cells is always related to the column in which they occur, except for free text fields. The term “Toyota,” for example, is therefore linked to “car brand.” These terms form a structure. This is used in the index to enrich unstructured information, which gives it the functionality of a basic thesaurus (taxonomy, list of keywords). The number of relevant results increases significantly and terms that are shown can be refined

Fig. 3.3 Combination of structured and unstructured data

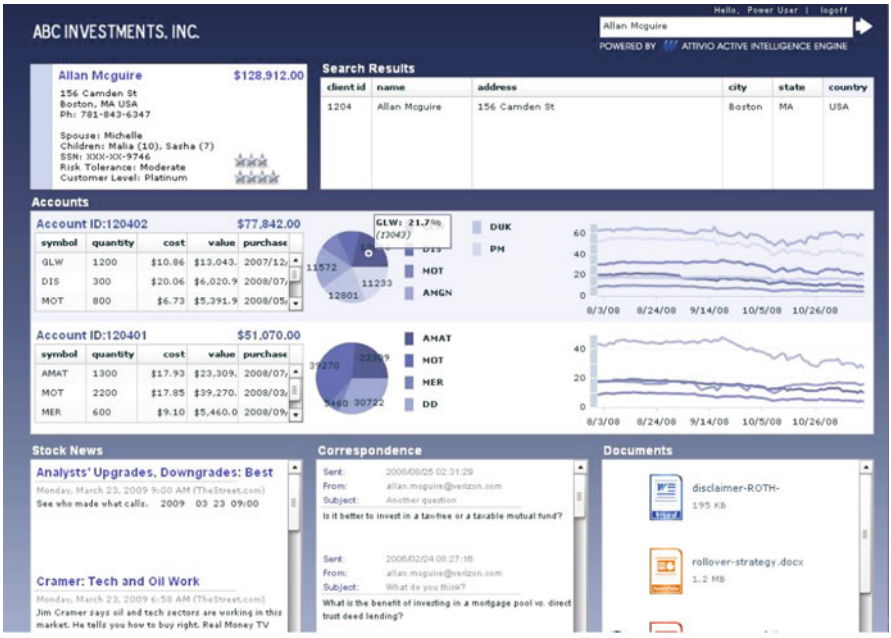


Fig. 3.4 EIM portal

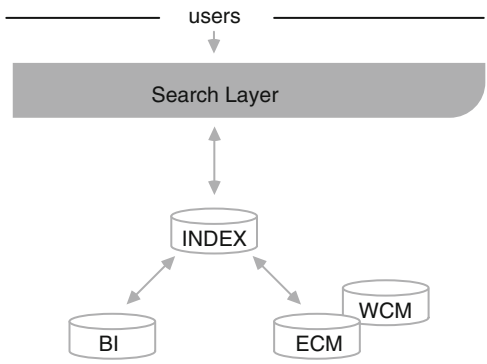


Fig. 3.5 General EIM architecture

using the entity "car brand." This so-called guided navigation based on entity extraction offers the user an advantage. Entity extraction means that the hits on "car brand" can be divided into corresponding terms (e.g., Toyota, Daihatsu, Volkswagen).

What does that look like? The look and feel of the results page can be put together by user based on the preferences in their profiles. This will be a composition of results from the database and unstructured sources. Figure 3.5 shows an example of a fusion of BI and ECM.

The main advantages of searching both structured and unstructured information are as follows:

- Expanded analysis capabilities (more coherent information for making decisions; e-discovery, risk management, fraud, and compliancy are improved/easier to substantiate with cross-analytics);
- Easier/simpler use of BI capabilities (usability and information access are improved);
- Improved efficiency and effectiveness in research and development investments (return on investment and return on time are both increased).

What do we need to achieve this? The key is a shared frame of reference. This can be filled with metadata. Metadata are the remedy that always helps of the solution. However, metadata in a structured environment have a different meaning than in environments that store unstructured data.

Therefore, there should be proper coordination within the organization so that metadata can provide the link between information flows in a uniform manner.

Metadata in structured environments show the relationship between the term in a cell (Toyota) and the columns (car brand) in which they occur. Metadata in unstructured environments show the relationship between a term (Toyota) and its position in a document (title, keyword, body), whether or not in relation to other terms (for example, thesaurus terms).

Metadata in BI serve as the technical and functional description of a data element. In addition, metadata also describe the relationship between tables and keys. Metadata in ECM are the notes to a document. Who wrote it, when was it written,

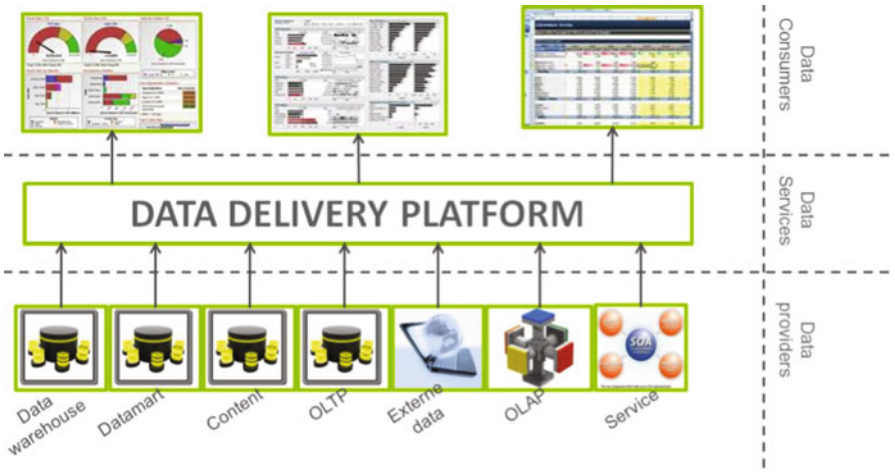


Fig. 3.6 Data delivery platform by Rick van der Lans

what keywords should be added, etc.? The information that is generated based on textual analysis is, of course, also an important form of metadata.

All these metadata are important for the integration of BI and ECM with the help of ESR. In ESR, the BI data are also a form of metadata. Dimensional data in particular are important in this regard. Dimensional data constitute the context of the data. The number "100" in the column turnover by itself has little meaning. It gets interesting when it is known that it is 100 euros, that the turnover applies to article x, customer A, supplier 2, employee m1, etc.

Metadata need to be coordinated well in order to relate the right pieces information to each other. Based on these metadata, one can provide the information that belongs together. Because an "all-in-one" index is implemented, all information becomes available with one query. The origin of the data is no longer relevant. This creates a universal index, i.e., unified information access (UIA).

This observation is also interesting from the perspective of the data delivery platform as described by Rick van der Lans in his article trilogy "The Flaws of the Classic Data Warehouse Architecture" (van der Lans 2009).

Van der Lans makes a distinction between data providers and data consumers. These two groups are linked to each other by the data delivery platform. Imagine this platform as a major index that can be queried by data consumers and that indicates the place where the answer to a question can be found. In the ESR world, applications based on this idea appear.

The search layer in Fig. 3.6 is the data consumer, the index is the data delivery platform, and BI, WCM, and ECM are the data providers.

However, the big question remains: How do we fill up the index in a way that the search layer can answer our questions from here? The example below provides insight into this technique.

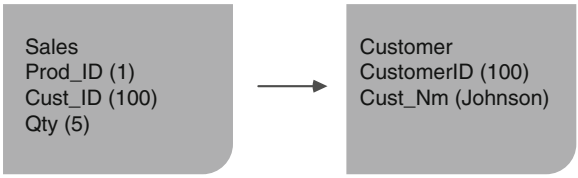


Fig. 3.7 Entity-relationship diagram (ERD) structured information

Table 3.2 Index of structured information

Origin	Metadata	Key	Content
CUST_ID	Customer number	100	100
CustomerID	Customer number	100	100
Cust_Nm	Customer name	100	Johnson
Qty	Quantity	100	5
Prod_ID	Product	100	1

Table 3.3 Index of unstructured information

Origin	Metadata	Key	Content
c:\documenten\document1	Product	Doc1	1
d:\management\document_conf	Product	Doc2	1
e:\email\12345.html	Customer name	Doc3	Johnson

First of all, we need technical and business metadata about the tables and columns. These are unlocked via the system tables of a database management system. Missing information can also be provided through, for example, a case tool. In addition, the tool must be able to read and index the content of the columns (Fig. 3.7).

Assuming that the names of the tables and columns and the relationships between the tables are taken from system tables, this means that the tool knows that Cust_ID equals CustomerID. This results in the following information in the index (Table 3.2):

A search for Johnson, based on this index, now shows that this customer (key 100) bought 5 units of product 1. This information can be combined with information derived from content inside or outside the organization through textual analysis. This could lead to the following information (Table 3.3):

Since the information in the tool is built up in the same way, regardless of its origin (structured or unstructured), it can be combined and integrated however the user desires at that moment. What is the distinctive quality of a unified index compared to an ordinary or standard index? In a unified index, terms are derived from both structured and unstructured sources. For structured information, this is the bare data combined with the relationships between the data and the table, column, row, etc. For unstructured information the position of a term in a document, sentence, or field and the links between the word and other terms are specified. A search for "bicycle manufacturer" may yield names of several bicycle manufacturers in a table in a structured database.

The structured data function as if they were a taxonomy for unstructured information. Based on the terms (in this case, names), a number of additional results may be found in the unstructured information, which may lead to interesting information. This makes it more valuable than ordinary search engines in which results are not combined. Whereas until several years ago the market was dominated by search engine providers that could combine several unstructured information sources into one index (integrated within structured or unstructured environments only), there are now products on the market that take a huge leap forward.

The disadvantages of working only with an index for structured information are that it lacks context and interpretation and analysis of data has not yet taken place. The disadvantages of working only with an index for unstructured information are that metadata and structure in the terminology are missing, the information is not suitable for decision making, and the results are only useful for the provision of information. The limitations of searching in both structured and unstructured information will be overcome by the use of a central index. Structured information provides structure to unstructured information, and unstructured information provides the context within which data become more meaningful.

3.4 Accessible

3.4.1 How Do We Make Integrated Information Accessible?

This section focuses on indexing data and how this is achieved when structured and unstructured data need to be processed simultaneously.

The new way of processing information obtained from indexing into new, more valuable information (more insight, faster processes, and better results) is to combine information.

Combining is merging the results from an index in response to a question to give a complete answer to the question based on the results. The question constitutes an important element in determining whether one will obtain the desired results. If one starts combining without knowing the question or without knowledge of the functional structures, one will get a Cartesian product (a Cartesian product is the result of a combination of two tables in which all records in one table are linked to all records in the other table). In an index for car brands, colors will occur.

Combining will lead to the following results (Table 3.4):

Entering a question can drastically reduce the number of possible results. The question of possible colors for European cars limits the number of results to four. Limiting the number of results adds value because the total amount of information can be processed faster.

This combination does not yet take into account the functional dependencies of car brand and color. What if Toyota does not produce cars in black? Then this combination can never be shown in the results of a combination.

Table 3.4 Combining

Car brand	Color	Combination	
Audi	White	Audi	White
Toyota	Black	Audi	Black
	Red	Audi	Red
	Blue	Audi	Blue
		Toyota	White
		Toyota	Black
		Toyota	Red
		Toyota	Blue

The use of a central index has the advantage that all the relationships between sources and between structured and unstructured information are tracked by the index. No results can be generated that are functionally impossible.

3.4.2 Data and Information

When discussing the maturity model, it was mentioned that in the fight for competitive advantage, it is important to be able to combine both structured (data) and unstructured (content) information. This can be done in several ways and at different levels.

At the ultimate level, all information is shown on screen with one question, and this information can be used both in daily business processes and in ad hoc analysis processes.

As Table 3.3 shows, combining can technically be done on two levels: on the metadata level and on the search level. The metadata level is the most structural way of combining and integrating results. It is not without reason that these two terms (data and content) also occur within the connect phase of the various EIM components.

An example of combination on the metadata level is the unified index (Table 3.3). Unstructured data are transformed into structured information based on textual analysis, and this information is combined with structured information from the BI environment.

This creates, so to speak, a derivative content information environment that is incorporated within the BI. This combined environment offers the possibility of incorporating new insights into reports.

It is always possible to create a link from the derived environment to the original document in the content environment. Thus the original document need not be copied and added to the BI environment. BI techniques can now be applied to the unstructured data, which offer completely new insights.

Integration on the search level is based on the idea of two separate pillars in the provision of information: structured and unstructured environments. No attempt is made to merge these two environments; the integration is based on recognition.

Fig. 3.8 Combination of structured and unstructured information

'Product number 1001' is viewed in a business application (Fig. 3.8). The screen shows a combination of the most important structured information and unstructured information that is available on the company network. Structured information is presented in the graph, and related reports are positioned next to the graph. The unstructured information can be found below the graph. With one search, the user is shown all available information and so can quickly gain insights. The disadvantage of this solution is that all business objects need to be recognizable. Two selections are made, one on the structured information, the other on the unstructured information. This is, so to speak, programmed into the solution. So in the example, the underlying code could have been: when search criterion includes the word "product," use the query "select * from product where product number = xxxx." This would take some intensive maintenance, but it could lead to some interesting results.

3.5 Relevance

3.5.1 How Do We Make Information Relevant?

This section discusses relevance as a different way of presenting results and compiling the results in a form custom designed for the end user. Human processing of data and information is central. When is information relevant and how is this determined? The human brain separates useful information from useless information.

The information that we gain by indexing and combining structured and unstructured data offers many possibilities. Users need help by determining what is important them; otherwise they will drown in the abundance of information.

It is hard to recognize interesting information in a sea of (un)desired information. Supporting people to find information quickly and easily is therefore an aim in unlocking information.

A good starting point is a personal approach in which the needs, wishes, and interests of the individual end user are taken into account. Users expect current and relevant content, and if they do not find this on a Web site or as a result of a search, they will find other ways to obtain their information. The customization of the presentation of the results is also called personalization.

Personalization in IT means making information personal. As a result, end users will quickly find information that is interesting to them. Personalization includes a set of methods and techniques to offer users information that is custom made for them.

Personalization also means that information can be filtered, prioritized, sorted, and presented based on the user's role.

How is personalization established?

- Information from or about the user; the user is registered in the information systems and derives rights based on this.
- Information entered by users themselves (profile); users can indicate what they want.
- From a third party (through identity); based on known data, information is made available to a group, department, location;
- Context; certain users make information available based on their knowledge of the interests of the other users.

Research shows that most problems in personalization have to do with retrievability and understandability. First, users must be able to personalize information and then understand what adjustments are possible. Proper explanation of the possibilities and quick results facilitate a smooth implementation.

Most users do not want to spend much time on customizing their profiles; they just want to find information. The design of the information source should be well put together, and the methods of personalization should add value and not be too complicated. Personalization can be beneficial to both businesses and users, provided it is well designed.

3.5.2 Filtering

Not all information can be made accessible to everyone. There are authorization schemes that indicate what information can be shown for each role. Nonaccessible information is filtered as a result of combination.

Many ECM systems (especially collaboration tools) can issue alerts whenever a document or content of a folder is changed (additions, deletions, new versions). The systems include rules for sending e-mails to the person who set/activated the alert or maintain a registration the user can retrieve in a separate folder or directory of a Web page. Consider the example of X. Internal research should reveal which roles within the organization use enterprise search and which groups of knowledge workers are there within the organization. After all, each group has its own approach to obtaining desired search results. This approach should be respected and implemented to make optimal use of the EIM solution. This can often be achieved by setting up group profiles in the search environment that are linked to levels and functions within the organization.

It is essential that an organization have a good idea of how to use its employees for maximum efficiency. Persona modeling can be helpful in this. Personas are often used for making IT solutions user-friendly, particularly user interfaces. Personas are established based on a target group study, after which a limited number of typical users are defined. Those users do not really exist, but, for the sake of the effectiveness of the use of personas, they are described as if they did. This means that a persona is described in terms of, for example, demography, needs, biography, preferences, and sometimes even photos. This puts a face on the persona, and in designing the user interface (for example a Web site) one can take into account the way this representative persona prefers to use the Web site. This, combined with a description of how the person wants to use IT solutions, form persona scenarios that in turn form the basis of the design.

In usability, the search behavior is considered a source for determining the desired interaction by users. By gaining knowledge about the way employees search for, use, process, forward, and provide information, it is possible to find a way of supplying information that is consistent with employees' methods.

The first step toward personalization usually entails connecting with the organization's previously defined access rights to systems and environments in the form of an active directory (AD, user administration of ICT department) or lightweight directory access protocol (LDAP). These are network protocols that describe how data from directory services are to be approached.

The best-known personalization techniques are as follows:

- Rules-based personalization: personalization based on a fixed set of rules (function, department, role, task); for example: report ABC must be presented to manager immediately after publication;
- Content-based personalization: personalization based on the interest history of a particular user in which similar information is provided; for example: since you have read documents about XYZ, you might also be interested in XYZ…
- Collaborative filtering: personalization based on the interest history of a particular user in which information is offered that users who are similar to this user also found interesting; for example: customers who looked at this document also looked at XYZ…

Rules-based personalization based on the LDAP or AD is most commonly used within organizations. In addition, content-based personalization in applications has become more common. This personalization technique is also called explicit personalization and offers visitors the opportunity to create their own profile (with areas of interest) using the application. This form of personalization is discussed under the terms alert, profile, and RSS feed. These types of personalization come in two variants: the push method (in which the system brings the relevant results to the user) and the pull method (in which user themselves retrieve the most recently saved relevant results from the systems). With push, the articles and information that fit the profile are automatically shown on the visitor's personal page. If new articles are posted, the visitor will automatically be notified by e-mail.

Content-based personalization and collaborative filtering are based on implicit personalization, in which analytic software is used to offer relevant information to users. This software tries to understand what articles users read, what products they buy, or what sections of a Web site they visit. This information is used to show, on future visits to the site, Web sites or articles that the software considers of interest to the users.

Implicit personalization is also described as recommender systems. These systems learn what information is of interest to users and apply this knowledge to prioritize information. These priorities are used to limit the set of information. Different criteria may be used to define these priorities, for example, learning from what the user found interesting in the past, the opinions of others with similar interests, or topics in the information that might appeal to the user.

To make predictions about the interests of users and the method of structuring and presenting information that most suits the user, an information system must learn from and about users. Therefore, research has been done on feedback by users to an information system. This feedback may occur because a user explicitly gives his opinion on the information that is retrieved or indicates his preferences concerning the structure and manner of presenting the information. Another approach to making predictions is by analyzing the way in which a user handles the information system. Without feedback, an information system cannot learn and, therefore, cannot adapt to user wishes and needs.

Interesting in this context is the fact that the same techniques that are used to gain knowledge about users are also used in improving results within search systems. This is called relevance feedback (see also Chap. 7, Sect. 7.4).

Filtering is done based on a profile. What functionalities are generally desirable?

- Users want to be able to subscribe or unsubscribe to notification services, to create and edit profiles, and to add and delete favorites;
- After visitor (users) have logged on, they can set up personal pages using a user-friendly wizard. On these personal pages, visitors are informed about information available on the site that tie in with their specified areas of interest, topics, or features;

- A thesaurus or list of keywords is often used as a basis for selecting areas of interest, so that terms may be used uniformly over the sources (build concepts; he terms fall within these concepts);
- Users want to be able to adjust what is presented in an application or on a Web site, select the language, locate help documents, and choose a specific wallpaper, with or without videos and with or without navigation (toolbars and other resources);
- Users can add results in the form of reports and interesting Web sites to their list of favorites simply by clicking one button. The addition of external links is also possible by means of an input form, which is processed by an application manager;
- Based on their profiles, users can gain access to special pages that are restricted for other users; these pages have so-called read access;
- Each type of access can be set individually by users. The administrator has far-reaching capabilities with respect to defining personalization rules. Personalization rules are centrally managed and applied to content at different levels, such as the page level and content element level, and within forms. Personalization may also be applied within text lines and phrases.

3.5.3 Prioritizing

Once a profile has been created, results can be sent to users in a specific order. Depending on the role of the user, it may be useful to prioritize information. This means that a weighting factor is applied to the results that remain after filtering. The weighting factor might stipulate that, for example, news about the user's own organization in specific newspapers should be prioritized in result sets that are presented to managers.

3.5.4 Sorting and Structuring Information

Besides limiting information based on interests, structuring (a subset of the) information is also a step in the direction of unlocking information. A clear structure that fits the user can improve information retrieval. The information can be presented in a desired order based on the weighting factor. Sorting is done primarily on the basis of date, source, or relevance of the results found. In addition, more refined ways may be used, for example, by grouping results by categories such as genre or topic or aspects such as price, title, or author. In addition to techniques for structuring based on information attributes, certain grouping techniques have been developed. Here, information is grouped based on the goals it serves; for example, a user wants to learn something new or relax.

3.5.5 Presenting

The presentation of information also represents a way of supporting people in accessing information. Therefore, research has been done on user interfaces of information systems. Information should be presented in a way that immediately makes clear what information is of interest to a particular person. Things like presenting priorities and giving explanations about priorities play a role in this. Also important is the way the applied structuring method is presented. The way information is presented largely determines the value that information adds to the activity that the employee must perform.

In practice, personalized information occurs in many places: on dashboards of collaboration tools, in virtually all Content Management Systems, in Document Management systems, and in search systems. In dashboards of collaboration tools, profiles and contact options, as on, for example, Hyves or Facebook, are available as standard items. As a result, the most important information is immediately available for an employee. This means that for every level of the organization (strategic, tactical, and operational) a type of dashboard occurs that is equipped to meet the needs of a given employee.

In search solutions, just as in ECM systems, a query can be forwarded to a colleague or other interested party through e-mail. Still, many questions remain to be answered: Can a search profile be saved so that the right form of presentation is used for each user? Can queries be saved for reused or to be shared with others?

3.6 Interpretability

3.6.1 How Do We Assess the Value of All Relevant Information?

Finally, in this section, the results of the EIM process are processed into information and knowledge (they are interpretable) that users can keep for themselves or share with others. These human processes come closest to information processing on computers. Interpreting means having access to all the available information in a form that fits the end user's processing capabilities. Interpretation is needed to make the best decision, initiate an activity, or manage a situation (monitor, verify, and so on).

EIM makes all data available that are relevant for answering this question, provided that the information was present and available within the organization.

Now that the information that is relevant to users has been filtered from the structured and unstructured systems, it is assumed that the information will lead to substantiated and faster decisions.

Whether or not EIM contributes to better decisions cannot be determined in this context. However, it can be determined that a decision is made based on a full(er) picture of the situation. More factors need to be considered: more considerations lead to more informed decisions.

This demands a lot from users. EIM is all about users' ability to interpret information well. Its usefulness is determined in that way.

3.6.2 Interpreting in the Maturity Model

The maturity model is divided into two sections, above and below the line. The focus of EIM can in particular be found in the *achieve* and *outsmart* levels. After all, in these phases the focus is on operational excellence, innovation, and smart decision making. But obviously it is also necessary to acknowledge aspects and make decisions in the lower levels of the maturity model. Information management is also needed here and will be a part of EIM.

Thus EIM is deliberately applied within the business processes in which an organization aims to attain the achieve and outsmart levels. All organizations wish to distinguish themselvesfrom the competition at those levels of the maturity model. Better opportunities within these processes lead to differentiation within the given industry. Better opportunities mean better and faster decisions.

EIM has everything organizations could need to help them improve performances compared to the competition. EIM's technical implementation forms the basis for improvements. Whether there really will be improvements depends on the technical possibilities. Organizations ultimately determine the results of EIM. If an organization cannot or does not want to change, then the return on EIM will be limited.

3.6.3 Why Is Technology Not a Problem?

Technology is not a problem because all technologies that are needed to make active contributions to the environment of information productivity already exist. The technical components have existed for many years, there is a mature software market for all individual components, and there is a solid base of knowledge in the form of professional services.

The preceding points are best demonstrated by looking at the enterprise software market. Ever since analyst firms like Gartner and Forrester started describing the potential advantages of an enterprisewide approach to information management, platform suppliers for enterprise solutions such as IBM, Microsoft, SAP, and Oracle have been busy buying logical components that, when combined, contain all functionalities needed for an EIM solution. Acquisitions have been numerous, and almost all the leading vendors in both the BI and the ECM markets were taken over by one of the major platform vendors. They all bought BI and ECM solutions. Microsoft's acquisition of FAST shows that the enterprise search market is next in line.

What seems to be lacking is a market participant with a healthy perspective on how these building blocks that are purchased should work together as an integrated solution for its customers. The platform suppliers all worked hard at integration on the software level, but because of that, they must still put forth propositions that explain what the (added) value of the combined approach might be for customers.

Enterprise search seems like the obvious facilitator in this process, but again that is reasoned from a technology vision, not a business vision.

In terms of help, the two analyst firms mentioned previously, Gartner and Forrester, do not offer much. They explain that customers themselves must first identify their information needs in their entire business and develop an information strategy based on those identified needs. This seems logical, but it has such a high level of abstraction that it is not useful or feasible for most entrepreneurs.

The following goals (requirements for information) should be pursued within organizations to successfully introduce EIM:

1. Information must be provided together with the context.
2. Information must be provided at the right time.
3. Information must be reliable.
4. Information should allow for cooperation.
5. Information from various sources must be combined.
6. Information must be easy to use.

3.6.4 Information in the Right Context

Information only adds value when it is used in the right context. Contextualized information is therefore essential in the use of information in daily processes. Integration of BI and content management, on the one hand, and processes and information, on the other, should form a unity of two parts.

3.6.5 Information at the Right Moment

In practice, this means information on demand. The moment information will be desired or required is difficult to predict, especially when the information is required for a customer contact, for example. But also in repetitive processes the moment at which information will really be needed is nearly impossible to predict. A repetitive process is in need of management by exception. Whenever the situation shows no default behavior, this must be reported. To prevent deviant behavior, the information must be available before the deviation actually occurs in order to avoid the deviation.

3.6.6 Information Must be Reliable

Knowing is better than not knowing as long as the information that leads to this knowledge is reliable. This is the foundation of information provisioning. Remove this foundation, and the information loses its value. Acting on the basis of unreliable information is akin to gambling: the outcome is unpredictable.

3.6.7 Information Must Allow for Cooperation

Information is important when parties are working together. This applies to the cooperation that takes place between customer and organization and supplier and organization, but also to employees in the organizations themselves. A customer places an order and provides a trigger within the organization to produce products and send them to the customer. The production of goods leads to a reduction in the stock of raw materials within the organization. As a result, the supplier of raw materials may receive an order. When the order is ready, the financial department generates an invoice and sends it to the customer.
Information that arises in different parts of an organization enables cooperation between employees in this total business process. Information is also exchanged such that employees can learn from each other. A senior employee may provide a junior employee with information so that both employees gain better insights and thereby achieve better results.

3.6.8 Combining Information from Several Sources

Combining several sources leads to new information that allows for better process support. Reports, documents, and articles can together provide a more complete picture of a specific problem.

3.6.9 Information Must Be Easy to Use

Information must be easily accessible. It should more or less automatically be offered to users. The system must know what users are doing and they usually compose the desired information. Based on this information, the system must provide the information in the right form at the right time.

Chapter 4
Business Process Management (BPM): The Information Loop

Anja van der Lans

Now that we live in an age in which cocreation is taken seriously, entrepreneurs must focus and anticipate, or they will go under. Organizations, within or outside their own chain, work together for products and services, subsequently find an appropriate business model, and offer their products on the market. This does not take a year of preparation. This may be done in months or, if possible, even weeks.

The development and structuring of business activities by setting up or changing processes must be simple, fast, and effective in order to fit in with this kind of cooperation. The sometimes fleeting existence of (co-)partnerships and the essence of doing good business, as a starting point, are the most basic parts of business process management (BPM); what makes us do what we do and for/with whom? Business processes are not only structures that ensure uniformity and effectiveness; they also connect people and departments, systems and techniques, but above all processes are used to cross link information.

BPPM is a field of study that deals with giving direction to and examining the interconnections among an organization's activities in a way that is in line with its goals. For 40years, IT has played an important role in the efficient performance of these activities, in supporting human-related activities, and in directing and controlling these activities. Like business intelligence (BI), enterprise content management (ECM), and enterprise search and retrieval (ESR), BPM is a field where technology, information, and organization are logically connected to each other.

Section 4.1 starts with an introduction and our vision of BPM as a part of enterprise information management (EIM).

Sections 4.2, 4.3, 4.4 and 4.5 tie in with the limitations of BPM from the perspectives of people, process, and technology to get to information as shown in the EIM triangle (Fig. 4.1).

A. van der Lans (✉)
Advisory Services, Incentro, De Meern, The Netherlands
e-mail: anja.vanderlans@incentro.com

P. Baan (ed.), *Enterprise Information Management: When Information Becomes Inspiration*, Management for Professionals 2, DOI 10.1007/978-1-4614-5236-2_4, © Incentro 2013

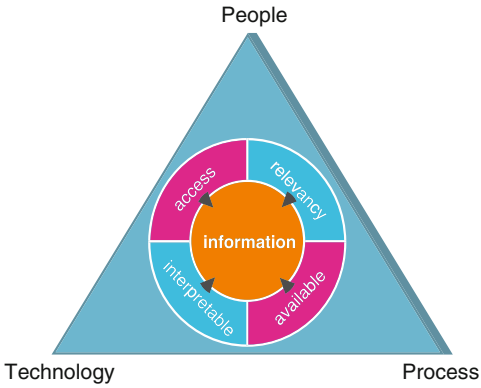


Fig. 4.1 EIM triangle

4.1 Concepts and Vision

4.1.1 What Is Business Process Management?

In every organization activities take place that are associated with the goals of the organization. Whether this is a factory where a product is made or a service provider committed to fulfilling a certain need. An activity usually consists of a number of steps that need to proceed in a particular order and according to certain agreements. These steps can be arranged in a diagram, a process scheme (flow). When the organization starts arranging its activities based on this scheme, it is called process management. An organization with the processes described has insight into the route along which activities occur, starts thinking about the optimal alignment of the activities (optimization), and tries to work faster, easier, or in a different manner (innovation).

Wikipedia describes BPM as follows:

Business process management (BPM) is a holistic management approach focused on aligning all aspects of an organization with the wants and needs of clients. It promotes business effectiveness and efficiency while striving for innovation, flexibility, and integration with technology. BPM attempts to improve processes continuously. It can therefore be described as a "process optimization process." It is argued that BPM enables organizations to be more efficient, more effective and more capable of change than a functionally focused, traditional hierarchical management approach.

This shows that nowadays, organizations attach great importance to processes that distinguish them from the competition and that generate higher profits from management. In the past, methods like workflow management, lean management, or Six Sigma were used to map and analyze processes. The focus was mainly on the process itself or on improving/controlling the quality of the results of the process.

As John Jeston and Johan Nelis wrote in their white paper "Management by Process," "Six Sigma is not a subset of BPM, it is a useful adjunct…." Meanwhile,

the field of BPM has taken some giant leaps in the development of a pure inventory of steps in the process toward a total package of process, organization, technology, and people. As a result, BPM has moved up to change management, which entails thinking about the consequences that decisions will have for people and overcoming the resistance that may accompany any changes.

Within BPM a distinction is often made between:

Primary processes (or production or operational processes), which include all activities where the output contributes to the end result for the customer; primary processes represent the organization's right to exist;
Management processes, which are all activities needed to manage the organization and the primary processes;
Supporting processes, which are all activities needed to facilitate the primary processes.

Nowadays, in BPM the focus is mainly on the coherence between the process and the other important components within an organization such as strategic alignment, governance, methodologies, information technology, people, and culture. This means that the organization is the sum of all processes that take place within the organization. Consequently, BPM can be defined as describing, implementing, and improving processes in relation to the people and technologies that are needed to perform activities.

Note the term *governance* in the preceding list. Governance refers to the procedure or manner of management, ethics, and monitoring of organizations. It is linked to decisions that define expectations, empower, or verify performances. It consists either of a separate process, on the one hand, or a specific part of management or leadership processes, on the other. In organizations, governance is often translated as having key performance indicators (KPIs) for the main processes and for measuring the processes and acting accordingly (Jeston and Nelis 2006).

Fewer and fewer organizations today are focusing on the previously mentioned types of processes; rather, they are turning their attention to the set of processes concerned with the customer, the so-called customer-oriented processes. In this regard, organizations need to keep a close eye on the relationship among the different processes in the process chain (what process is performed and delivered to the next link? what are the responsibilities?), but the following kinds of questions should be addressed: Why does this organization exist? What are its mission and vision? What are its goals and what is its chosen strategy? All this, seen from a customer's perspective, determines whether an organization is performing the right activities and whether those activities meet customer demands.

Ownership and division of responsibilities are crucial for BPM to achieve the ultimate goal of process descriptions – ensuring business effectiveness and efficiency.

BPM makes business processes more transparent and the organization as a whole more flexible (agile). The best result with BPM, the most value-adding result, is produced when the organization's strategy leads directly to decision making on the processes and to an integrated process architecture and a structured approach (Fig. 4.2).

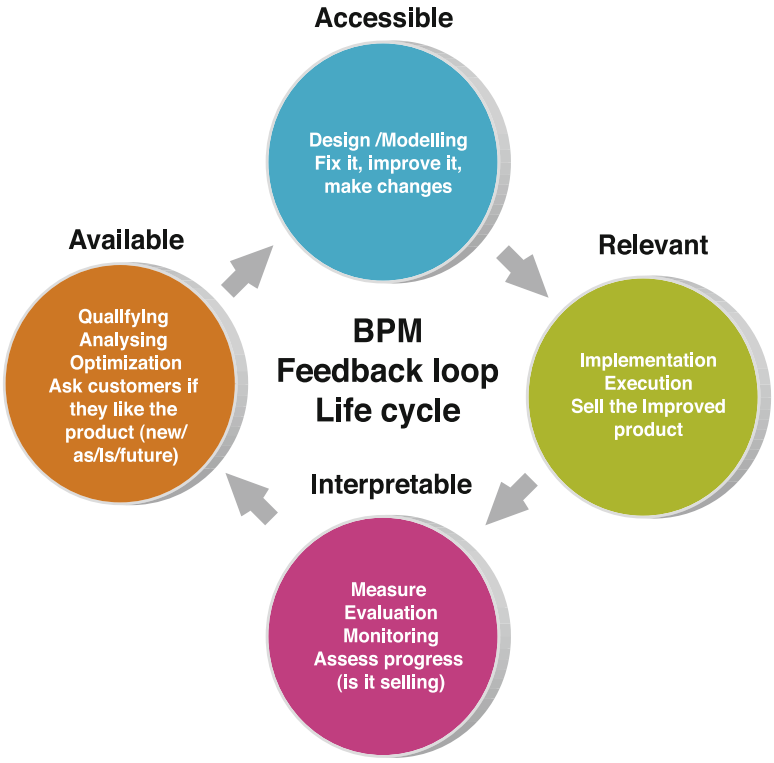


Fig. 4.2 BPM feedback loop and life cycle

The structured approach that over the years has led to optimum results in organizations is based on strictly using a feedback loop in which the cycle of qualifying and analyzing, via design and implementation, leads to measuring and evaluating. The ultimate goal is to use the conclusions from the evaluation to adjust the process and, in so doing, start a new cycle of the loop. The components of the feedback loop form the core of Sects. 4.2–4.5.

4.1.2 BPM Vision

BPM plays an important role within the EIM concept. For businesses, the vision of enterprise information means that a given employee always has access to information when it is needed. The availability of information alone does not constitute BPM. The business itself defines the power of BPM, in other words, its ability to make decisions based on the information available. Obviously, the available information and ability to make decisions must be in line with each other.

Drawing up an inventory of processes is important within all components of EIM (BI, ECM, ESR, and BPM) for finding out how EIM can contribute to these pro-

cesses in terms of the integration of the different components. The processes are essentially the links in the information flow between the separate components. The point, for example, in an invoicing system where incoming bills are recorded and then electronically distributed according to an authentication scheme – all this is done on the basis of the steps in a predefined process. The data about the payment (which employee of what department made commitments for what amount of money?) end up in a database so that reports can be created. After a while, this database becomes part of a source system for a data warehouse and may become part of a data mart to be used by management to draw conclusions about the extent to which the finance department is doing its job. Accounts can be linked to, for example, order administration and the underlying documentation of products that are ordered. The possibilities are endless, and an important goal within EIM is to match the integration between data and information (structured and unstructured, formal and informal) with the needs of the organization and employees in their daily work.

Usually, process optimization serves as the motivation for formulating and implementing an information strategy. The current processes within an organization will reveal problem areas regarding information provisioning. This will allow for investigating the impact on the overall processes and the improvements that could be made to optimize information provisioning.

The EIM maturity model offers good guidelines to show what the organization is capable of and what it needs to reach its full potential. Research may offer insight into what is required to bring the organization and the information management to the desired, higher maturity level.

From a people perspective, BPM means that all employees understand given information and are trained to make the right decisions. They enjoy the full support of management, and the organizational culture is based on transparency and the will to improve the business processes in such a way that the improvements are in line with the business goals.

BPM from a process perspective means that the business processes can be optimized and that the decisions that need to be taken to achieve this are supported by figures from the source systems or data warehouse architecture. Where in this process is there room for improvement? Why is this not optimally used? The fact that answers to these questions are still open does not mean that the organization must go back to using stopwatch techniques to minimize lead times; it does mean, however, that certain parts of the chain could work more efficiently if information in the processed were (re)used.

BPM from a technological perspective means that data are stored in several locations. In recent decades, the BPM field has been especially well positioned from this perspective, and this has led to a number of concepts that can be used to supply needed information to achieve the process and people perspective.

As was mentioned previously, in BPM, processes can be in a certain phase (qualify, analyze, design, implement, measure, and evaluate). Based on evaluation, a new analysis is conducted and the cycle starts all over again. The number of times the feedback cycle is looped shows the degree to which the organization is dedicated to optimizing its processes; more loops means that the organization will show more maturity in relation to its processes.

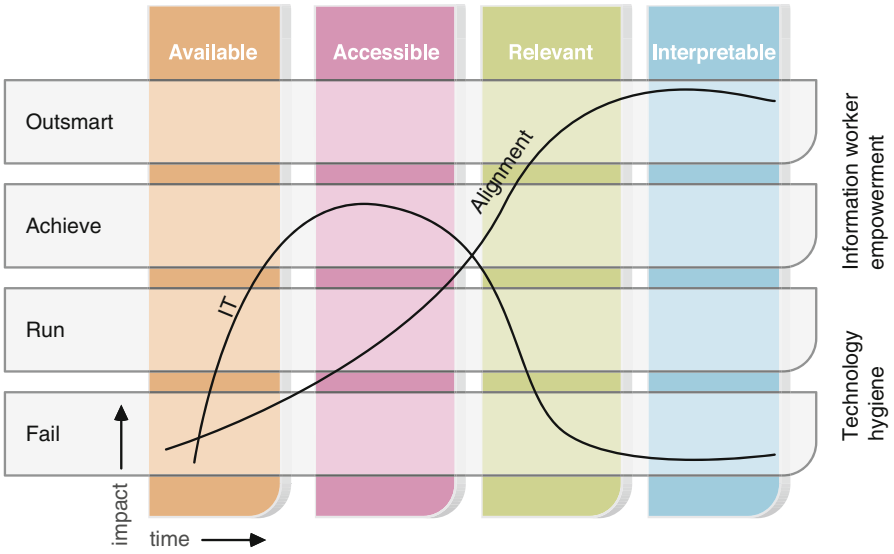


Fig. 4.3 EIM maturity model and constraints

Processes are usually described in terms of steps and in terms of the role/position that is responsible for a given step. Often it becomes clear that there are gaps between steps, that steps are performed twice by different people, that the process is not as clear as it first seemed, that processes never or constantly change, or that the goal of the processes (what are we doing and what do we want to achieve?) is ambiguous. These are all aspects that indicate the maturity level of the processes in an organization and that offer opportunities for process optimization.

This shows that change is a constant factor within BPM. This is not a process of change as in Kotter, for which the necessity should be considered; the process should be supported by the appropriate decision makers and requires a vision. Nevertheless, communication with all workers in the organization, and especially management, regarding the processes is the key to acceptance and success with implementation in BPM.

Communication provides greater understanding to the employees who work with the (sub)processes. Often, they are the ones who know the details about what is not going well and what elements of the processes could be improved to optimize the business and, in doing so, ensure the continuity of the organization (Fig. 4.3).

Four levels are used within the maturity model. These are fail, run, achieve, and outsmart.

The lowest level of maturity is fail, which simply means there is insufficient information available to certain processes to run smoothly. The next higher level means that the information for a specific area is just enough to execute the associated process satisfactorily. This level is called run.

The next level of maturity is called achieve. The focus of this level is efficiency. This level is about executing the primary and important processes as quickly and accurately as possible, making optimum use of available information.

The highest maturity level is outsmart. This level is about using the same information as your competitors and, perhaps, making smarter decisions. Organizations that have their primary processes at the outsmart level establish a strong competitive position.

When thinking about the curve in the maturity model to represent BPM (which is not drawn in the model now), it immediately becomes clear that processes (and therefore BPM) from the origin of the graph (limited time and impact; bottem left corner of the graph) has the same curve as IT and then develop in alignment with the needs of the information worker. At first, the focus is on collecting and mapping the processes in which IT is dominant but when it comes to interpreting and changing the processes, people predominate.

With this in mind, it may also be clear that the path to the outsmart level is not a straight line or achieved in one iteration. It is more like a spiral staircase, very slowly winding its way up. It is even possible that the gradation is so small that one needs to take a few steps back to make adjustments so that a new optimum may be found.

Maturity can be determined based on various aspects, but for both EIM and BPM information management and information value are important. Therefore, we will limit ourselves to discussing only the aspects information management and information value.

With this principle in mind, the stages of the maturity model within BPM are characterized as follows:

	Information management	Information value
Fail	Processes are (still) unclear; there is no program or project that focuses on improving the strategic processes. Processes are not documented and work instruction may not be up-to-date. There is no overlap between processes, which drives up costs and makes it harder to adjust processes	There are no process owners and nobody has a mandate to make the necessary decisions The transition of process steps between departments has not been regulated, and the information transfer is insufficient. Business logic and rules are not established. Compiling and filing of reports is effectively and efficiently regulated
Run	Processes have not been standardized but have an ad hoc character; however, they provide sufficient results. There is still much overlap between processes, and now and then small adjustments are made to a process Descriptions are made though still without a connection with other processes and some work instructions are communicated	Basic coordination of certain processes is initiated by the coordinators, though still without a mandate to implement decisions. Reports slowly start to get standardized. A policy is set for saving and destroying reports

(continued)

	Information management	Information value
Achieve	The organization has a clear focus on important processes, which are unambiguous and have a clear goal. Projects aim for a complete overview of the existing processes so that the impact of adjustments can be determined more easily and the information supply supports fairly standardized processes	(Sub-)processes are merged and process owners are embedded in the organization so that processes can be optimized and the decision making on these matters can proceed efficiently Business logic and rules are defined and serve as a basis for the standardization of processes
	Processes are described and communicated so everybody knows the expectations and goals of the work	Reports can be flexibly adapted to changes and stored under archival arrangements
Outsmart	The organization is constantly working to improve processes to enhance customer service. Processes and the organizational architecture are viewed in an integrated manner, and improvements are made continuously by means of an agreed-upon progra/set of guidelines. The entire value chain is described in a process-oriented way. This allows for an integrated information supply that can be easily adjusted. All processes are analyzed with a view to reusing them. This leads to processes that are strongly integrated, easy to support, and easily maintained	Information flow within the processes runs smoothly across departmental boundaries using workflow systems to further guide the process. Lifecycle management as a discipline is fully supported Business logic is defined in terms of supporting IT, which allows workflow to be used and adjusted more easily A robust review process is in place for compliance with policy. Reports are managed and produced effectively and efficiently and can be easily adapted to changing demands
	Process owners are actively involved in optimizing processes and improving the information supply	

4.1.3 Under What Circumstances Is BPM Required?

The righteous BPMer will say “BPM is always required” and he or she may have a point. There are situations, though, where it pays off to look into processes and see whether there are indications that the processes are not optimally adjusted to the goals they should be serving at that moment.

When managers are suddenly seen walking around with all kinds of spreadsheets or having people build small database solutions, this could be an indication of more or different needs within the organization to map out or control processes. Another indication might be that the number of overtime hours or the backlog is not decreasing. A sign that something is amiss could be that the same work is being done in several places within the organization or that work must be redone because something went wrong the first time. Such moments in an organization provide opportunities for improvement from BPM.

In addition to this, there may be a need, from management's perspective, to look into the processes or to adapt them. Often a merger or takeover has occurred (how do we interweave the processes of the two partners so that we benefit from the collaboration?) or a reorganization, change in strategy, or new laws and regulations that the organization must comply with.

Also, it may happen that the quality of the information from management provides insufficient possibilities to make decisions about the processes. This is often described as a lack of reliable management information or poor(er) performance of the systems. Naturally, management wants to know the cause of this, and therefore processes and reports on the processes will be examined.

From an employee perspective the processes should be described clearly in manuals (or digital variants) with descriptions and work instructions to ensure that, in case of high personnel turnover or a growing staff, new employees quickly become operational and structured training can take place.

Organizations benefit from well-oiled and flexible processes. They can quickly identify customer dissatisfaction with products or services and quickly make adjustments to their processes. Customers often appreciate when an organization acknowledges its mistakes and also demand that defective products be recalled, but then the problem must be corrected quickly and never occur again. The organization must be able to respond adequately, often with a change in process as the practical outcome.

4.1.4 Lifecycle Versus Workflow

In practice, workflow management is often mistaken for lifecycle management (see also Chap. 5 "Event Content Management"). A workflow that determines who can do what at a certain moment after all previous steps have been taken can be a blessing, for example, in a car factory where it is unthinkable to place the hood before installing the engine. But this can also be a heavy burden for employees. For employees, workflow management means waiting to perform assigned duties and sometimes even having to assign tasks to themselves to perform multiple actions. If a delay occurs in a previous stage of a process, employees cannot independently accelerate this process.

This may also mean that employees themselves create a bottleneck when they go on vacation or get sick. After a while, the workflow is divided among others, but often this does not happen automatically. Needless to say, workflow changes are not always received with applause.

A system with a built-in lifecycle management is much simpler as far as prescribed order and rules are concerned. The possibility of doing something with a product (often this is integrated in an ECM system, and then comes documentation) depends on the phase that the document is in. In the creation phase, for example, one can create a document, but afterward it can only be read, not edited (formal document); then, in the archive phase, the document is only available to members of management upon request.

4.1.5 Processes and Related Concepts

After the Industrial Revolution it gradually became clear that an organizational structure that was based on predefined processes offered a significant competitive advantage. The way these processes should be defined has been the subject of study for a long time. Large companies in particular started to map the processes at their own discretion and in doing so created interesting methods that have inspired many generations of BPMers to implement great optimization measures. Some terms and their relation to BPM are discussed below.

4.1.6 Business Process Reengineering or Redesign

Business process reengineering (BPR) is a management technique and methodology whereby an organization fundamentally and radically restructures its business processes in order to bring about great improvements in the organization. At first, BPR was a method in which all business processes came under the microscope to restructure the entire organization. BPR does not only influence the organizational structure; it also relates to the restructuring and changes in management style and organizational culture.

The main principles of BPR are as follows:

The organization is a collection of processes that can be restructured in a systematic manner.

The nature of change is revolutionary and consists of a transition from functional-oriented business units to multidisciplinary teams. The bureaucratic culture throughout the organization should be changed and customer satisfaction should be the starting point in every process.

Change starts with higher management. Upper-level managers should set an example for the organization. Change do not occur at once; it is a continuous process of adjusting.

Michael Hammer, founding father of this method, came to the conclusion that automating business processes alone would hardly have any effect on their efficiency. In his findings and examples from Ford Motor Company, he indicates that a different approach is required for the effective deployment of IT (Hammer 1990).

Hammer believed that automation combined with restructuring of business processes leads to much greater efficiency and effectiveness than automation alone. Hammer recognized that IT could make radical improvements provided that the traditional ways of thinking about how organizations should operate were abandoned.

Although over the years there has been much criticism of the methodology of BPR , it has led to more sophisticated models that are able to analyze business processes and make them more efficient and effective. Examples are Total Quality Management (TQM), BPM, and Six Sigma.

4.1.7 Total Quality Management

TQM is a management stream with continuous improvements being made to company performance and a focus on meeting client demands and fulfilling the business strategy. Obviously, processes play an important part in this, but this stream sees process as part of a greater whole that leads to a (more) efficient organization.

Company performance metrics include leadership, quality management of suppliers, defining vision and planning, evaluation, process management and improvement, product design, improvement of the quality system, employee involvement, recognition and reward, education and training, and a customer-oriented approach.

The central idea is that it is more efficient to perform all activities right the first time. This saves the organization time on corrections, failed products, and service (such as warranty repairs). Eventually, this will save the organization money. It is applicable both within production companies and service organizations.

Some TQM streams use quality circles, for example Deming's TQM. Unlike the radical BPR, TQM provides incremental changes– so not a revolution but evolution.

4.1.8 Workflow Management

Workflow is a business management concept that stands for a sequence of interrelated steps. It is the visualization of a series of edits that could be the work of one person, a group of people, a staff organization, or one or more simple or complex mechanisms. Workflow may be seen as an abstraction or real work. In control technology applications, workflow may single out a certain aspect of real work under a chosen aspect, thus serving as a virtual representation of actual work. The flow being described may refer to a document or a product that is being transferred from one step to another.

Workflow management is the management of the movements of information, often with the help of automated means. Workflow management ensures that the right information is transferred from one department to another according to the rules of the company in a transparent and efficient manner.

The term is most commonly used in ICT contexts, such as workflow systems or as part of ECM systems. Business processes are stored in workflow systems and ensure that an order, complaint, or other task is processed in the correct order by dividing them into subtasks. The relevant departments handle and approve these tasks in the appropriate order. The status and the course of a task can be checked anytime.

In addition to regulated processing of tasks, additional advantages may be a reduction in costs, improved quality of service, or increased control for the organization over the industrial process. In some cases these improvements can be explained by the fact that the processes in question require less direct involvement of management staff.

4.1.9 Lean Management

Lean manufacturing, or lean management, is a management philosophy that considers all things that do not create value a target for elimination. The method is derived from the Japanese car manufacturer Toyota. This "lean production" would lead to an increase in quality and a decrease in costs, which would lead to an improvement in company performance. Waste may have various causes, one of which is irregularity in processes, for example, variation in measurements. Lean management distinguishes eight kinds of waste: defects, overproduction, transportation, idleness, storage, movement, overactivity, and unused creativity and capacity.

The advantage of this method is that all of a company's energy and creativity is focused on optimizing all aspects that are perceived as valuable by the customer – such as price, quality, delivery, maintenance, and environmental charges – and whose costs are justified. The introduction and optimization of elements often bring about improvements in safety, industrial hygiene, and ergonomics.

The increased interest for process optimization resulted in a situation where the lean principle is now also applied in nonindustrial environments. For example, lean is also applied in supporting processes such as ICT and sales and marketing, but also in business processes like contact centers.

4.1.10 Six Sigma

The Six Sigma management strategy was originally developed by Motorola. It is applied in many sectors of industry. Six Sigma aims to improve the quality of the results of business processes by detecting and removing the causes of defects or errors so that variation in processes is reduced.

It consists of a collection of quality management methods, including statistical methods. Every Six Sigma within an organization follows a predefined series of steps and has quantifiable financial targets (cost reduction or profit improvement).

The term *six sigma* originates from terminology used in the production industry, specific terms associated with statistical models, or production processes. The maturity of a production process can be described with a "sigma" rating that represents the efficiency in the percentage of error-free products that are produced. A six sigma process is a process in which it is expected that 99.99966 % of the products are error free (3.4 errors per million). Motorola aimed for "six sigma" for all production processes, and this became the nickname for all business and technical activities used to achieve this goal.

4.1.11 Value Chain or Supply Chain Management

The value chain is a business management concept that links the strategically relevant activities of an organization. This concept may be used to describe a

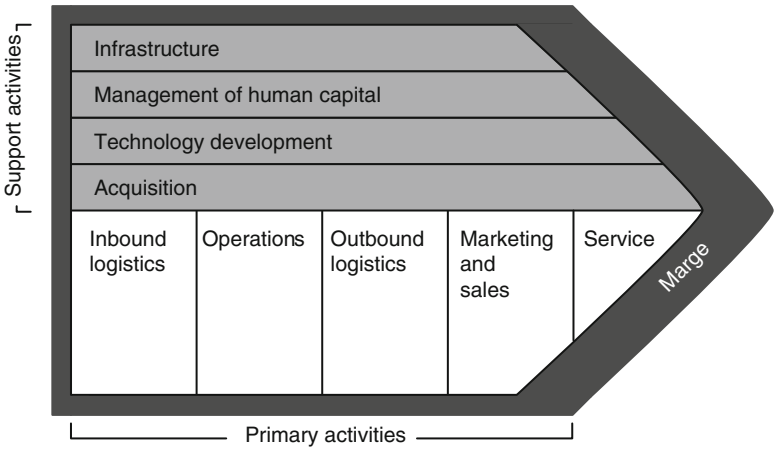


Fig. 4.4 The value chain of Porter

management tool for systematic research of activities or for interactions with third parties. In those parts of BPM discussed earlier, we saw that activities and interaction were usually defined in process flows. The value chain is related to the supply chain and is then mapped out as a process with a fixed order (see figure below). You could therefore consider the value chain a "super flow."

As stated previously, a value chain is a chain of activities. Products are processed successively through certain activities of the chain and gain value at each activity. The chain of activities as a whole adds more value to the product than the sum of the individual activities. It is important to distinguish the concept of the value chain from the cost structure in the implementation of activities.

This concept was first described in 1985 by Michael Porter in his book *Competitive Advantage: Creating and Sustaining Superior Performance*. Porter also claimed that competitive advantage could not be fully understood as long as a business was seen as a whole. It is essential to systematically examine all activities performed by an organization to determine the sources of competitive advantage. Therefore, all elements as such were put under the microscope and related to other parts of the process (Porter 1985).

A value chain divides a company into its strategically relevant activities to provide insights into the cost behavior and the existing and potential sources of differentiation. A company gains competitive advantage by performing these strategically important activities in a cheaper or better way than its competitors.

To find out exactly where profit can be made, a company will do everything it can to produce products as efficiently and effectively as possible at the least cost and the greatest value. The supporting activities are of minor importance in this model. They add little value and will therefore not receive the same attention as the primary activities (from management but also in terms of finance and number of employees/staff are concerned) (Fig. 4.4).

The value chain shows the primary activities, the supporting activities, and the margins of the organization. Net profit is the difference between total revenue of the organization and the total costs incurred to generate this revenue.

4.2 Availability

4.2.1 How Is Process Information Made Available?

Availability has to do with being able to use data that are generated inside or outside an organization and that can be used within the information management system of the entire organization. To determine which data from the source systems are needed to meet the information demands of (a department of) an organization, we need to know the elements that comprise this information need as well as the business rules that are applied to obtain the desired information.

To meet this information need, it is necessary to determine how any missing information can be added to the processes that are present within the organization or how new processes can be implemented within the existing environments. A common issue in this regard is information about customer satisfaction. All organizations want to know how (potential) customers experience the organizations. There are often no moments of measurement in the business processes to capture this information. New processes need to be added to the workflow to for this to happen (Fig. 4.5).

This section outlines the conditions for bringing process-related information together, which in BPM is often described as qualifying and analyzing. In other words, we are talking about availability from a human point of view.

What actually becomes available about the processes is, among other things, obtained through a workshop, analysis of the Strengths, Weaknesses, Opportunities & Threats of an organization (SWOT analysis), or a description of the organizational context (e.g., force field analysis). All this is done to determine which factors and parties influence the process being described.

Processes can be described in many different ways. In practice, this is usually done using a fixed method and the description is then documented in an application. Different kinds of relationships are defined and described in a responsibilities table that defines who is responsible, who has final responsibility, who is consulted and informed (RACI model), who is the owner of the process, and what activities must be performed in the process and in what order. This is also the time to find out whether

Fig. 4.5 BPM lifecycle, available

the activities are unique or recurring, whether they are optional or mandatory, and whether or not they add value to the chain (e.g., copy, sort, archive, transport).

The processes may be a description of the current or future situation (new or improved). If a process is already running, a description of the future situation is the result of a decision to apply a change. The result is usually a process optimization (e.g., a certain part needs to be improved, speeded up, have fewer steps, be done in a different order) or a totally different approach (e.g., random quality checks instead of checking each product).

By qualifying we mean collecting the needs of the customer to determine if we must, can, or want to fulfill them. This often leads to a requirements analysis with a survey of the needs in order of importance for the customer. With this approach, very specific goals can be set and the chances of a successful implementation increase. Also, it will lead to better alignment with the business goals and architecture of the organization, and an optimal design can be determined based on these conditions.

In a way you could say that the conditions that the design must meet are mainly determined during this stage of the loop. The conditions are usually culturally defined, that is to say, depending on the culture of the organization. Important aspects are the motivation of the goal (why do we want this?), interaction (who does what and in what order?), cultural (what is allowed and what is not?; business rules), and behavior (is there any resistance or is it strongly supported?).

Besides that, the impact of working with fixed processes will play a role. The IT-landscape needs to be mapped out, and any overlap with the selected processes, the extent to which selected processes interface with other processes within the organization, and management's control processes need to be identified. The goal is to supply management with information on (core) processes. The combination of these factors should lead to a Specific, Measurable, Attainable, Relevant, Time-bound (SMART) design.

So, making process information available entails (1) collecting information following the steps in the process, (2) processing the information in schemes, and (3) sharing the results of the measurement at the end of the feedback loop. The "making available" step in terms of qualifying and analyzing is therefore the first step in every description and change of a process (and the conclusion of the final step is to start describing once again), which makes it an actual loop.

4.3 Accessibility

4.3.1 How Do We Make Process Information Accessible?

The second part of the value chain focuses on the design of processes. The accessibility of information plays an important role in this (Fig. 4.6).

Design helps to increase the transparency, effectiveness, and efficiency of processes. A design enables an organization to better align information supply and IT with business processes. This makes it a very helpful communication system.

Fig. 4.6 BPM lifecycle, accessible

After all the information about the processes has been collected, sorted, and selected to apply to the goals we want to achieve, it is important to think about designing the most optimal process. This phase is called design or modeling. What actually happens is that the organization will look for improvements and make adjustments in the current design or come up with a new design in a schematic representation of the desired reality.

A model can be created of a business process in the form of a scheme. Different model types are used to achieve this. The most familiar model, the DEMO, is described in more detail below. The remaining model types will be described only briefly, and an example of a scheme will be given.

4.3.2 DEMO Model

DEMO is a methodology developed in an academic environment for designing, engineering, and establishing organizations or networks of organizations. The interaction patterns, shown as transactions between parties, form the foundation.

Jan Dietz is the designer of DEMO (Design and Engineering Methodology for Organizations). In this method, the actions of people are central in terms of "communicative actions": communication is essential for the functioning of the organization; agreements between employees, customers, and suppliers are achieved by communication. The same applies to the acceptance of delivered results – no result without communication (Dietz 1996).

The DEMO methodology is based on the following principles:

The essence of an organization is that it consists of people that act and negotiate with competence and responsibility;
Modeling business processes and information systems is a rational activity that leads to uniformity;
Informatics models should be understandable to all concerned parties;
Information systems should psychologically "fit" their users.

The DEMO approach provides a coherent understanding of communication, information, action, and organization. The scope has shifted from information systems

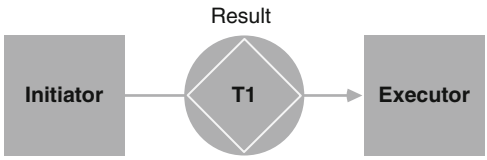


Fig. 4.7 Example of DEMO transaction diagram

engineering to business systems engineering, which indicates the significance of a clear understanding of both the information systems and the organizations.

When someone (a person) wants to realize that someone else creates the desired result, the communication on this will start with a request. The person receiving the request can respond to it by making a promise in return. After a while, when the desired result has been worked on, it can be announced that the desired result has been achieved. If this result is accepted by the person who made the request, the result is a fact. In DEMO, the pattern in communication between two people that is described here is called a *transaction*, and a chain of transactions is called a *business process*.

Figure 4.7 presents a diagram of the principle of a DEMO transaction between two actors.

The result of a transaction can be described as a fact. Object role modeling (ORM) is used in DEMO for this purpose. Because of the fact that with DEMO a business process can be described as a chain of transactions and the results of these transactions can be described with ORM, a clear connection between a business process and information is created.

Over the years, this methodology has grown into a basis for many organizations searching for a more pragmatic approach doing business for application in everyday practice. This approach mainly focuses on management and implementation of such programs with a strong focus on business processes. Determining the scope of a process is crucial because in organizations processes are often integrated in such a way that changes quickly affect other processes and systems. An organization rarely starts without a process history, but then a method will be applied in an existing context based on a very wide range of issues and goals. A connection will be established with subjects such as project management, governance, compliancy, (process) architecture, service orientation, business cases, and requirements management.

Increasingly, organizations are coming to acknowledge the added value of thinking in transactions and seeing processes as a management steering tool. The combination of solid theory and practical application unlocks the concept of "management by transaction" for a large group of organizations and businesses.

4.3.3 Data Flow Diagram (DFD)

A data flow diagram (DFD) is a graphic representation of the data flow through an information system. It is a commonly used technique within a structured analysis. Another name for a DFD is a data flow scheme (source: Wikipedia) (Fig. 4.8).

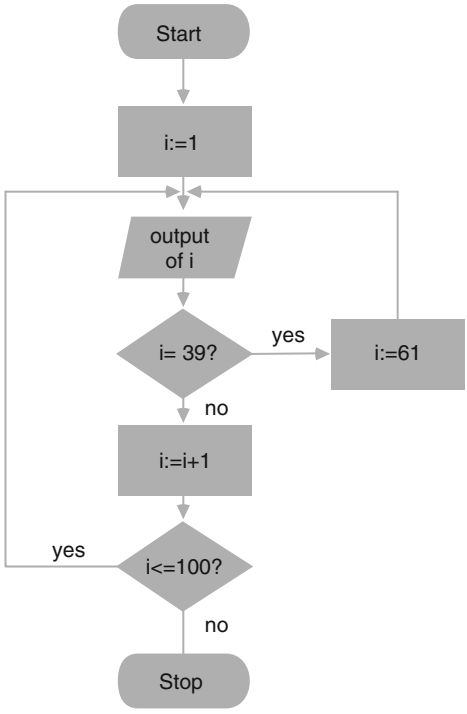


Fig. 4.8 Example of data flow diagram

4.3.4 Flowchart

A flowchart or flow sheet is a diagrammatic representation of a process. It is used mainly to help visualize a process or to find errors in a process (Fig. 4.9).

These diagrams are used, for example, to demonstrate how to assemble a bicycle or piece of furniture or to clarify how to make the right decisions on a tax form. These diagrams may range from simple schemes that are in fact nothing more than a phased plan to production schemes consisting of multiple pages.

4.3.5 Input-Process-Output Model (Also Known as Black Box)

The input-process-output model, also known as the IPO + S model, is a functional model and conceptual scheme of a general system. An IPO chart identifies a program's inputs and outputs and the processing steps required to transform the inputs into the outputs (source: Wikipedia) (Fig. 4.10).

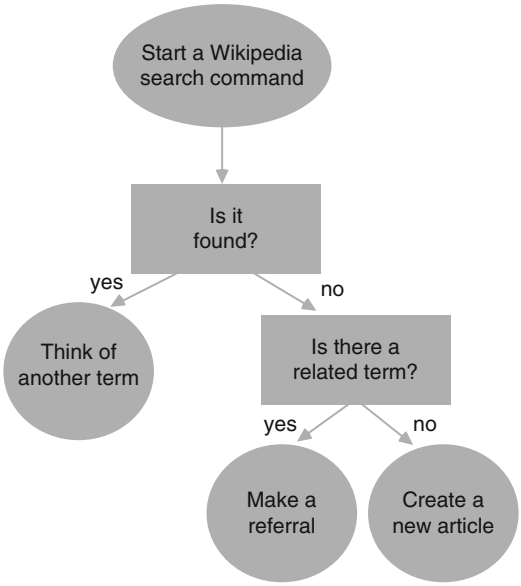


Fig. 4.9 Example of flowchart

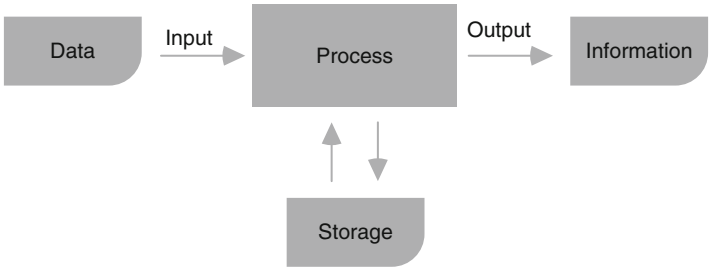


Fig. 4.10 Example of black box model

4.3.6 Supply Chain Model

The supply chain is a network of organizations, people, and activities that exchange information or products. This network is a logistical chain that ensures that products or services reach the user. The network transforms raw material and semiprocessed material into end products. Various subprocesses can be identified in a supply chain (Fig. 4.11).

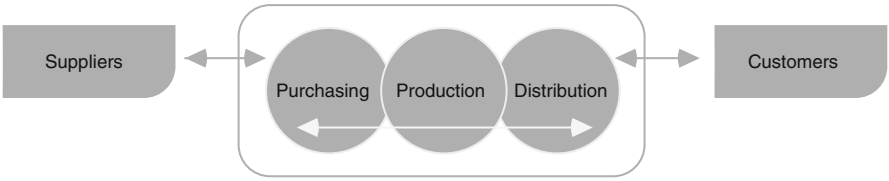


Fig. 4.11 Example of supply chain modeling

Fig. 4.12 Example of IDEF(0) model

Actors in the chain may account for one or more of these subprocesses. When one actor accounts for several successive processes, this is called vertical integration (source: Wikipedia).

4.3.7 IDEF Model

Integration Definition (IDEF) is a function-modeling methodology for describing manufacturing functions that offers a functional modeling language for the analysis, development, reengineering, and integration of information systems, business processes, or software engineering analysis (source: Wikipedia) (Fig. 4.12).

The examples shown here are very small versions of the actual models, which are often so complex that one could easily cover a wall with the schemes.

A commonly used way to create these kinds of schemes is by organizing interviews or workshops with the people directly involved (as already mentioned in the previous paragraph). This generally is a time-consuming activity with a lot of alignment, in both substance and decision making, of efforts of various people or departments. During any discussions a lot of time is lost in a large number of (operational) details and much time is spent debating all the possible exceptions.

Often it is easy to discuss the optimizations with the people directly involved, but the realization of true innovation/changes in this setting is much harder. In addition, it is difficult to get the responsible business manager involved (the owner, the person who bears final responsibility for the process, if this person has already been defined).

Finally, it is difficult to monitor whether and to what extent the developed processes are and continue to be in line with the business goals and business strategy. In short, from an organizational perspective BPM involves much more than just drawing process flows. And so far we have only discussed BPM in more or less traditional organizations, with solid processes, robust partnerships both internal and external, and employees who have sometimes worked at the same company their entire working careers. You can imagine the impact of frequent changes in the external and internal environments on the processes in real time and the way they are described and adjusted on a day-to-day basis.

4.4 Relevance

4.4.1 How Do We Make Process Information Relevant?

Relevance has to do with having information available at the right moment, in the right place, for the right person, in the right form (Fig. 4.13).

After a model has been established and all transactions are mapped out, it is time to actually execute the process. Here we are concerned with the presentation of results and compiling results, custom designed for the end user. Within BPM this is also known as Implementation.

Fig. 4.13 BPM lifecycle, relevance

In the implementation of process information, it is important to check the extent to which it is possible to present the information in such a way that some goal is achieved. The goal is usually to deliver a product or service in the shortest possible time and at the least cost in accordance with user preferences.

Human processing of the data and information is central here. When is information relevant and how is this determined? The human brain filters the available information into useful and useless information.

The purpose of applications that support the activities performed by employees in the form of workflows is therefore to conform with the logic of what is relevant to the organization.

The conditions under which information is processed determines, to a large extent, what is relevant and what is not. Until now, we have looked at BPM from the perspective of a traditional organization with well-defined processes that has as its ultimate goal the optimization of its own business management results.

But besides traditional organizations, there are several other players in a market that play this game under different rules. Think of Internet companies who focus on other areas of the value chain, or the new world of telecommuting where employees no longer need to come to one particular location to do their jobs.

Nowadays, cooperation depends more on opportunity and varies greatly. This has consequences for the processes, the individual steps in the processes, the speed of the processes, and the order in which steps are taken.

The new organization is a network organization. Work is done within the organization but also between organizations. Also, virtual organizations arise where principles such as outsourcing, mobility, flexibility, and teleworking are dominant. The conditions in these new forms of working and working together change so quickly that proper process inventory, much less process optimization, no longer plays a role. The only constant factor in success is that the organization can act quickly and the knowledge that tomorrow will be different than today. The saying "the only constant is change" definitely applies to such organizations.

4.5 Interpretability

4.5.1 How Do We Assess the Value of all the Relevant Process Information?

The final aspect of the human–information–technique interaction that we will discuss is interpretability. Depending on the way information is made available, a tool will be chosen to assess the information (Fig. 4.14).

The result of BPM is first processed by a piece of software and then by a human into information and knowledge (interpretation), which makes it possible to measure a process's effectiveness and to evaluate whether any improvements can be made to it.

Fig. 4.14 BPM lifecycle, interpretability

We gain our knowledge through measurement. Knowing is having all available information in the form that suits the possibilities of the user's working methods. Knowing is essential for taking the right decision, initiating an activity, and forming an overview of the situation (monitoring, control).

The benefit of measurement is that it helps to gain insight into business operations. The questions "How long is the complete product cycle?" and "How long is the cycle supposed to be, according to the design?" should be easy to answer. Any discrepancy in the answers must lead to an evaluation and a proposal to adjust the cycle.

The measurement data also offer the opportunity to monitor the KPIs, identify and correct any bottlenecks, or see where work is piling up and whether it should be given to someone else or to another machine.

In short, measurement data offer the opportunity to intervene in a process, in the short or long term, by starting a new feedback loop.

But what if the measurements from the processes contain large amounts of data that are relevant but still do not give a clear picture within the period in which the decision must be taken because of the enormous amount of data? Think about logging and analyzing (big data) data within processes. In recent years, big data have become a hot item. The performance requirements set by the organizations can in many cases no longer be achieved with the "standard" solutions. This is why we are seeing more and more cross linking between large amounts of process data and business intelligence solutions (of which more in Chap. 6, "Business Intelligence").

Chapter 5
Enterprise Content Management (ECM): The Unstructured Part

Anja van der Lans

This chapter on enterprise content management (ECM) explains the meaning of this concept and describes the types of ECM solutions and the forms in which the information lifecycle occurs in these solutions. It also looks at how the enterprise information management (EIM) maturity model is used in ECM.

Sections 5.2–5.5 tie in with the EIM triangle (Fig. 5.1). The first part in Sect. 5.2, *Availability*, focuses on creating content and preconditions that enable the saving, processing, and retrieval of content in ECM environments. In Sect. 5.3, *accessibility* is discussed which falls under the umbrella term *relate*. Relate means linking internal relations between content. In Sect. 5.4, *relevancy* is discussed under the term *assemble*, a different way of presenting and composing results by conveniently restructuring existing content to create new content. Lastly, the results of the collection of content are processed into information and knowledge (*interpretability*) that one can keep for oneself but that can also be shared with others.

5.1 Concepts

5.1.1 What Is Enterprise Content Management?

ECM is the umbrella term for managing streams of content from its origin (unstructured information components) until the moment it is produced or published.

The concept of ECM describes the strategy, methods, and the set of instruments that are used to collect, manage, save, store, and publish/unlock content (unstructured information and documents) related to organizational processes. Among other

A. van der Lans (✉)
Advisory Services, Incentro, De Meern, The Netherlands
e-mail: Anja.vanderlans@incentro.com

P. Baan (ed.), *Enterprise Information Management: When Information Becomes Inspiration*, Management for Professionals 2, DOI 10.1007/978-1-4614-5236-2_5, © Incentro 2013

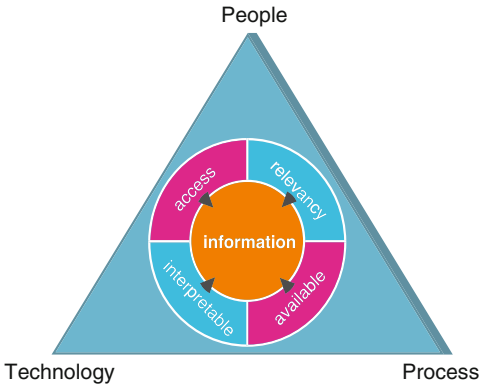


Fig. 5.1 EIM triangle

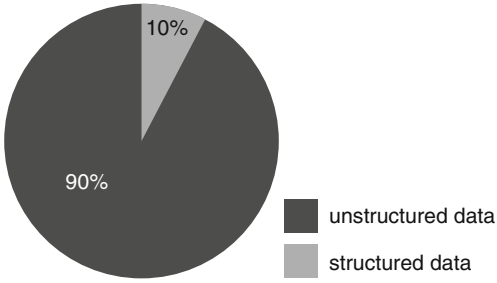


Fig. 5.2 Structured and unstructured information within an organization

things, ECM covers document management, digital asset management, records management, Web content management, collaboration, and imaging.

The definition of ECM explicitly states that ECM entails the use of unstructured information, which is information that is saved not in a data structure with definite form but rather in a text or any other medium. Figure 5.2 shows that the overwhelming bulk of information available within an organization is unstructured. The percentage of unstructured and structured information will differ by organization. Generally speaking, decisions are made on the basis of structured information in databases, putting unstructured information in second place.

Why, then, pay so much attention to ECM?

ECM enables the unambiguous recording of unstructured information (mainly documents) coming from organizational processes, preferably at a central place that is easy to search so that information can easily be retrieved.

Because of the rapid growth of unstructured information, documents are less surveyable and different versions of documents occur. This results in inefficiency and, possibly, unnecessary additional cost for the organization for redoing certain work.

An important element in ECM systems is the retrieval of saved information. Every ECM solution has a built-in search engine. This search engine is usually a derivative product (OEM) of those on the enterprise search engine market. We will

not go deeper into searching in this chapter. All details of searching will be discussed in Chap. 7.

The Association for Information and Image Management (AIIM), the umbrella organization for this field of study, states the following about ECM on its Web site (www.AIIM.org):

> Enterprise content management isn't about technology. ECM means being able to get control of your business processes. It means improved efficiency and reduced cost. It also means you can satisfy statutory and regulatory compliance requirements and be in effective control of your business. In the end it boils down to improved financial performance and genuine competitive differentiation (AIIM 2010).

All these elements of ECM will be discussed in the following sections.

5.1.2 Enterprise Content Management

As mentioned previously, ECM is a field of study in which unstructured information is central. Unstructured information is everything that is not "data," in other words, it is not stored in a fixed structure in a database. The number of files that are considered unstructured is limitless (e.g., Microsoft Office files, Web pages, images, audio files, video, e-mail, PDF files) (Fig. 5.3).

ECM is used as a framework for document management, digital asset management, records management, Web content management, collaboration, capture and imaging, enterprise information portals, and workflow management. The ECM area is broad and deep, meaning that many different elements and solutions are offered on the market. Each element has its own terms and principles. The elements will be discussed briefly.

5.1.3 Capture and Imaging

Literally, this means capturing and making an image. In practice, it means converting a paper version of a document into a digital version. By creating images of documents, it is possible to save them in an ECM environment. The term includes techniques such as processing, compression, saving, printing, and displaying images.

Many organizations have vast amounts of paper archives. Once digitization is considered, the idea of wasting paper and of diminishing the paper usage to a minimum comes up. Digitizing paper documents as such is not too difficult; nowadays every photocopier has a scan option. But when it involves multiple papers, different file formats, and not only an image of the content but also making the content searchable, this can become a project in itself. Even more, this could simply become the most expensive part of a change project, and in most cases it has a profound effect on employees' perception of change.

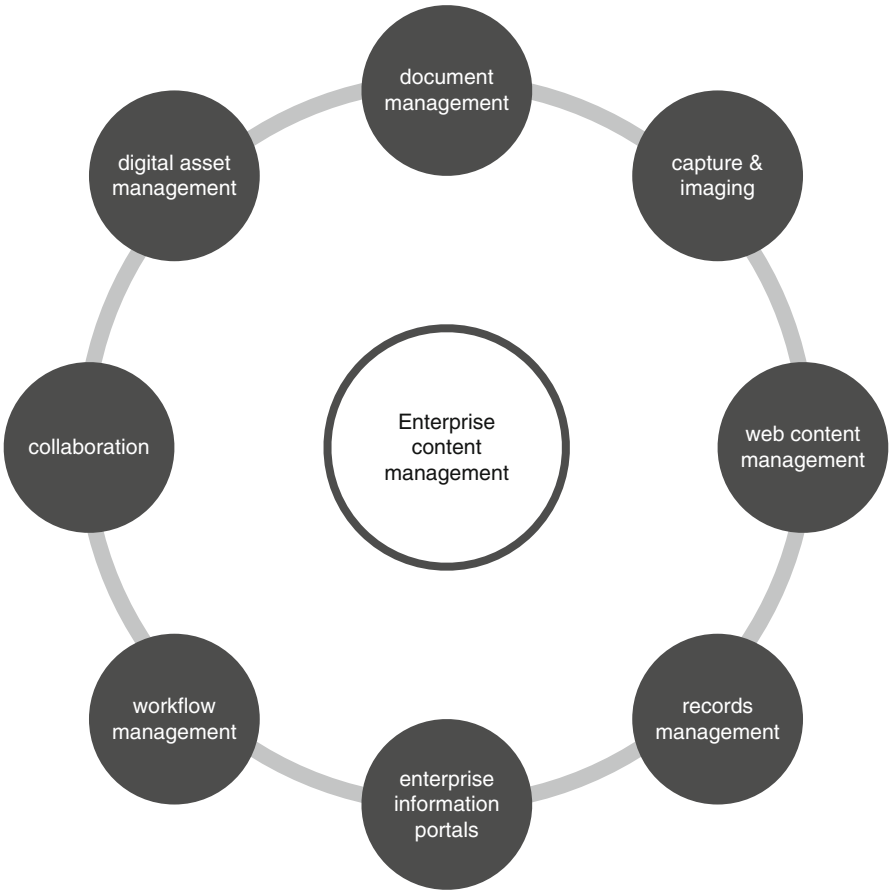


Fig. 5.3 Elements of ECM

Converting paper documents into digital variants is called substitution. Substitution means replacing documents for reproductions. The original document is then destroyed, so the reproductions fully replace the originals.

As mentioned previously, converting paper documents into digital versions can be a time-consuming, expensive, and massive operation. In addition, the result of scanning, the digital version, will not always be machine readable. Depending on the quality of the document (too much use of color, handwritten texts, and "faded" ink all impede recognition) a recognition rate of 95–98% is commonly reached.

The digital version of a document will often be saved as a PDF file. Again, several formats are imaginable. The PDF is more or less the standard for the long-term, and therefore sustainable, storage of documents.

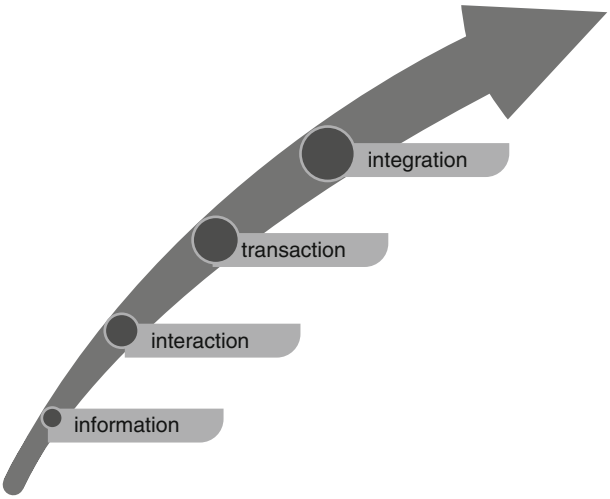


Fig. 5.4 Stages of growth in Web site applications

5.1.4 Web Content Management

To many organizations, Web content management (WCM) and content management are virtually identical. Content management in the definition used here involves more than just WCM. WCM is that part of a field of study that unlocks content through Web technology.

There are three levels of Web applications, each with its own target groups:

- Intranet,
- Extranet,
- Internet.

An intranet is used to reach groups that are internal to an organization, essentially the employees. An extranet is meant for visitors with whom the organization has a special relationship. These Web sites are not accessible to others; login codes are required to enter them. The Internet is used to share information with anyone interested in the company. However, the environment should be well managed. The early release of information or even false information may have major consequences for the organization. Some of these risks can be reduced by the use of a Web content management solution with a workflow management system. The content of the Web site is related to the stage the organization is in with respect to its use of Web technology.

Gartner (2007) acknowledges four stages of growth (Fig. 5.4):

- Information: the Web site only gives information. There is no way to interact with the organization. Interaction takes place over the usual channels such as telephone, fax, or face-to-face meetings. This type of application was commonly

used in the early days of the Internet (1990s) but has now almost completely disappeared.
- Interaction: Web sites expanded, and this expansion spurred a growing need for the capability to search for specific information and to react to that information. Search functionality and e-mail were added to information Web sites. Direct communication with the organization was now possible. By adding interactivity, an organization could reduce its costs (lower occupancy at the customer service department) and add value to the site for its visitors.
- Transaction: the addition of transaction applications allowed visitors to complete a transaction using the organization's Web site. Examples include buying a book, traveling, and downloading music. Because of the addition of transaction capabilities, the Web site became, besides a cost-saving element, a possible source of extra revenue. To prevent customers from being misled, security standards and certification marks were developed.
- Integration: integration makes it possible to process data from a Web site directly to the underlying systems, transforming the process into a completely atomized one. Manpower became unnecessary to enter data from the Web site into the underlying systems, and the 24-h economy could be fully supported. This applies to Web sites like Amazon.com, Bol.com, and many others.

5.1.5 Record Management, Archiving, and Document Lifecycle Management

AIIM states the following on its Web site[1] under the heading 'Managing the Information Management Lifecycle' (AIIM 2010):

Records are nothing new. They have always been a necessary and required aspect of business. However, the arrival of the digital age has raised some unique issues. The volume of information being created and circulated has exploded. Within this deluge of documents, messages and agreements are items whose importance might require them to be raised to the level of a record. Once raised to this state, they need to be preserved for an appropriate period of time, and then disposed of. Traditionally, these records would be created and stored on paper. Now they are almost certain to be created electronically, and increasingly they are being stored electronically for their lifetime. In addition, there is little point in preserving a record if it cannot subsequently be found if at some point it needs to be referenced.

Record management (RM) used to be defined as the management of records that were no longer used in daily processes but still had to be preserved. These are the semicurrent or inactive records in the archives of an organization. In a modern view, RM refers to the total life cycle of a record, from the moment it is created until the moment of destruction.

[1]www.aiim.org

> ISO established the following definition (ISO 15489–2001): "Record management is the field of study that is responsible for the efficient and systematic management of creating, receiving, maintaining, using and destroying records, including the process of managing the evidence of and information about business activities in the form of records."

A record is defined as information that is maintained as evidence and information for an organization or person in relation to legal obligations or transactions of an organization. In the lifecycle of documents, their storage in an RM environment is its final stage, the end of the cycle. The possibility of recording the lifecycle of a document in an environment is also advantageous for determining and dealing with a specific workflow.

Organizations either must, because of government regulations or compliancy, or want to, for reasons of evidence or process improvement, decide on the maximum retention period for documents in ECM before the ECM solution is brought into effect.

Each organization will decide differently, but generally speaking one could make a list starting with management/board documents (preserve) and ending with restaurant menus (don't preserve). Thinking about and agreeing on these terms results in a so-called retention policy in which guidelines for storing as well as destroying different kinds of documents (paper and digital) are laid down.

Once the policy is clear, it can be put into the RM solution, so automatic destruction can take place (if you wish, of course). This is the only way to prevent the unlimited growth of document storage in an organization.

5.1.6 E-mail Archiving

Up until now, this chapter has principally discussed meeting legal requirement, guidelines, and standards concerning paper documents and digital documents. Of course, this also applies to the storage, preservation, and retrieval of e-mail. E-mail is getting more and more important in daily communication between people and growing exponentially; therefore, e-mail archiving is becoming more important.

While storing e-mails, organizations come across limitations and obstacles to storage in the form of PST files, access to old e-mails, and a lack of backup discipline.

For a few years now, e-mail has been accepted as evidence in court. The data retention law applies to both e-mail and paper documents. Companies that are unable to submit e-mail data when summoned can face charges and fines that can amount to substantial sums of money. Companies should therefore have a tool to fulfill this legal obligation. Reliability of e-mail storage is essential.

The principle of reliability is broken down into four components: authenticity, integrity, confidentiality, and availability.

Reliability of e-mail evidence influences the storage of messages in a digital archive system. After e-mail messages are entered into the storage system, the authenticity (is this the original message?) and integrity (is this message complete

and not damaged?) of digitally stored messages must be ensured. This places demands on the protection of the archive against unauthorized access and viruses. In addition, e-mails should remain readable (sustainability). Finally, the system should have a logbook that automatically records all changes to the system with respect to stored messages, such as, for example, deleting, indexing, and copying of messages (audit trail).

It is obvious that confidentiality of e-mail traffic is important. Accessibility to stored e-mail by unauthorized users puts enormous pressure on reliability.

5.1.7 Workflow Management

Workflow management (WFM) is a complete or partial automation of the business process in which documents, information, and tasks are transferred from one person to another following a set of predefined rules. Putting it all in a fixed order is only possible when the process is completely predictable.

Many kinds of processes lead to many forms of workflow. Four types of process flow are recognized:

- Group related: aims to have people work together toward a common goal. Usually, only deliverables are defined. The choice of process to get there is discretionary. The Internet and messaging play an important part.
- Ad hoc: users can define and modify processes themselves. This kind of flow process can respond to changing circumstances and is therefore very flexible. In this context, compliancy and security deserve special attention.
- Administrative: the movements of forms within an organization determine the process. Many people are part of this process.
- Production: automation of tasks is emphasized. The organization tries to achieve the highest possible productivity.

In implementing workflow, it is often pointed out that it would be useful to be able to check the status of a document in the flow. In other words, how far along in the process is this document. Based on the predefined lifecycle, transition of status can be realized by document during each step in the process, which eventually leads to a status of "completed."

This document lifecycle has quite a different meaning than the one mentioned previously, in which the focus was more on the status of saving or destroying. Both can be part of an ECM system.

5.1.8 Teams and Collaboration

Collaboration is a process whereby many people work together toward a common goal by sharing knowledge, learning, and reaching a consensus. Collaboration is supported by a specific application called groupware.

Groupware (also known as group support systems or collaborative software) is a kind of software that facilitates the reaching of a common goal by a group of users. Collaboration tools have become very popular, especially in the last few years, since they make group work easier. Learning and sharing are supported, and more and more social network functionality is being added.

It is slowly becoming clear that software used on the Internet to exchange information is integrated into collaboration tools. A wiki is a perfect example of this, but there are other software packages for brainstorming sessions or, in a broader sense, e-mail and joint calendars or blogs, instant messaging, Twitter, social bookmarking, and so forth. Personal profiles and contact options as seen on Hyves and Facebook are available as parts of the dashboard of collaboration tools.

Collaboration tools are increasingly being used to supplement or replace document management systems. Many collaboration tools build in basic functionalities of document management systems in their applications such as version control, processing, and authorization.

EIM is all about using information to outdo the competition. Social networks are a new way to use information within an organization. A social network is a group of individuals communicating about a certain topic. In practice, EIM and social networks affect each other (Kooij 2009).

Knowledge workers use a professional network of likeminded people to exchange knowledge and information. A social network helps to build a network. The technique makes it easier for people to keep in touch with digital contacts. Social networks allow for rapid spread of information. Twitter is a good example. Social networks can lead to the success or failure of a product, service, or event. Because of the speed with which information is spread globally, this is an important element in the information supply of knowledge workers. Knowledge workers should be given the chance to provide themselves with the necessary knowledge through these channels.

5.1.9 Digital Asset Management (DAM)

Digital asset management (DAM) consists in recording, annotating, cataloging, saving, and loading digital sources that carry little or no proprietary metadata or text like photos, videos, or music. This is also called media asset management (MAM).

DAM also refers to the protocol for downloading, renaming, rating, grouping, archiving, optimizing, maintaining, and exporting files. There is a target version of a digital document, referred to as the essence. This version has the highest resolution (highres) and the highest representation possibilities. Depending on the demands placed on the representation, copies with lower resolutions can be derived from the essence (lowres, midiumres, and thumbnails). Besides the size of the file (based on resolutions), conversion of a file extension is also important. These so-called renditions are used for further processing or publication in other media or (storage) environments.

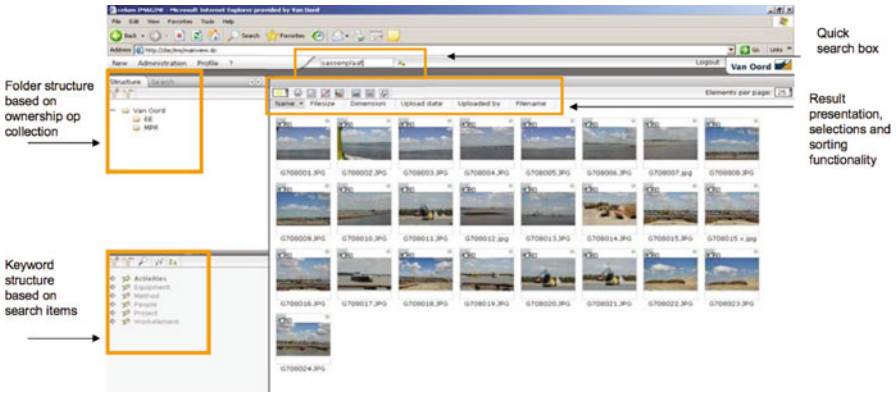


Fig. 5.5 Presenting results in a DAM

Obviously, having metadata about digital assets is crucial. Without metadata it is almost impossible to retrieve a file, making it unsuitable for reuse. Each file comes with a number of data fields that are stored with the file so it can be retrieved, but it is also possible to add extra fields for subject, title, classification, and so on.

Organizations that purchase DAM systems usually buy or receive the rights to files (as part of a publication or by hiring a photographer or videographer). These rights are valuable and should be registered and protected. Registering and managing these rights are also part of the DAM system (Fig. 5.5).

There are four types of DAM systems:

- *Brand asset management* systems focus on the facilitation of content reuse within large organizations. As the name implies, such systems mainly aim to make everyone in an organization use the same house style outside the organization. Content is largely marketing or sales related (for example, product imagery, logos, and fonts).
- *Library asset management* systems focus on the storage of large amounts of media, for example, in video and photo archiving, and making it available to others.
- *Product asset management* systems focus on the storage, organization, and version control of frequently changing digital sources in a production area, for example, managing CAD drawings of installations, factories, ships, and so on used to create two- and three-dimensional images.
- *Digital supply chain services* are used to push digital content out to, for example, digital retailers.

5.1.10 Document Management

All organizations that want to give substance to managing documents use document management (DM). Documents in a broader sense are very close to the concept of content. Nevertheless, the focus of DM is more on internal documents than on

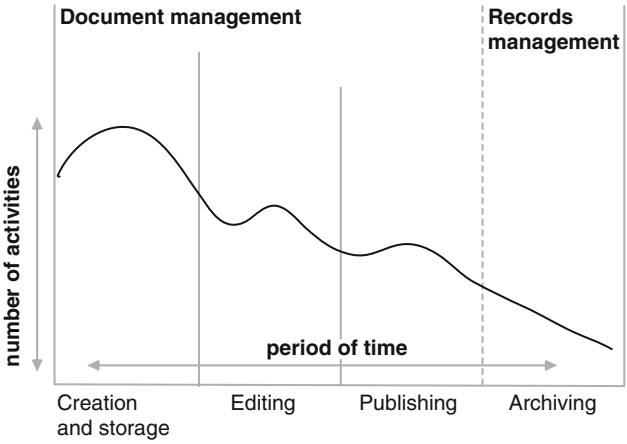


Fig. 5.6 Graphic representation of document lifecycle

content management, and content from various sources will always be brought together in a document (digital or physical).

DM has the following characteristics:

- Check in/check out: editing a document while it is still available to others; the copy is temporarily made available to the editor;
- Version control: the latest version is shown, earlier versions remain available, and differences between versions can be made apparent;
- Layout: a viewer shows what the document looks like with respect to, for example, different layout designs.
- Searching: documents should be retrievable based on content and metadata;
- Metadata: characteristics of the document are stored, like keywords, author, title, date of creation, date of editing, and so on;
- Authorization: defining who can do what with a document prevents unauthorized usage.

DM is also about:

- Where to store documents – in a centralized or decentralized location?
- How to store documents – with or without metadata or hierarchical characteristics?
- How to retrieve documents – with or without specific indexes, by searching or also by navigating?
- How long to retain documents?
- How to guarantee authenticity – through an audit trail or otherwise?

In this part of ECM it is once again important to set up the DM system in accordance with the document lifecycle (Fig. 5.6) and the intended workflow.

In particular, the document creation stage, storage, editing, distribution, and publishing are part of DM. A selection must be made from the collection of documents based on retention periods. Documents that are slated for long-term storage are entered into the RM system.

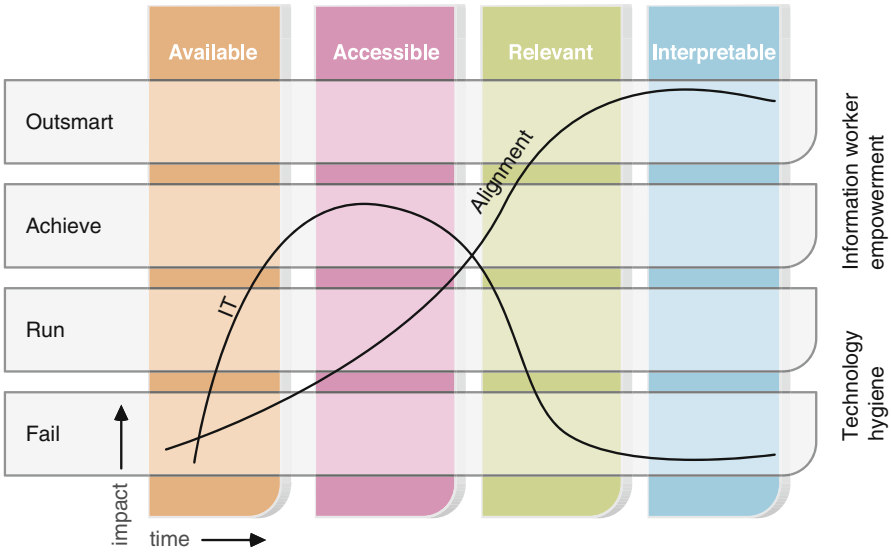


Fig. 5.7 EIM maturity model

5.1.11 How Does the EIM Maturity Model Influence ECM?

The term ECM covers many different types of information flow, each with its own problems and solutions, but it can also be subdivided into levels of development (maturity levels). Each level has its own specific characteristics and demands a certain approach to reach the next level (Fig. 5.7).

The operation of the EIM maturity model is described in Chap. 3.2.2, but what are the characteristics for ECM or, in other words, how does the model affect ECM?

5.1.12 Fail

Information is stored in folders by department on one or more file stores without uniform structure. Each department chooses its own structure and determines access to the folders. Information is double stored; navigating through the folders is the only way to retrieve information (availability).

5.1.13 Run

Storage in DM and (Web) content management becomes more important for certain departments. Applications are purchased to achieve proper storage. Templates for making common documents are created, and use of the templates is mandatory (accessibility).

5.1.14 Achieve

The transparency mentioned in the previous stage clarifies what needs to be done in an organization. If the information found by retrieval shows that activities are performed in suboptimal or superfluous way or if they are done twice, this will lead to efficiency measures being taken (relevance).

5.1.15 Outsmart

The organization supports initiatives that facilitate the exchange of information between departments, employees, and clients and allow the exchange to be carried out in different ways to suit all tastes. DMS is replaced by a portal or suite, also known as collaboration tools, generally with 2.0 tools like wiki, blogs, and so on. Acting quickly to gain a competitive advantage is paramount (interpretability).

5.1.16 Processes

As unstructured information results from unstructured processes, the number of requirements that are set for a certain activity will vary. In the following paragraphs we will investigate how processes and the roles people play in those processes affect the activities of unstructured information.

You might think that ECM involves a lot of tools and processes. That is true, but every tool has a specific role to play in the lifecycle of information. Major suppliers of ECM solutions focus on increasing the completeness of their offerings and call this ECM suites. What does the lifecycle look like when a suite is fully supporting it? Instant information from blogs, chats, and e-mail are usually irrelevant within an hour or day after creation. The running time of content could be anywhere from seconds to permanent storage.

When information is shared in a collaboration tool, announcing this usually has a practical value of 1 week to 1 month. Publishing, on the other hand, has a cycle of 1 month to a year, and then content management and document management can help in monitoring the process. In some cases it is useful to be able to retrieve documents for reference or tangible evidence of the progress of a process or historic event. A record management system (RMS) helps to reconstruct events or the course of events. Finally, there is always a limited set of documents stored permanently, usually with very restricted access. Examples are deeds of foundation, statutes, distinctions, minutes of important conversations, and so on.

5.1.17 Why Would an ECM System Be Desirable?

In this chapter we have investigated the forms in which ECM occurs and we have looked at the functions and functionalities of the various solutions. The diversity in supply alone shows that there is a need for solutions. An organization often searches for a reason for or makes a business case to justify a solution. The most common reasons given in a business situation are as follows:

- Compliance: legislation requires organizations to provide insight into their documents and meet the demands of future checks, for example, Sarbanes-Oxley Act (SOX), environmental law, or personal data protection.
- Efficiency: process improvement, availability of information. Organizations store documents in different places, use several versions, and do much work redundantly or simultaneously in different departments; a common platform for document storage and availability makes this less likely to happen.
- Consistency: many organizations want to determine how they are viewed by the outside world and therefore prescribe one way of implementing internal activities. Employees know exactly what is expected from them. Accurate recording in an ECM system can be helpful.
- Customer service: the quality of the helpdesk and good customer service are becoming increasingly important in customers' experience. Organizations are trying hard to solve customers' problems as quickly as possible, preferably by phone or e-mail. This requires up-to-date, relevant, and correct information. ECM supports this need.
- Archiving: when an organization has cabinets full of paperwork, a change in location can often provide a pretext for selecting, reorganizing, and digitizing documents. Reduction in storage costs can then be a reason to purchase an ECM system.
- Consolidation: organizations that are merged or taken over often use different systems. This can cause many problems. A new ECM solution can improve efficiency and, at the same time, make life easier for new employees because they will learn to share knowledge and be able to work well together in the organization (a form of knowledge management).

In ECM as in document management systems the focus is on quality improvement, improving efficiency in processes, and satisfying laws and regulations, whereas different reasons often exist to work with Web content management. These include the following:

- Cost savings: setting up a Web site correctly and delimiting roles and responsibilities clearly can save a lot of money.
- Income generation: a good Web site attracts customers and keeps them coming back to place orders; this represents one more way to generate additional turnover.
- Image creation: showing what the organization does and what it stands for leads to transparency, and the outside world's confidence in the organization will grow.

5.2 Availability

Under ECM, availability means making resources available to others besides oneself. If information is to be stored only for oneself, an ECM system is usually unnecessary.

Content can appear in many different forms, but the reason for its existence is usually the same: to carry out a process or activity in an organization for which an employee is responsible. Content is not a goal in itself but, rather, the result of a step in a process. The byproducts of content, in the form of documents or pages with information, are saved in computers.

To enable cooperation among colleagues, efficiency is fostered when employees can see what their coworkers have done within a given process and the knowledge they have gained (target group). We can speak of sharing knowledge and experiences. The sharing of experiences causes a target group to do or imitate something. This broadens the target group's knowledge. It is tempting to do things as prescribed or defined by some content. Content leads. In an organization, content will often appear in the form of, for example, reports, manuals, work instructions, system documentation, and product and service descriptions. This content will be used both inside and outside the organization, so different target groups will use the same kinds of content. Nevertheless, product descriptions aimed at clients will differ from product descriptions aimed at vendors or technical assistants. It is therefore important that content be in line with the expectations and demands of the target group.

Before content is created, the following questions should be answered:

- What is the goal of content?
  The goal of content is to be a reflection of a process, to set down and share knowledge, or to be saved for another time. Either way, rules should be established about, for example, where to record the content (usually for each kind of content), what metadata should be recorded for each kind of content, and the structure of the different versions. These rules form the basis of the further design of ECM or DM systems.

- Who is going to use the content?
  It makes a huge difference whether the content will be used only by you; the department, group, or team that you work in; the entire organization; customers and suppliers; or by anyone who wants to. In all cases it is important to realize that content should be made available when it needs to be viewed, used, or edited. To bring this about efficiently, it is essential to add certain data to the content that will make content retrieval easier. It should, for example, at least have a title that covers the content and a logical storage place and structure.

- How can and may the content be used?
  The question as to who will have access to the content and what actions may be performed with the content is settled depending on the user and confidentiality of the content. This could be nothing (no access), read only, edit and save under a

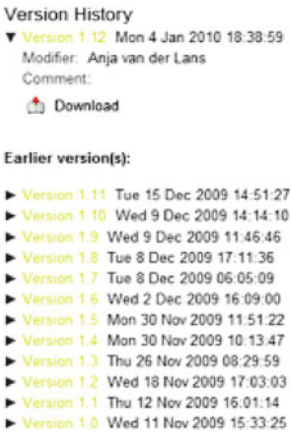


Fig. 5.8 Survey of available versions of a document

different title, or edit and save as a new version. All these variants indicate that careful thought should be put into the authorization on documents. A securely built authorization table with all possible variants of rights for each type of document is needed to ensure that the right user gets access to the right information at the right moment.
Three elements characterize availability in DM:

- *Check in/check out*
  When a user wishes to edit a document, it can be useful not to use the original but a copy (check out). All other users can still see the document, but they cannot edit it simultaneously with the first user. Once the first user has completed a new version of the document, he or she puts it back (check in) and the new version is visible to others.

- *Version management*
  In many organizations, uncertainty about versions is the reason to consider using DM. If documents are not stored at one central location, chances are that different versions of documents will circulate, end up in different places, and become uncontrollable. Version management requires agreement regarding the extent to which editorial changes will be shown in subsequent versions (Fig. 5.8).

- *Layout*
  A document can be created in, for example, Microsoft Word. This is an editable file type and therefore unsuitable for external use. The same document is then used to make a PDF file. Sometimes another format is needed for another target group. All these so-called renditions are stored in the ECM system. Renditions are documents in similar versions but in different formats.

  In a way, saving several copies of one document is contrary to the idea behind ECM systems. After all, ECM systems are used primarily to create uniformity and clarity, where the refrain is "save documents only once and in one fixed location." Some flexibility is sometimes required to use an ECM system. A good

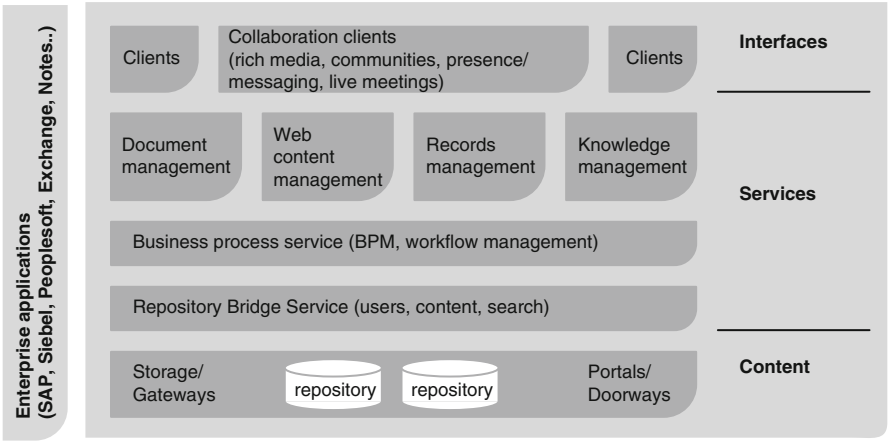


Fig. 5.9 ECM architecture (Open Text Corp. 2005)

system will support organizations in their quest for flexibility (for instance, by allowing several documents to have the same name and version in different renditions).

Other layout applications within ECM solutions are also examples of this flexibility. Templates for the layout of documents, linked to or included in ECM solutions, monitor the layout of documents. The same goes for the use of styles in a WCM solution, in which colors, fonts, font sizes, and page layouts are fixed, which leads to a balanced and consistent look on the Web site.

5.2.1 Database

When content has been created and links added, it is stored in a database. The architecture of the database defines the connection between all elements of an ECM or a WCM system.

Figure 5.9 shows a more or less standard outline of an ECM suite architecture consisting of three layers: content, services, and interfaces. Content is the layer in which the content is stored in databases (repositories). A gateway serves as an entrance gate to another network, in this case to systems for content storage. A doorway to other portals is also possible from here.

The second architecture layer maps out the services. In this image, there is a large diversity of ECM components such as DM, WCM, RM, and Knowledge Management (KM), which are built on a workflow application that is available for all components. There is also an element called repository bridge services that connects the content layer with the services layer.

The interface layer allows users to get access to the various service components in a form that suits them (taking into account, for example, authorization, profiles, and settings).

Lastly, on the left side of the diagram a possible link with enterprise applications is drawn that must ensure that users get access to content or metadata from, for example, Enterprise Resource Planning (ERP) systems.

5.3 Accessibility

Once documents and information are stored in a DMS, they are available to everyone with access rights. Accessibility can be expanded by adding search functionality and metadata.

- *Search functions*
  Content stored in a central storage with metadata has greater retrievability than content stored without metadata. A document management system (DMS) is equipped with a search function. This is often a scaled-down version (OEM) of the search engines that can be bought on the open market for enterprise search engines. The exact meaning of enterprise search and what functionalities are available will be discussed in Chap. 7.
- *Metadata*
  Metadata have a special meaning in ECM. They mean not only "data about data," as is common in business intelligence (bibliographic or formal metadata), but also data are added to the contents of contents (content-related metadata). They are, so to speak, the dimensions of a document or other content item (Fig. 5.10).

By metadata we mean all information that says something about the content. This could simply be a title, authorship, date, or version but it also might be keywords, a summary, document status, or any other specific characterization. What is important depends on the organization and user group. Nevertheless, some standards have been developed for metadata.

The most common standard is the Dublin core metadata element set. It consists of 15 optional metadata elements that can arbitrarily be repeated or omitted.

1. Title	2. Contributor	3. Source
4. Creator	5. Date	6. Language
7. Subject	8. Type	9. Relation
10. Description	11. Format	12. Coverage
13. Publisher	14. Identifier	15. Rights

The application of metadata for unlocking document content is called classifying or tagging.

- *Authorization*
  The importance of setting the rights on documents was mentioned previously. Some information may not be edited, while other information should be updated regularly or it may only be edited by a select group of people. These kinds of

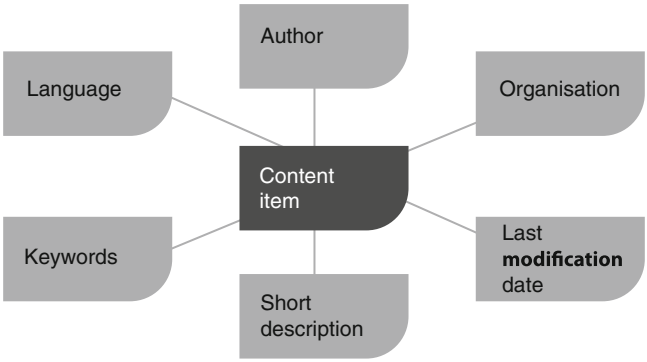


Fig. 5.10 Metadata in content management

agreements are laid down in an authorization table that is used to anchor the rights in a system. In many cases it is useful to take as a starting point the access rights on systems and environments that the organization has already established in the form of active directory (AD; user administration of the ICT department) or LDAP (lightweight directory access protocol). These are network protocols that describe how to approach data from directory services.

Besides general access to an ECM system or parts of the system based on who uses it in the organization, the document type can also determine access. A medical document, for instance, is only visible for the treating physician and the patient herself. Others, like the director of the healthcare clinic where the doctor works, have no access to the documents. With certain "type" classifications for documents, authorization and protection are established, and in some cases a process in a workflow application is also (automatically) initiated.

- *Linking*
  Linking a document with its related components is central in the *accessibility* (Web) content management stage. Some examples are linking mail files to their attachments, defining the relationship between different versions and renditions of a document. In addition, links are made to the (original) location of a document (digital or nondigital). This process is called linking or connecting. Especially in WCM, this is a commonly used term. Making a link or reference is called linking.

5.3.1 Link Management

Content is fixed in objects. The way content can be approached is determined by so-called links. The place where a link leads to is indicated by an "anchor link." The location indicated by an anchor link is called an identified element.

Another way of linking is by using positionable elements. In this form, the link refers to a position in an object. This will only work when you are sure that the contents of the object will not change. If any words or sentences are edited, removed, or added, then the link will no longer lead to the original sentence. Results become unpredictable and are recognizable as “page not found” or dead link/broken link reports on the Internet.

Links create logical connections between separate objects (parts of content). Link management is an important element in managing content. When a content object is removed, the referring links must also be adjusted to avoid undesired results (broken links).

A data model is designed to enable document editing and to make sure that documents can be retrieved easily. Link placement is important for information reuse. A link is a kind of reference to places where more information on a subject can be found.

When we talk about content creation, we usually mean creating documents. But of course, Web content is also created. Besides creating content, managing relationships in the form of links is also important for Web content.

Link management allows for monitoring the integrity of links between available pages. Problems can occur at several points: when editing an object or link, when removing an object, or when removing a link.

Editing a link usually does not lead to integrity problems because both the link and the object(s) have to be available at that time. The relationship should remain intact. When an object is modified, the anchor link can be removed so that the link no longer has a reference and an error occurs when it is used. An object has no knowledge of all the (monodirectional) links that refer to it. Undesired results can therefore not be ruled out. Documentation on the applied links and rules for their use significantly reduce the number of times that they occur.

When an object is removed, the same problems occur as when it is modified.

When a link is removed, a separate page can be created (orphan). This means that the page is no longer accessible and a lot of information is lost. Again, documentation can eliminate part of the problem. Many WCM systems have a dead link or broken link checker that regularly checks the links and locates orphans or unavailable pages, so it can be managed.

5.3.2 Web Content Management

WCM is often seen as the environment for managing Web pages. Content itself is irrelevant. In contrast to this view, only content is important in the framework of EIM and is referred to as such when WCM is mentioned in this book.

A WCM system often uses an object model. This is a predefined model, and content is provided by entering it through the user interface content management system. The connections between the objects must be known in order to unlock the data in the system. Depending on the information needed, the contents of the table are uploaded in a data warehouse or stored in a content repository.

Uploading, for example, the content of a Web site is strikingly different from most sources of operational processes (OLTP). That is how HTML is used in attributes. Two fields of the table are included to insert the contents of an article. For convenience's sake, we assume that only text is inserted. Although the HTML code contains useful information (such as font size, color, and so on), it is primarily intended to put the contents of the messages into a central storage location. Besides content, sufficient metadata must be supplied to be able to execute analysis, things like the validity of the message, the title of the message, the author, the date/time combination when it is published or edited, and who has access to the document (everyone or only a restricted group).

5.3.3 Document Management

There is often a relationship between documents in a DM system. When it involves e-mail, this could be a relationship between the message and the attachments or a document that is referred to in another document. In all cases it must be determined that such a relationship exists. In most DM systems, this is done by filling in the name or names (titles or unique numbers) of the documents involved in the related field.

5.4 Relevance

Whether a stored document is relevant depends on the situation. Each user will try to find the document by searching or navigating in the ECM system. This part of relevancy is covered in Sect. 7.4. In ECM, make document information can be made more relevant by assembly.

Assembly entails editing and merging content to create a new document, a new composition of pieces of content, or another form for existing content. After content is created, stored, and provided with relationships with other documents, it is time to see how all the components of a document can be brought together. This is called assembly.

A content item can play a role in zero, one, or more published documents. In other words, the content may not have been published (yet), it may be intended to fill, for example, one Web page, or it could end up in several publications in many ways.

Assembly takes place at different levels in an organization when reports are made that include images, graphs, references to Web sites, and so on. This has a relatively big impact within WCM because the layout is being looked at after the creation of content. Mainly the structure of a publication is determined during the assembly process. For a Web site this is the structure of the site (along with the navigation options) and the structure of page types. All sorts of extras, like contents, registers, site maps, and so forth, are also determined here.

In the design of the content and to add metadata, the demands imposed by assembly in the form of structuring are taken into account. The primary focus is on the granularity of the content (coherence and connections between components). These are chapters, sections, and paragraphs in the sense of visually observable coherence, but also other forms that are more common in ECM (including relationships as in links).

Functionality and design are added to files in the form of scripts and style sheets, nowadays mostly defined in XML. This can be complex depending on the format of the content used in the publication. An example is a Word document that needs to be converted into another format like HTML, SGML, XML, and the like.

Assembling paper output based on digital content is a specific area of expertise. Digital content consists only of content objects and links between them. A link that jumps to a related subject can be translated as a reference. A link that provides deeper insight into a certain concept can be translated by inserting text at the starting point of the link. So in every situation solutions can be found for the proper implementation of a link. The dynamic character of the link must be translated to a static form at that moment.

In modern ECM systems, the environment is equipped to properly store content, so the possibility of reuse is optimized.

Reuse occurs in the following forms:

- Making a new product by reconstructing the whole file using parts of a document;
- Converting one format into another to reach a different public or to guarantee the spread and protection of information (for example, Word files that are converted into PDF to prevent editing);
- Publishing parts of the content for internal use and parts for the Internet.

All these forms occur, and the possibilities are increasing since the standardization of exchange forms is also increasing. The possibilities for assembly of new and reused products will become more and more important.

5.5 Interpretability

Collection content is processed into information and knowledge that one can keep for oneself but that can also be used to share with others. To be received by others in a proper way, the information must be understandable in the right size and format, at the right time, and in the right place (storage medium). Then the information can be interpreted correctly, which can lead to new ideas and innovation.

Sharing is spreading data, information, or knowledge among interested people. After pieces of content are created, stored, linked to each other, and assembled into a new product, it is time to share them with others. Sharing can mean that the author offers the content to its users (sending: push mechanism) or that the user must gain the content himself (collect: pull mechanism).

Examples of sharing are as follows:

- Sending links to colleagues,
- Activating alerts,
- Making an RSS feed,
- Publishing content (determine the version for use by employees or on the Internet by making it public).

Basically there are two ways of sharing information: pull or push.

5.5.1 Sending Links to Colleagues (Push)

Instead of sending a document as an e-mail attachment, a link containing a reference to the location of the document in a database can be included. This reduces data traffic and the risk of scattered storage in separate new versions.

5.5.2 Activating an Alert (Push or Pull)

Many ECM systems (especially collaboration tools) offer the possibility of setting up alerts or alerting when a document is modified or when the content of a folder has been changed (adding, removing, creating new versions). The system saves rules for sending e-mails to the person who has set up or sent out an alert and keeps a record that a user can retrieve in a special folder, directory, or section of a Web page. Examples are my recently edited documents or *Recently added* list.

5.5.3 Making an RSS Feed (Push)

The really simple syndication (RSS) protocol has become a popular standard for achieving syndication, in other words controlled distribution, of content. An automatic push from a system to users is called a feed. RSS feeds can be read by several popular news readers and feed readers, recognizable by a logo.

RSS is mainly used for weblogs, forums, or news sites to keep informed of the latest articles and news. Weblogs and forums are kept up to date with specially developed publication software. This kind of software generates not only regular (X)HTML output but it also quite simply, or even automatically, RSS feeds.

RSS feeds (XML files) are mostly generated by the publication software of the person who maintains a site. This form of publishing is not only restricted to people who make weblogs, the so-called web loggers. Major publishers like the *Washington Post* and *New York Times* publish their copy not only in newspapers or on Web pages on the Internet in HTML but also in RSS format.

RSS feeds can be read using special RSS readers, but they can also be read through special Web sites. RSS readers (pieces of software) come in many shapes and forms and with different preferences, paid or free.

5.5.4 Publishing (Push or Pull)

In a DMS, publishing means releasing a version to employees as the preferred version for use. Other versions are still stored in the DMS, but they are usually no longer used by others. Some users find it very important to know what older versions might exist, for example documents that are sent to customers in different versions. They want to be able to open and edit an older version so they can make, for example, a new brochure for a target group based on the same principle. This cannot be done using a published version because this version is authorized and cannot be edited by others. So publishing a document gives a certain status to a document.

Within a WCM system, publishing means releasing a page for users. Until the moment of release nobody except the WCM application user can view the content in a browser environment. After publication, the content is available and visible for everybody.

Publishing involves two steps. First, the required content must be selected and then merged (assembly) to produce the desired result, the publication. This final step is called publishing.

In the workflow of the publishing process, publishing is the final stage of the process.

Chapter 6
Business Intelligence: The Structured Part

Peter van Til

This chapter on *business intelligence* (BI) investigates the possibilities that BI offers to meet information needs that arise in the pursuit of different business aims.

Section 6.1 uses a conceptual approach to BI. This means that it mainly elaborates on what BI is and what methods are used to solve information issues. The ideas of Kimball, Inmon, Linstedt, and Van der Lans cannot be missed in this regard. Sections 6.2–6.5 tie in with the EIM activity model. Section 6.2, on *availability*, highlights the presence of the required data in source systems. Section 6.3, on *accessibility*, explains how the data in source systems can be made available to end users while ensuring that all the (historic) data from the source system are preserved. Section 6.4, on *relevance*, discusses how to ensure that all information is made available in the proper form to the right person in the right place at the right time. Section 6.5, on *interpretability*, focuses on understanding data and the way these data can be visualized. Section 6.6, on *trends*, discusses a number of common applications and developments in the field, such as, for example, BI in the cloud, big data, and social BI.

6.1 What Is Business Intelligence?

BI is a field of study that makes structured information available to the right people in the right way at the right time.

In almost all cases, a *data warehouse* environment serves as a source for *BI*. The necessary data are transferred from a source system to a data warehouse, and from there the *data marts* are filled and then made available to the end user for compiling the required information (Fig. 6.1).

P. van Til (✉)
Business Intelligence, Incentro, Nootdorp, The Netherlands
e-mail: peter.vantil@incentro.com

P. Baan (ed.), *Enterprise Information Management: When Information Becomes Inspiration*, Management for Professionals 2, DOI 10.1007/978-1-4614-5236-2_6,

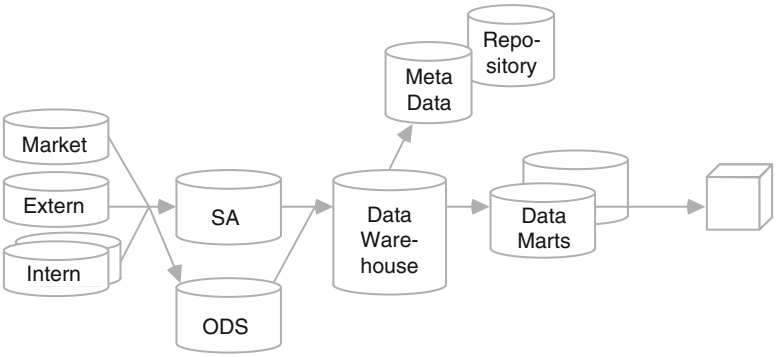


Fig. 6.1 Data warehouse and BI environment

The source systems, leftmost in the illustration, form the basis. The data that are needed as support in decision making are extracted from the source systems. Often this is done periodically (daily), but it can also be done in real time. The data are stored in a staging area or operational data store to be processed in the data warehouse. The data warehouse comes with a metadata repository. It is important to have such a repository because the data recorded in the warehouse will be used throughout the organization, so it should be clear what the meaning of a particular data element is. The data warehouse can be considered the conscience of an organization. All the facts that the organization has realized over the years can be found here. It is essential that this conscience be 100% unambiguous and 100% complete. This conscience can be used to reconstruct the image of time in order to recognize the challenges and opportunities that the organization is facing today. It allows an organization to better anticipate a situation. Not all data are required for this, and therefore so-called data marts are created from the data warehouse. Here, data from the data warehouse are copied and saved in a way that makes them interpretable for various BI tools. These tools give employees the opportunity to visualize data in a way that is most desirable for employees and for solving a problem. This could be a report, an online analytical processing (OLAP) application, a graph, through a dashboard or a scorecard, on a laptop or a mobile device. The result is an optimal decision-supporting environment.

Two aspects are very important in making information available: *accessibility* and *availability*. The terms *interpretability* and *relevance* are important in decision making. Within the field of BI, availability plays an important role in unlocking source systems. The required data must be generated somewhere inside or outside the organization. After the data have been unlocked, they are stored in a data warehouse. Here, they are integrated with data from other source systems, and, finally, they are made available to the organization through a so-called data mart.

Now the technical part with the data is done and it is up to the organization to benefit from it. It is important for the organization that all relevant data (for a particular problem) be present in the data mart and that the organization be able value them properly.

The field of BI is based on cooperation between technique, processes, and employees (Fig. 6.2).

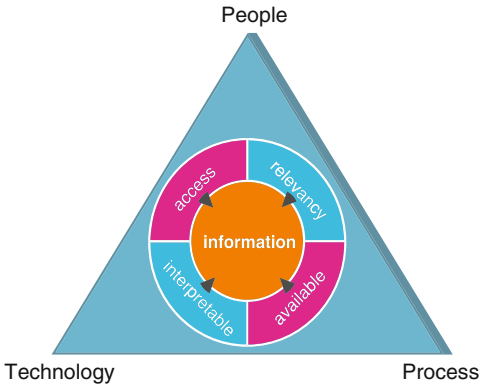


Fig. 6.2 EIM triangle

6.1.1 Vision

BI plays a major role within the EIM concept. The vision of enterprise information requires that the employee in question always have access to information when necessary. We cannot speak of BI solely based on the availability of information. The BI is determined by the business itself. The organization's ability to make decisions based on this information is crucial. Obviously, the available information and the ability to make decisions based on this information should tie in with each other. The maturity model for EIM offers good guidelines to show what an organization can do and what it needs to do so. Also, EIM scans offer insight into what is needed to raise the organization and its information management system to a desired, higher maturity level.

From a human perspective, BI means that a person is trained to make the right decisions. He or she has the full support of management, and the culture is based on transparency and the will to improve business process, in which the improvements are in line with the business goals. To the organization, BI has to do with culture rather than technique. Technique is merely a starting point, a necessity. Everyone works to improve the organization and to achieve business goals set by management. BI supplies the necessary information.

From a process perspective, BI means that business processes can be optimized by making decisions based on the numbers from the source systems or data warehouse architecture. Is there a way to make the process even more efficient? This does not mean reverting back to stopwatch techniques to minimize lead times, but we can show that at certain points in the chain work can be done more efficiently by using information in the process.

From a technological perspective, BI means that the data in the various sources are made available to the end user – directly or via an integrated environment (data warehouse) – in a way that is workable for this person. In recent decades, BI was mainly positioned from this perspective by a number of gurus. This has resulted in several concepts that can be used to provide the information needed to achieve the process perspective and human perspective.

From this vision, it becomes clear that the organization must value BI properly and positively influence results based on BI. These requirements form the basis of the maturity model that plays a role in the vision. Different expectations may be set depending on the maturity level of the organization. It would be unfair to employees to have expectations that exceed the maturity level. This leads to unnecessary tension and irritation.

Within the maturity model, there are four levels: fail, run, achieve, and outsmart.

Maturity can be determined for many different aspects, but in EIM and BI information management and information value are very important. Therefore, we will only discuss these two aspects here.

	Information management	Information value
Fail	Management only focuses on information that is needed for business processes that are its responsibility without paying attention to the coherence with the rest of the organization and its environment	Information is considered a burden, and employees are kept busy documenting information. Workarounds are used without concern for possible consequences for the underlying processes
Run	The organization has mapped out the information flows. There is insight into duplicate processes and their consequences, and it is understood how the information is used throughout the process	Information has been given a formal meaning within the organization. Standards are introduced, which leads to less discussion about certain terms. Reports become less ambiguous. The term *key performance indicator* (KPI) is carefully introduced
Achieve	Data quality is an important topic in this phase. After all, it can improve the effectiveness and efficiency of the processes. Data stewards are/will be appointed to manage the improvement of quality and to anchor this term in the processes	Information flows are transparent; their quality is ensured. Benchmarking (compare your own situation with that of your competitors/colleagues) increases efficiency. Information quality improves the quality of decision making
Outsmart	Information is a production tool, and the organization is constantly looking for opportunities to do more with this means of production. New technologies are used if necessary. The aim is to better meet the information needs of tomorrow. Anticipation is an important concept	Data are used to improve products or services to customers in such a way that a significant difference with the competition arises. In other words, there is a competitive advantage. 360° profiles of business objects and application of new technologies are important, and the organization is engaged in innovation to maintain its competitive advantage

This maturity model can be used to determine how the organization operates and what can be expected from it. More importantly, the organization is offered the opportunity to choose the level on which it wishes to operate, and it may decide what is needed to reach this level. So it gives direction to the information policy and the expectations of the environment.

6.1.2 Information Architectures (Concepts)

The data warehouse and BI environment shown in Fig. 6.1 are a type of architecture that is commonly used in the field. It is derived from an idea of Bill Inmon. The "father of data warehousing" came up with the Corporate Information Factory (CIF) concept in the 1990s, which he later developed into the DW2.0 platform. This idea became the starting point for Dan Linstedt's data vault.

The CIF was in fact a reaction to Ralph Kimball's data warehouse bus architecture. Kimball, also called the father of the data mart, created data warehouses on a departmental level. He did not need the kind of data warehouse shown in Fig. 6.1 but instead went straight from the staging area to the data mart. This led to a conglomeration of data marts, which Kimball considered a data warehouse. This idea is still commonly used in many organizations.

These ideas have been developed since the 1990s. At that time, the automation of basic processes was in full swing, and soon the capacity of hardware and middleware ran up against limits and extremes. Nowadays, the limits of the components used in networks are not what they used to be, and as a result, the opportunity to implement real-time capabilities has increased substantially. This is also the principle behind Rick van der Lans' ideas regarding the data delivery platform. Van der Lans says it should be of no importance to the user where the data come from as long as she has access to them whenever it is needed.

Let's list the various ideas, starting with Kimball and Inmon. After that, Lindstedt's ideas will be explained, and finally we will end with the concepts of van der Lans.

6.1.2.1 Data Warehouse Bus Architecture

Ralph Kimball used the data warehouse "bus" architecture. The term *bus* is borrowed from energy suppliers. Here, a bus is a shared component to which everything is connected and from which it gets energy. In a data warehouse, this bus is formed by so-called conformed dimensions. These are dimensions that are part of several processes and are commonly used. Kimball uses a data warehouse bus matrix to determine which data should be considered as having been formed by conformed dimensions and which should not. Figure 6.3 is an example of such a matrix.

The data warehouse bus matrix is the general data architecture for the BI environment. Every implementation of a business process builds on the results of a previous implementation.

There is a short architecture phase to make sure that the development of data marts does not lead to *stove pipe* solutions. In this phase, for example, the bus matrix is created, the conformed dimensions are designed, standard definitions are established for measurements, and the infrastructure is set up. This phase shows, among other things, the points at which the data marts touch/overlap. Based on this information, data marts can be developed individually, and the already developed

Business process/event	Date	Policy holder	Coverage	Agent	Policy	Claim
Close policy	✓	✓	✓	✓	✓	
Prolongation	✓	✓	✓		✓	
Calculate commission	✓	✓	✓	✓	✓	
Submit claim	✓	✓	✓	✓	✓	✓

Fig. 6.3 Data warehouse bus matrix

parts of the data warehouse bus can be put to use. These coordinated data marts are called *supermarts* (Kimball 1998).

A data mart consists of one or more fact tables. Every record in a fact table has a key that consists of the combination of keys from the related dimension records. Each fact table contains zero or more measurements that apply to any combination of dimensions. The model for one fact table is also called a star model. The technique that is used for modeling the star schema is called dimensional modeling (Kimball 1997).

A conformed dimension is a dimension that has the same meaning in relation to any sort of fact table. In general, a conformed dimension is identical for every data mart. A major responsibility of the central data warehouse team is to establish, publish, maintain, and enforce conformed dimensions. Typical examples of conformed dimensions are customer, product, supplier, promotion, date, and time.

Without a strict connection with conformed dimensions, a data warehouse cannot function as an integrated unit. When a dimension such as customer or product is not shown as a conformed dimension in the architecture, the related fact tables (for the individual product tables or customer tables) cannot be linked to each other. If such linking is done anyway without the use of a conformed dimension, then the chances of producing incorrect information are high. Most conformed dimensions define everything on the lowest detail level. A customer dimension will include every individual customer, and for each customer it will record a large number of descriptive attributes. These attributes can come from several source systems. Therefore, several sources are integrated in the staging area so that they refer to one and the same customer. The most important activities in the staging area are sorting, cleansing, searching and checking relations, and combining.

The preparation of standard fact definitions is done during the architecture phase, prior to the development of the data warehouse. Standard definitions are important when certain terms are used at different places within an organization. Examples of standard facts are turnover, profit, standard prices, and standard costs. If no consensus can be reached concerning the meaning of some measurements, all measurements as such will remain in the data warehouse.

The architecture starts by filling up so-called *first-level* data marts. These are data marts that are linked to one single source system (Fig. 6.4). This means that the facts

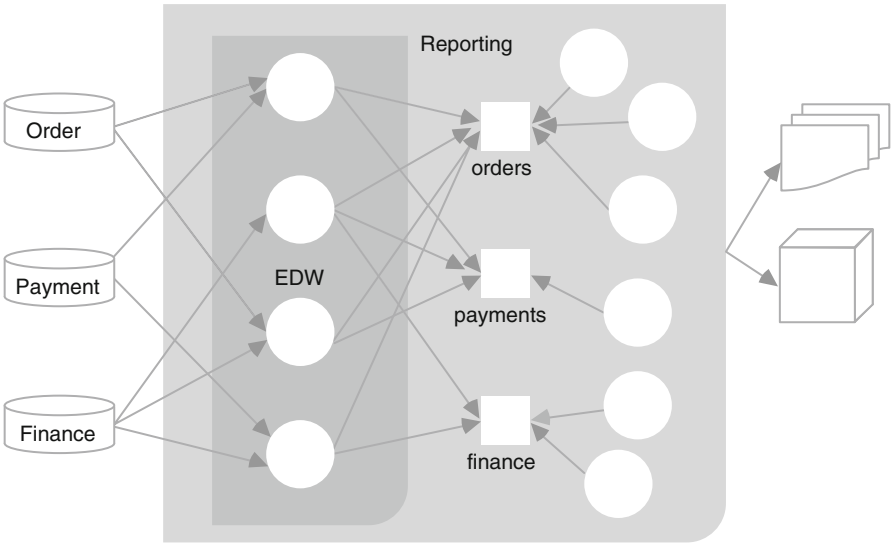


Fig. 6.4 Data warehouse bus architecture

in the fact table, as well as the ordinary dimension, come from one source system only. Starting in this way avoids the danger that the implementation of the first increment will be too ambitious. The user experiences the new information components as an improvement and is therefore pleased. While the user starts exploring the new features, the data warehouse team can continue developing more difficult information requests. *First-level* data marts can be combined into *second-level* data marts like an output data mart. This is a combination of costs and revenues and gives insight into the customers, suppliers, and products that contribute most to the efficiency of the organization.

6.1.2.2 Corporate Information Factory

Bill Inmon uses the *corporate information factory* (CIF) as the reference architecture (Fig. 6.1). CIF is the central point of information that enables an organization to respond optimally to changes within the organization and beyond.

Business processes are supported by so-called *stove pipe* systems. These systems aim to increase the efficiency of one specific process. The organization's environment as well as the essentials within the organization and the effect of current decisions on the near future are visualized by BI applications.

Responding to changing customer demands is covered under a field of study that combines *BI* and *business management*. Based on findings from the *BI* field, *business management* comes up with applications within the business. The *CIF* provides a robust foundation to support *BI* and *business management.*

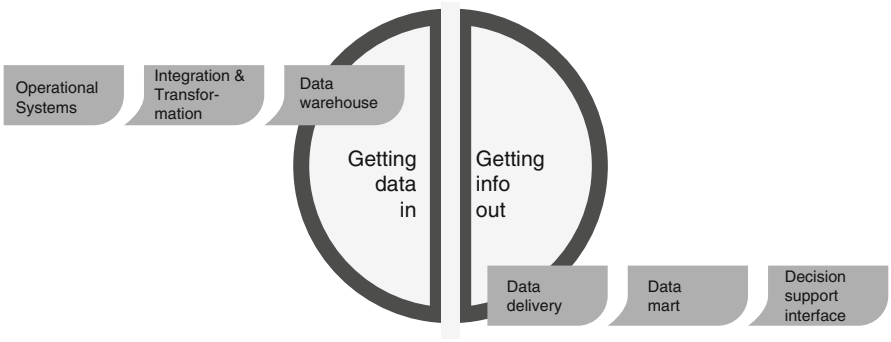


Fig. 6.5 Getting data in versus getting information out

The *corporate information factory* can be divided into two fundamental processes (Fig. 6.5):

1. Producers who *obtain data* by placing them in context for use by the entire organization.
2. Consumers who *obtain information* by providing support to *BI* and *business management* issues.

Different elements of CIF support the daily activities of different types of users. Inmon distinguishes four user types, also called *personas*. These are *tourists, farmers, explorers,* and *miners.* A tourist is someone who wants a quick impression of the environment and determines the most important elements. A *farmer* has a structured way of working. He makes sure the environment is clear – no weeds (*bad data*). An *explorer* wants to know the ins and outs of an extraordinary situation and solve problems in his working area through such knowledge. In doing so, he exposes aspects that are not very obvious. A *miner* takes it a bit further and explains the causes of the extraordinary situation. He or she formulates assumptions and hypotheses and tries to have them confirmed by the available information.

The data warehouse of a CIF is designed to support strategic and tactical decision making. It is not suitable for supplying real-time access to detailed transaction data. Therefore, the operational data store (ODS) is added to the architecture. This represents a remarkable difference in architecture compared to Dan Linstedt's data vault, which will be described later in this chapter.

6.1.2.3 DW2.0

DW2.0 is a new model for data warehousing that can meet current user demands and needs. The model takes the various types of data into account, as well as their structure and how they are related. As such, DW2.0 constitutes a solid foundation for filling in the current information supply.

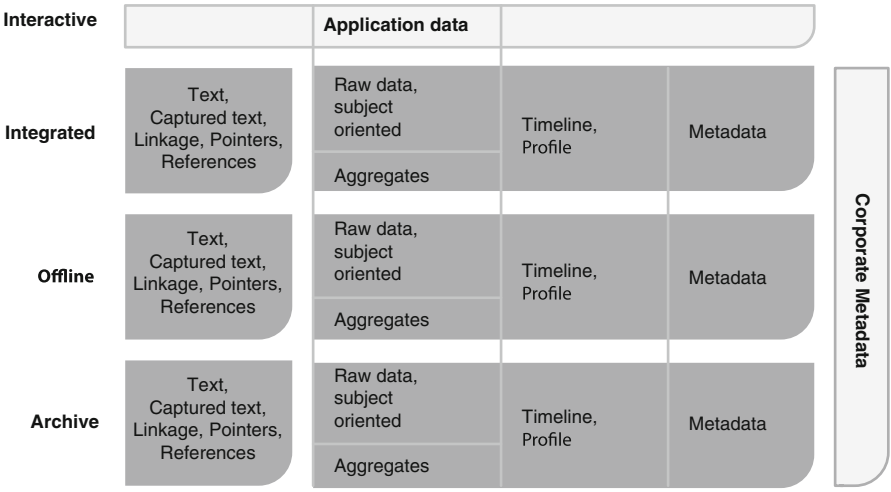


Fig. 6.6 DW2.0 reference architecture

The most striking difference between CIF, a first-generation architecture, and DW2.0 is in the recognition of the *data lifecycle*. Data start a new lifecycle the moment they are included in a *data warehouse*. This lifecycle is supported by the four sections in DW2.0. The first section is the interactive section. Data are contained very quickly following creation. When the data are no longer the subject of change, they are transferred to the integrated sector. Here the data are integrated with other data. The data remain there until their use decreases. Reduced use occurs when the data's added value for the organization starts to diminish. This happens as the data get older. In general, use starts reducing after 3 or 4 years. From the integrated section, data are transferred to an *offline* section or archive. The *offline* section is an extension of the integrated section. This section is only used for extremely large amounts of data whose use regularly fluctuates. If there is no fluctuation and the data are rarely used, then the data are probably no longer needed in everyday processes or analysis (Fig. 6.6).

DW2.0 is an architecture for structured and unstructured data. Unstructured data must go through an integration process before being included in the DW2.0 environment. The integration process prepares the data to enable analysis of unstructured data.

What are the advantages of DW2.0 for businesses?

- The costs of *data warehousing* are limited.
- The environment is held together by metadata. As a result, data are less likely to be lost or to lose meaning.
- Data are quickly accessible. In the first generation of *data warehouses,* data were buried under loads of – at that time – unneeded information. This resulted in performance problems.
- The necessity of archiving is acknowledged.

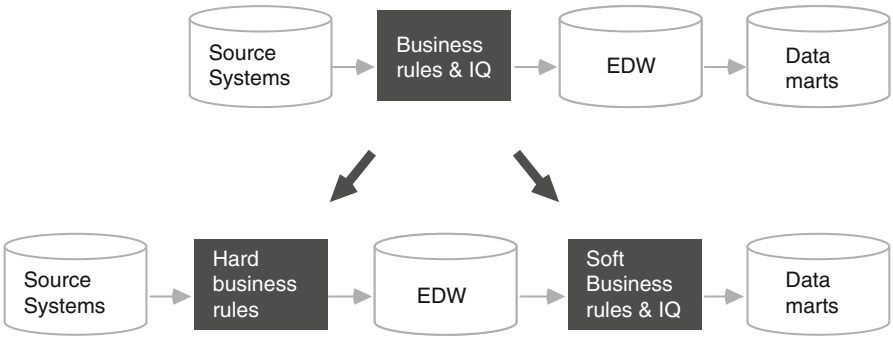


Fig. 6.7 Compliant versus noncompliant data warehouse architecture

6.1.2.4 Data Vault

Dan Lindstedt's approach to *data warehousing* is called a *data vault*. The concept is based on the principle that responsibilities should be placed where they belong. This makes the data warehouse *auditable*. *Data governance* is possible because one can determine what was delivered by various parties and when this was done. The environment is *compliant.*

The upper flow in Fig. 6.7 represents the traditional way of working with a data warehouse. This is the *strategic-only* way of working. This means that this way of filling up an enterprise data warehouse (EDW) is only suitable for batch processes that provide the support for strategic decision making. The transformations result in such a slowdown that real-time support cannot be achieved. The lower flow represents the new way of working in which *active/real-time* data warehousing is possible, and this provides a *single point of record/fact*.

Organizations are constantly trying to support the changing market as well as they can. The rules that are applied also change because of these changes. Changing rules lead to different results in EDW. Today's truth is not, by definition, tomorrow's truth. Consequently, it is hard to discern which data formed the basis for the data recorded in the data warehouse. In other words, the data warehouse is not *auditable*. At a time when more and more near-real-time functionality is expected of a data warehouse, it is hardly possible to perform major transformation steps when data are provided for the purpose of being loaded to a data warehouse environment. It is important for both reasons (auditability and near-real-time functionality) that the data stored in the source system be copied in the data warehouse without any transformations. The *single point of truth* is hereby replaced by a *single point of facts.* Or, as Barack Obama states in his book *Thoughts on Reclaiming the American Dream*, "Everyone is entitled to his own opinion, but not to his own facts." Because all data in a data warehouse are exactly the same as the data in the source system, it is now impossible to establish an audit trail in a source system. The facts are no longer subject to discussion.

The separation between hard and soft business rules has to do with the functioning of the data warehouse. Everything concerning integrity is covered by the hard business

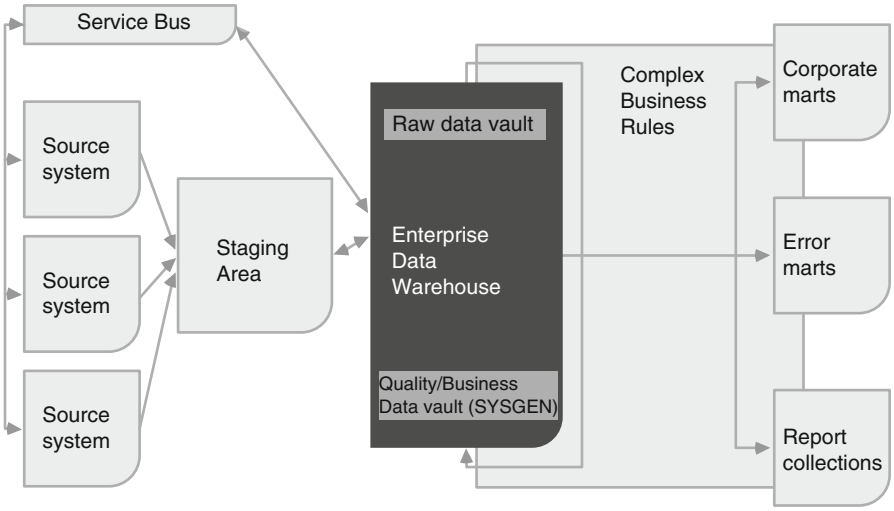


Fig. 6.8 Data vault architecture

rules. All other aspects fall under the soft business rules. A hard business rule is a business rule that ensures that data can be loaded at any time. One might think of filling in empty fields so that joins between database tables always go well. Changing business rules has another advantage: gap analysis can be performed to reveal the differences between how a business operates and how it thinks it operates (Fig. 6.8).

When we place business rules before the storage of data in an EDW, the EDW will show how the business *thinks* it operates. After all, all transformations have already taken place. When business rules are placed after storage in an EDW, the EDW will show how the business *actually* operates. This can lead to interesting insights.

6.1.2.5 Data Delivery Platform

The *data delivery platform* (DDP) (van der Lans 2009) is the answer to the growing interest in operational BI. *Operational BI*, combined with strategic and tactical BI, requires a different architecture, less redundant storage, a combination of structured and unstructured data, specifications that may be used by several tools, and a separation between data storage and data presentation. The DDP focuses not on data storage, but on the integration of various sources in the organization.

The DDP separates data storage from data usage (Fig. 6.9). All data usage goes through the DDP, which maintains connections between various data providers. It is clear to the user where the information comes from. New technologies are making redundant data storage increasingly unnecessary.

The DDP simplifies the whole environment. Parts of the existing architecture can be remediated because more and more information is extracted directly from the data source. Fewer components means simplification and less maintenance efforts.

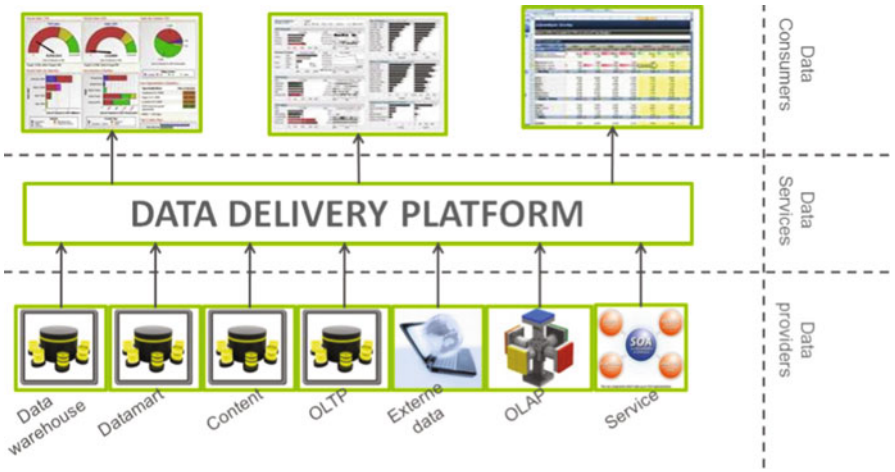


Fig. 6.9 Data delivery platform

The DDP does not imply elimination of the need for extraction, transformation and load (ETL), *enterprise information integration* (EII), (virtual) data warehouses, and service oriented architecture (SOA). ETL as a tool for integrating data will remain necessary, as will ETL as a tool for moving data, but the amount of data that plays a role in this will decrease.

The DDP should manage a metadata repository that can be approached by all (ETL) tools.

EII is a technique by which several heterogeneous databases can be considered one logical database. EII provides data integration *on demand*. DDP is EII++.

Virtual data warehouses are created by covering several source databases with a layer of software so that it looks like there is one central storage. Objections have been raised to having virtual data warehouses work as a real data warehouse (Inmon, 2009). In particular, the required real-time data integration and the extent of the required resources cause problems. DDP does not prescribe the use of a virtual data warehouse.

The SOA concept is based on the concept of the service itself. A service is a secured amount of functionality that makes its results available for users. It is therefore easy to integrate DDP in SOA.

New technological capabilities make this platform more and more realistic. A federation server adds content to the data services layer. New storage techniques provide an acceleration of access to data, and in-memory processing makes storage devices partially or completely redundant.

The data provider layer will continue to contain the components mentioned in Fig. 6.9. However, the components will only contain the data they need. Increasingly, data will be stored only once. Data warehouse techniques will only be used when

the source system does not have the required information. This will greatly reduce the time required to provide information to the end user. New capabilities lead to new requirements and to new architectures. The application of new technologies will mean that architectures are adjusted to help the organization reach its potential.

The field of data warehousing and BI is a combination of people, processes, and technology. It is not surprising that especially in data warehousing technology prevails. In BI it is more about a person using technology to her advantage.

Because a data vault architecture adds most value to the information architecture of an organization, it is used as a reference architecture for a further description of the field (Fig. 6.8).

6.2 Availability

Availability has to do with data that are generated inside or outside the organization and that can be used within the information management of the entire organization.

To determine which data from which source systems are needed to meet the information needs of (a part of) the organization, one should first know what elements the information need is composed of as well as the business rules that are applied to obtain the desired information.

When the required data elements are known, they can be transferred to the first data store within the data warehouse environment, the staging area. Specific techniques will be used to do this, depending on the type of storage in the source systems. The vast majority of data elements are stored within an organization in relational databases, and these can be transferred to the staging area through an SQL process. Based on metadata, SQL processes may be generated, resulting in a flexible solution and a decrease in implementation time of the new source elements.

Special techniques like unlocking source data are required for, for example, big data environments and social media. This will be further discussed at the end of this chapter.

Until now it was assumed that all elements required for meeting information needs were present. This need not always be the case. If certain data elements are not available, then the information need cannot be met. One must first determine how this missing information can be added to the processes within the organization or how these new processes can be implemented in the existing environment(s). A common example is information about customer satisfaction. Every organization would like to know how they are perceived by (potential) customers. Often, there are no measurement points inserted in the business process to capture this information. Therefore, new processes should be included in the workflow.

An example is the company Carglass. Whenever a windscreen is replaced or the existing windscreen repaired, an e-mail is sent to the customer to ask about his experiences during the repair or replacement of the windscreen. When this is done, the required data are generated and become available to the organization.

6.3 Accessibility

The process of making data accessible can be split into three aspects: loading the data into the various data stores, modeling the data for the purpose of this storage, and verifying the data quality. The loading of data is done using extraction, load and transformation (ELT) processes. An ELT process is a variant of the well-known ETL process. Modeling is done based on a data vault of a star model, depending on the store in which the data will be kept.

6.3.1 Data Vault

An EDW uses a data vault and in the data marts (corporate marts) a star model is used.

A data vault is designed for capturing history within an EDW. A data vault also provides an extensible architecture that makes the management of the environment very simple. This has to do with the way the data are spread out over tables. In a data vault, there are basically three types of tables: HUB, LINK, and SAT. A HUB table contains the identifying elements of a business object. Business objects are terms used within an organization like, for example, customer, product, product group, supplier, order, and invoice. A LINK table contains the links that arise between tables. The link between a customer and a product is laid down in a LINK table. An order line is the link between a customer and a product. All descriptive elements are recorded in a SAT table (SAT is short for SATELLITE). For example, a customer's address would be included in a SAT table. A SAT table is linked to a HUB table or a LINK table. This results in a model like that shown in Fig. 6.10. For each HUB or LINK, multiple SAT tables can be designed (Fig. 6.10).

All history is laid down in SAT tables. For each change within the elements, a new record is added to the SAT table, and the validity of the data is registered based on start and end dates. Within a star model, this principle is also known as a type 2 slowly changing dimension.

New data can be linked to the existing model based on LINK tables, which leads to extensibility. The existing model is not modified in the process so that the already implemented information need is not affected. Only the part of the data warehouse with new entities is tested, which is considerably quicker. The test is further limited by the standardization of tables.

6.3.2 Extract, Load, and Transform

A variant of the well-known ETL is used in data vault architecture: ELT. This change in location where certain things are done has major consequences for the management

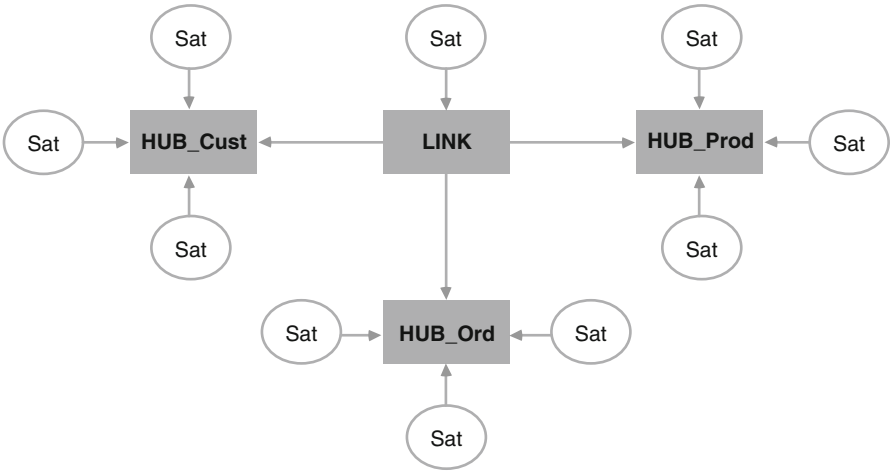


Fig. 6.10 Data vault data model

and flexibility of the solution. The ELT process makes sure that data remain unchanged and that they are recorded in the data vault in a standard manner. This part of the data vault is called a raw data vault. The data will be recorded in the data vault without paying attention to the content of the data elements. This is a deliberate choice. This provides the possibility of tracing all data in the data warehouse back to the source systems. It is, so to speak, a one-on-one copy. This is important for the auditability and traceability of the total environment. After all, the data warehouse team (getting data in) is responsible for capturing and retaining data and not for the content. This is the responsibility of the data or process owner that generates the data. Assessing the content/value of the data is further discussed during the determination of the relevance of information.

6.3.3 Data Vault

In a HUB table, only the identifier of a business object is defined. This is the primary key and cannot be changed. The functionality is the same for each HUB table and is basically very simple. All records from a specific supply are compared to the existing content of the HUB table. If a characteristic already occurs in the HUB table, further processing will not be necessary. If it is new, it will be added to the table (Fig. 6.11).

Certain technical elements can be added to the HUB table such as record_ID, loading date, and the origin of the key of the business object. If a standard process can be used to update the characteristics of a business object, then that means the process has already been tested and the testing can be limited.

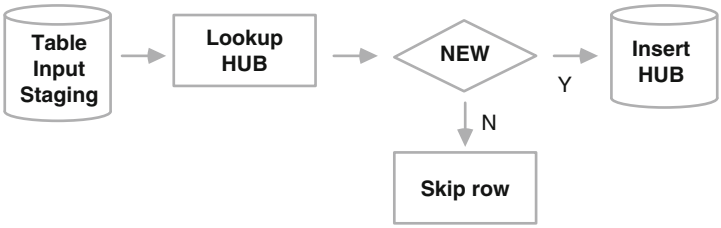


Fig. 6.11 Loading process HUB table

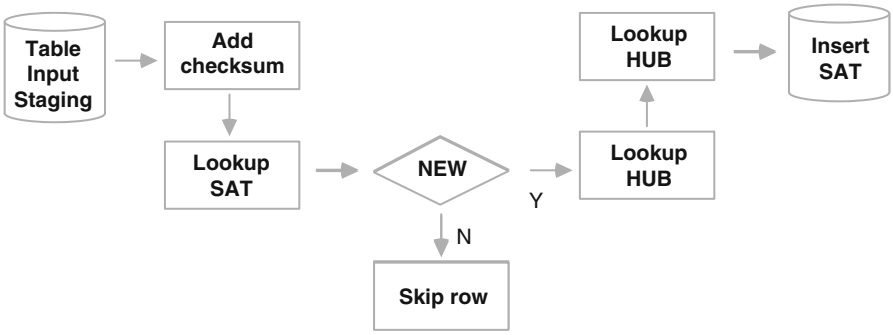


Fig. 6.12 Loading process a LINK table

The same applies to a LINK table. After all, a LINK table serves as a link between two HUB tables. Therefore, a LINK table only includes references to the record_ IDs of the HUB records that are linked together. Since history is recorded in a data vault, a link cannot be changed or deleted once it has been established. A record_ID, loading date, and origin are added to the IDs of a LINK record. If a link is already established, further processing will not be necessary, and if one is not established, then a record should be added. So a LINK record basically has the same functionality as a HUB record.

To limit the number of lookups, a so-called checksum is used. This is a unique code for the shared values of certain elements. In this case, the values of the keys from the HUB tables need to be linked. If this checksum already occurs in a LINK table, then further processing will not be necessary. If it does not occur yet, then there is a new LINK and the record_IDs from the HUB records will be collected and the results will be written in the LINK table. In Fig. 6.12, a LINK is formed by two HUB tables, but basically the number of HUB tables that can be linked is unlimited.

Although a SAT table is a bit harder, it is also based on the same principle: a type 2 slowly changing dimension. Once again, only new situations must be registered. Besides the technical elements mentioned previously, a validity period must be added to each situation. Based on the checksum (determined based on all collected data elements), it is determined whether or not it has already occurred. If the checksum has occurred and the validity period has not passed yet (date valid until 9999-

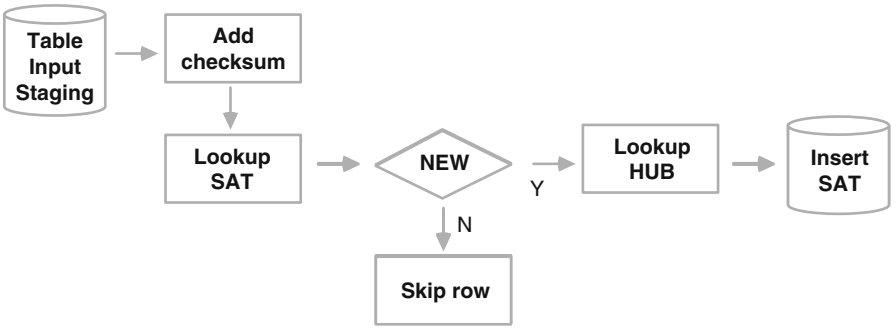


Fig. 6.13 Loading process an SAT table

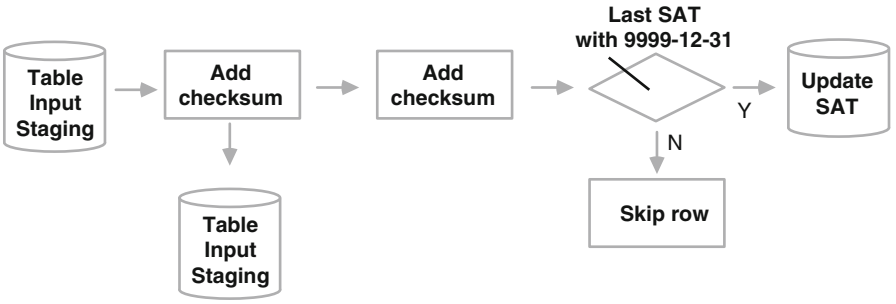


Fig. 6.14 Update process on SAT records

12-31), then further processing will not be necessary. If the checksum has not occurred, then a record will be added (Fig. 6.13).

A SAT record belongs to a HUB or a LINK. This means that the SAT record can refer to only one record. This is done using a foreign key (referential key).

After a SAT record is added, there may be several records for one foreign key with a date valid through 9999-12-31. This means that the validity periods overlap. This should not happen and must be corrected. This is done using an update process (Fig. 6.14).

The input is formed by all HUB_IDs with at least two SAT records (from the same SAT table) for which the end date of the validity period is 9999-12-31. All these SAT records are read and sorted based on the start date of the validity period. The end date of the records is changed with the start date of the validity period of the next record that refers to the same HUB record or LINK record. When there is no subsequent record, the end date remains unchanged on 9999-12-31. This creates a timeline with consecutive validity periods.

The four processes that are needed for maintaining a data vault are now known and can be copied an unlimited number of times for every HUB, LINK, and SAT.

This does not mean that the data are available to the end user. There are two more hurdles to overcome. The first is data quality, and the second is the corporate mart.

6.3.4 Data Quality

A data vault that includes unmodified data is a raw data vault. This is unlike the quality data vault or business data vault. In the business data vault, data are added that meet the data quality requirements set by the organization. The data from the raw data vault are checked using quality rules or business rules. These terms are synonymous with each other. Data that do not get through this screening will be included not in the corporate mart, but in the so-called error mart. This error mart is the source for a bug report to the data owner, who in turn can correct the data in the source systems. The corrected data are included in the raw data vault during the next loading process.

Several tools are available to implement business rules. It is important that there be a separation between defining the rules (which must/could be done by the business) and implementing them (which must/could be done by IT).

The results of the rules are laid down in a business data vault. These are separate tables (SATs) that are integrated in the raw data vault. To distinguish them from the other tables, these tables get special names (for example, QSAT or BSAT). Unfortunately, the processes that generate these tables are not as standard as the processes for raw-data-vault tables. After all, business rules vary from one business to another. So this is always custom-made within the total solution. It can be filled in whenever a lack of quality is detected or when the need arises to implement a certain business rule. The raw data are always there, so the data in the business data vault can always be regenerated.

6.3.5 Corporate Mart

The separation between identifying data (HUB), relational data (LINK), and descriptive data (SAT) creates a data model that is not suitable for processing with a BI tool. A BI tool is an application that enables end users to gain insight into data and make decisions for improving the organization based on this insight.

Such a tool needs a so-called star model – also called a dimensional model – to perform optimally. The data must therefore be restructured into a data mart. Because basic data are stored in a data vault, a data mart can be (re)generated at all times.

Restructuring data from the data vault to a star model in the corporate mart (data mart) is a technical process in which it is important that HUBs and their SATs can be converted into dimensions and that the transactional LINKs and their SATs can be converted into fact tables. As a result, the data model looks a lot simpler (Fig. 6.15).

All dimensions have a 1:N relationship with the fact table. The identification of records from the fact table is a combination of all keys referring to the dimension tables. A dimension can consist of a combination of data from several HUB and

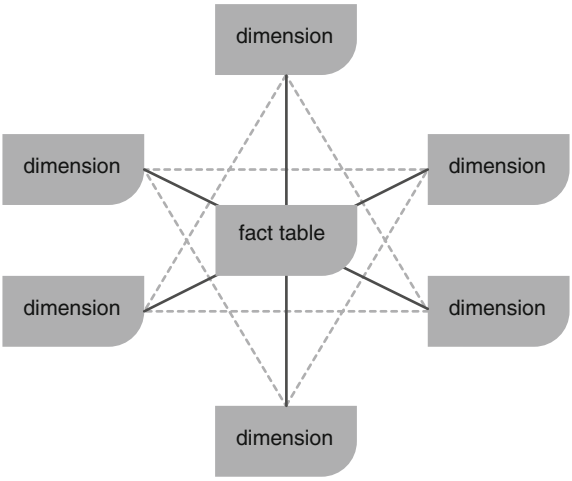


Fig. 6.15 Star model or dimensional model

Fig. 6.16 CXAIR 1

SAT tables. For example, a HUB's and SAT's product and product group can be combined into a dimension "product." This combination creates hierarchies within a dimension. These hierarchies are necessary to be able to drill (changing the level of detail) within the data of the fact table.

Within a corporate mart, data are included that are needed to meet the information needs of a specific user group inside or outside the organization. A corporate mart may consist of multiple star models. This results in shared dimensions, that is, dimensions that are the same for multiple fact tables. These are the so-called conformed dimensions. Based on these conformed dimensions, facts from several fact tables can be combined so that new information can be created.

6.4 Relevance

Relevance has to do with information being available at the right moment, in the right place, for the right person, in the right form. To achieve this, the ideas of van der Lans come in handy. The challenge lies in choosing the sources that are most relevant to the desired information. Sometimes this is a data warehouse, sometimes a direct link to a source system is established, sometimes a Web service is used, and sometimes a combination of these possibilities is used to obtain the desired information.

It is important to choose the right (combination of) sources. There are no rules that dictate a certain choice, but the key to finding the best practice is determining the shortest route. If the required data appear to come from only one source, then this source can be used directly to meet information needs. We now ignore the fact that because of performance problems an alternative needs to be found. If a combination of sources is needed but there is no need for historical data (apart from the history in the source systems), then an operational data store (ODS) may be used. If both a combination and history are required, then a data warehouse may provide a solution.

When the source has been chosen, the end user should be able to quickly make the correct selection to meet the information need. This selection can be made automatically in advance or manually at the moment the information is compiled. An automatic selection can be made when it is clear that only information from the last 3 months is needed or when the user is only allowed to see data of her own department. Manual selections are often defined as parameters for reports. Before a report is compiled, the user must indicate what selection criteria should be applied. Select a country using a drop-down box and then a choice of province/state and a radio button for a detailed report or just a summary (totals). This quickly provides the end user with relevant information.

The right place is always via the Web. There should be no restriction for the end user regarding the place and time he or she wishes to have the information available. The office PC, the laptop while working on site, the tablet in the living room, or the smart phone – all are devices on which the end user should be able to obtain the desired information.

The right combination of sources, selections, and (mobile) devices provides the necessary infrastructure to increase the relevance of the information. Relevance also depends on how and where the information should be used in the organization – strategically, tactically, or operationally. Strategic information must generally support business goals, whereas tactical information should be tailored to optimize business processes. Operational information provides insights into performed activities and statuses. Examples could be the number of products produced, the number of hours spent, or the number of complaints that have been addressed. The place where the information is used determines the relevance of the data.

6.5 Interpretability

The final aspect of people–information–technique interaction that will be discussed is interpretability. This concept makes use of several BI tools, tools ranging from simple reporting tools to more advanced OLAP and data visualization tools. Which tools are chosen will depend on the way the information is made available. This choice requires cooperation between IT and business. IT will do the technical management of the tool, and the business will have to work with it.

The business has a number of roles that must make decisions, and these decisions are made based on different types of information:

1. Standard information that is made available periodically;tThink of a daily occupancy survey or the weekly turnover reports lying on a manager's desk in the morning;
2. Ad hoc information that is requested in response to situations that arise in the organization; one sees "interesting" things and focuses on them;
3. Detailed information that can shed light on particular points of a situation so that explanations of the situation may be given;
4. Information about patterns in data that can be used to optimize a business process or predict the effects of a situation.

The use of information as mentioned previously is based on the personas described by Inmon. These are, in order, the farmer, the tourist, the explorer, and the miner.

Obviously, the different ways in which information is needed require different tools to support these in the best way possible. A number of tool types are discussed below, making it clear when a particular type may be chosen.

For standard information it is best to use a reporting tool. A reporting tool ensures that information will always be presented in the same manner. The end user cannot change the information displayed; everything is static. Reports can be personalized so that they only show information that users are entitled to. Think of a team leader who can only view detailed information on his teams. He can only view aggregated information about other teams. Examples of these reporting tools are business objects, WebFocus, Microsoft Reporting Server, Pentaho report designer, and Microstrategy.

The information from a tourist is hard to define, but it will often contain comparisons, for example, information about the current period compared to the previous period or the same period last year. This information leads to comments like "I did not expect this. What is going on?" Tourists are people who have extensive experience within an organization. They can analyze and compare the numbers in a report to reality or to expectations based on their experience.

An explorer wants to further examine a situation discovered by a tourist and to find an explanation for things. Traditional reporting tools lack functionality to do this. A suitable tool type would be OLAP tools. Such tools show the starting point of a number of measurements,for example, the turnover per product group per year.

An end user can change the starting situation by choosing a different detail level or making another intersection. From this starting situation the user can change the year level by showing the turnover per quarter. This is often done with a (double) click on the year. From the quarter level one could go to a month, week, or day level. This can also be done for a product group and refined throughout the product level. The end user can also add information so that other information is generated. By adding store information one can obtain insight into the turnover of products in a given region. Depending on the data in the OLAP model, other insights are possible. Changing the detail level is called "drilling down." Adding information about which a selection is made is called "slicing and dicing." Examples of these tools are Cognos, Microsoft Analysis Server, and Oracle Discoverer.

A miner tries to find connections between the explanations found by the explorer. Which situations led to the discovery and what is the connection between these situations and how do we handle them in the future? Finding these connections eventually leads to an improvement in management. Examples are fraud detection and credit card issuance. The organization of shops is another example; this is done based on links in sales transactions. Algorithms that are needed to find connections are part of statistics, and high-level business knowledge is required to make optimal use of these connections. Examples of data mining tools are Weka, Rapidminer, and Kxen.

OLAP tools offer end users a flexible way of dealing with information. However, this flexibility is limited to the data stored in the OLAP tools. It is often also impossible to use search terms on several groups. For example, problems like "Provide an overview of all soft drinks with the word 'juice' in their names" or "Who had a subscription to the Dutch children's magazine *Donald Duck* both last year and this year?" are hard to solve with an OLAP tool.

For these kinds of requests, a BI search tool is useful. The starting point for a search tool is the data, while in many BI tools the form in which the data appear is the starting point. A search tool basically works the same way an Internet search engine (Google) does. Examples of these tools are Attivio and CXAIR.

Suppose there is a magazine publisher that notices subscriptions to its magazines have fallen off recently. At an annual meeting, the CEO and the c-level managers of the company decide that the sales of subscriptions to weekly magazines should be increased by 10%. The CEO receives a weekly survey in which sales of subscriptions to various magazines are compared with the status of the previous period. This shows that sales will remain roughly the same over the periods. The head of marketing is given the task of exploring how to go about promoting sales. He thinks it is odd that sales have not risen since so much effort is put into marketing and selling new subscriptions.

He divides the available subscription data into two sets. One set provides insight into this year's subscribers and the other into next year's subscribers. Based on these data sets, it is concluded that sales will keep pace (Fig. 6.16).

Then the marketing director asks himself: What subscribers have renewed their subscriptions? He gets an insight into this by combining the two separate data sets (Fig. 6.17). When overlaying the two sets of data, the tool suddenly yields much more information. The subset on the right shows the subscribers that had a subscription last year. The left set shows this year's subscribers. In the middle are the subscribers that have already renewed their subscription.

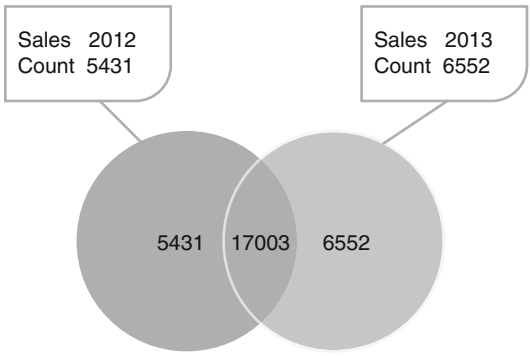


Fig. 6.17 CXAIR 2

So the right side shows the results of the campaigns to sell new subscriptions. On the left side, we see which subscribers would not like to renew their subscriptions.

With this information on his screen, the marketing director immediately calls a meeting with the team leaders of the weekly magazines and devises a strategy to persuade "dropouts" to renew their subscriptions after all.

From data to information to knowledge to action, that is BI.

For the CEO, it is important that a list of files can be provided in which information with regard to the critical success factors (CSFs) of the company is presented. This technique is called *balanced scorecards*. The user has all information about the state of the CSFs and key performance indicators (KPI) on one screen. The screen consists of four quadrants (employee, process, innovation, and financial). For each KPI on the screen the budget, the realization, the valuation of the realization (traffic light), and the trend (arrow) are known. A well-known product in this area is Bizzscore.

The scorecard is not such a useful medium for analyzing the state of CSFs/KPIs. Therefore, analysts or division managers have access to a dashboard tool. These tools show the situation on a particular subject on one screen. Here, several numerical and graphical reports are combined into one information screen. Examples of dashboard tools are Qlikview and Tableau.

In this way, it is possible to find optimal support for each information problem. This means there is no standard solution within an organization. IT will have to get used to the fact that several tools are needed to support information workers.

6.6 Trends

6.6.1 Cloud

"BI in the cloud" is increasingly mentioned as a possible alternative platform for BI environments. Just as often, this is followed by arguments against it: security and performance. The underlying argument probably is "unknown, unloved."

Nevertheless, we must consider the possibilities and impossibilities of this application, especially now that organizations are increasingly using cloud solutions to support their daily processes, for example, their sales forces or time registration. But also applications related to document management are available such as Google Docs and Microsoft Office365. More and more applications will force BI to shift toward the cloud as well. Subscribing to business applications has financial advantages compared to the current licensing structure governing software purchases. Compared to ownership, it is a much more flexible architecture. The problem of cloud BI often lies in the data. Many organizations are reluctant to store their data in external location that are not predefined. And they should be. But if the security is guaranteed at least as well as it is for storage within the company, then this argument will cease to hold. It is the elusiveness and the idea of not being in control of data that leads to a reluctance to store data in the cloud. But it is simply a matter of time before companies start to do that.

6.6.2 Big Data

In the last few years, the field big data has been booming because of the enormous amounts of data that need to be stored from various applications. Think of web analytics, the application of smart meters in the energy industry, RFID technology in logistics, or research findings in science and the pharmaceutical industry.

Let us first state that the technique of processing big data is something every organization should master. In coming years, this problem will arise in more and more organizations. There is no escaping it.

After processing has taken place, it is important to remove data from the daily environment as soon as possible. Archiving, using techniques like hadoop, is an option, but the data should be separated from the data warehouse and only the results of processing should be stored.

The phenomenon of Web analytics has received special attention because of new legislation on privacy. It is no longer legally permissible to use IP addresses to make special offers to (potential) customers, and many Web analytics principles are based on the use of the IP address. Workarounds will probably be found, but the issue remains a tricky one.

In the pharmaceutical industry a lot of research is done on medicine. Enormous amounts of data are needed to determine the effect of the medicine. Processing these research data is an extensive task, and processing them into the desired information is a very complex undertaking. Often, this requires some data mining because patterns in the data form the basis for the functioning of the medicine.

6.6.3 Social BI

In recent years, social media have come to play a major role in everyday life. Few people can do without Facebook or Twitter. People are online 24/7 and communicate as never before. Much communication is about companies and their products, which yields valuable information for the marketing department. Many organizations have one or more full-time employees who check messages and, if possible, respond to them. In this way, contact with customers is very direct and the impact of negative publicity is limited and may be converted into positive publicity (if the reaction of the company is accepted).

So far, this is not a BI application. BI comes in place when all the messages are used for analyzing, discover trends and improve the business processes. Analyzing all these messages is a daunting task for marketers. Questions like "How often is the company's name/product mentioned in this medium?", "Was it a positive or negative remark?", "What is the position of the person who sent the message?", "How quickly did the organization respond to the message?", "How well does the competition do within this medium?" etc. Analyzing social media data raises many questions but just as many answers. That is because of special applications such as uberVU.

6.6.4 Mobile BI

One aspect of social media that is emerging with a strong presence is mobile BI. Everyone should have access to BI applications via their smartphones, which allow users to see what's going on in their world when and where they want to. Many BI suppliers have responded to this need. As a result, the accessibility of data will increase significantly and many more employees will want to have access to data. Making decisions is no longer reserved for managers but is a task for all employees of an organization. Good ideas do not depend on place and time.

Chapter 7
Enterprise Search and Retrieval (ESR): The Binding Factor

Anja van der Lans

This chapter on enterprise search and retrieval (ESR) discusses the capabilities of search engines to process content but also people's ability to find what they are looking for. Both the state of technology as well as psychological aspects play a role in this.

Section 7.1 uses a conceptual approach to ESR. This means it mainly elaborates on what enterprise search is, what methods are used to optimize results, and how the technique works in general. Sections 7.2–7.5 tie in with the EIM triangle (Fig. 7.1). The first part in Sect. 7.2, *Availability*, focuses on searching as a technical action of the search engine (bringing the content to the index). In Sect. 7.3, *Accessibility* is discussed under the umbrella term *Finding* as the human effort required to extract content from a machine by means of an index. *Relevance* is discussed in Sect. 7.4 in connection with the term *navigate*, which is a different way of presenting results and compiling results in the form of a clear structuring of the content. Lastly, in Sect. 7.5, the results of searching are processed into information and knowledge, which users can keep for themselves or share with others. We will focus on the aspect of *interpretability*.

7.1 Concepts

7.1.1 What Is Enterprise Search and Retrieval?

The term *enterprise search* is used to describe an application with capabilities to search for information in organizations (the search functionality itself and the results can be public). This is in contrast to the two other main types of search engines: Web

A. van der Lans (✉)
Advisory Services, Incentro, De Meern, The Netherlands
e-mail: anja.vanderlans@incentro.com

P. Baan (ed.), *Enterprise Information Management: When Information Becomes Inspiration*, Management for Professionals 2, DOI 10.1007/978-1-4614-5236-2_7, © Incentro 2013

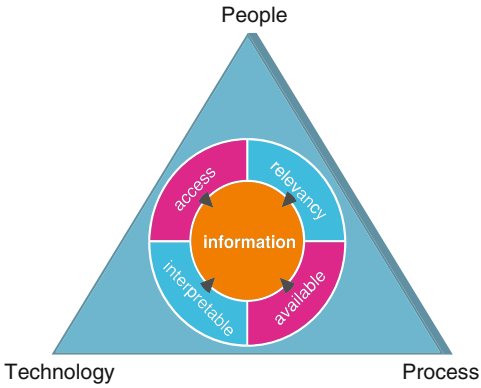


Fig. 7.1 EIM triangle

search and desktop search. Sometimes these types overlap, and search systems may contain multiple types of search engine. Enterprise search is about identifying specific documents (content) and making them available to organizations by means of an index, search capabilities, and result summaries that are presented to an authorized user.

> In September 2009, the Wikipedia entry on information retrieval (IR) was as follows: "Information retrieval is the science of searching for documents, for information within documents and for metadata about documents, as well as that of searching relational databases and the World Wide Web. There is overlap in the usage of the terms data retrieval, document retrieval, information retrieval, and text retrieval, but each also has its own body of literature, theory, praxis and technologies. IR is interdisciplinary, based on computer science, mathematics, library science, information science, information architecture, cognitive psychology, linguistics, statistics and physics."

As the preceding description of IR makes clear, a distinction is made between Web search, personal search, and enterprise search. Each type of search belongs to a specific content environment (also known as an information domain).

Web searches are used to search the Web or Web applications, for example, Google, Yahoo, and many other Web search engines. They are characterized by public access to the information, which means that all the available information is accessible to everyone.

Desktop search, which is also called personal search, is about indexing and searching personal information on a desktop computer and in the e-mail inbox of a specific person, in which only the information on the desktop that is saved by this specific person can be unlocked (this could be personal information but also business-related information).

If documents from a personal environment need to be shared, then they must be saved in a location that is accessible to others or distributed through, for example, e-mail.

Enterprise search is somewhere between public and private when it comes to unlocking information. Consequently, the amount of information (often a lot), the many different sources and document types, the protection of the sources, and various legal structures for the information need to be taken into account.

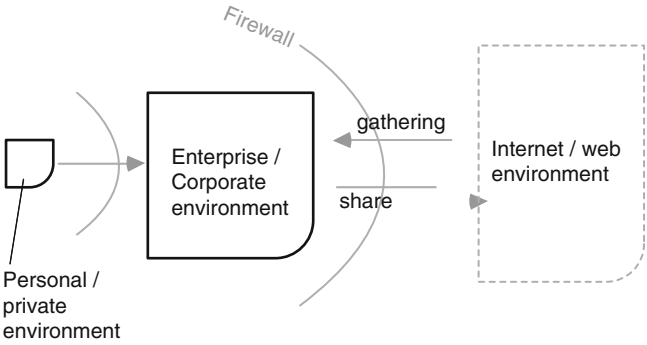


Fig. 7.2 Types of content environments

In addition, the weighting of the relevance of the content in sources plays an important role in the presentation of results. These aspects have been discussed frequently in the former chapers (EIM, ECM, BI) and they form an ongoing challenge in the field of ESR.

7.1.2 Three Types of Content Environment

In this book IR and enterprise search (ES) are considered to be one field of study because IR theories are used in ES, and therefore the two areas cannot be seen as being separate from one another (Fig. 7.2). The term ESR is used to indicate that no distinction is made between IR and ES.

Like most fields of study that blossomed in the heyday of the computer industry, ESR's foundation was laid long before the 1970s. The basis of the current structure of metadata, the basics of search and find, was formed by the unlocking mechanism of traditional libraries. In particular, IR requires a fair amount of mathematical understanding to be able to fully comprehend what is happening.

In this book, the idea is expressed that enterprise information management (EIM) can only occur when business intelligence (BI) and enterprise content management (ECM) are well integrated by means of a common set of terms (metadata) and when it can search this set and the related text (search) (Fig. 7.3).

7.1.3 Vision

An increasing number of companies have large collections of documents that they have built up over long periods of time. These companies have an increasing need to have access to this information or these documents, regardless of the storage location. The idea is that the end user does not need to know where information is stored, as long as it can be found easily.

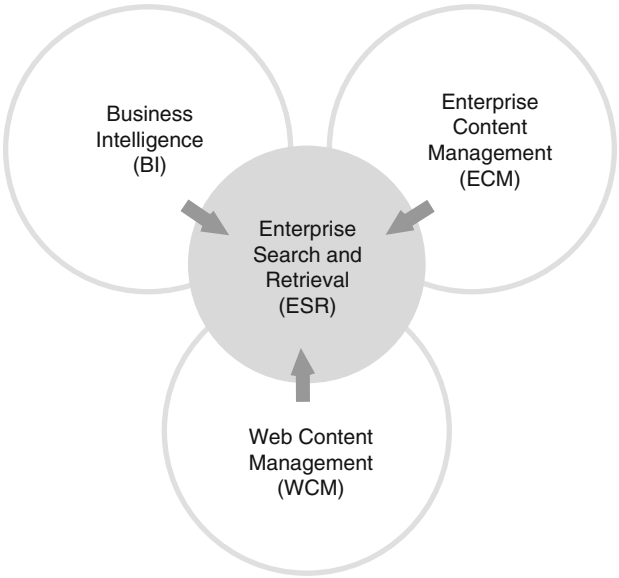


Fig. 7.3 Central structure of index from various information environments

In this context, the following subjects are often discussed:

- Protection of stored documents: during storage restrictions are placed on documents regarding who has access to them. These restrictions must always remain in place (examples are access control lists and active directory);
- Reduplication of information: if documents are stored in several locations, the end user only wants to see the documents in the results once (determine similarity rules);
- Relevance ranking: the most important results of a search must be presented first (specify sources and fields for the right order).
- Presentation of results: ways of presenting results in a navigation structure or list or based on another kind of structure;
- Profiles: with a profile, a user can be made aware of new content or results can be clustered into groups of terms that are meaningful to users (personalization);
- Integration of different sources: present search results from different source systems to users as one result set or in separate results for each source (federated versus integrated search);
- Interweaving of enterprise searches with Web searches and personal desk searches: there is an increasing need to include data storage drives of personal desktops or laptops in indexes and at the same time be able to see what can be found on a given subject online.

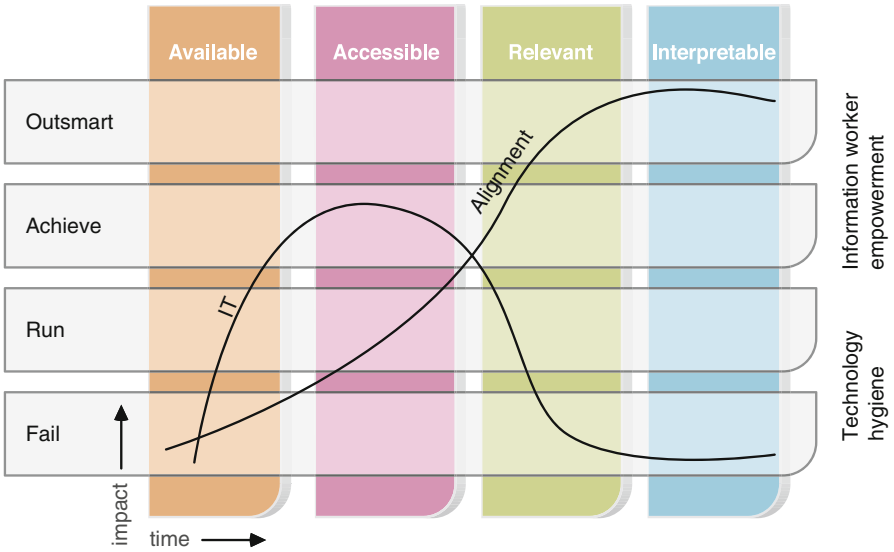


Fig. 7.4 EIM maturity model

It is expected that in the near future results lists will be created with one search interface, regardless of the original storage location of data (structured and unstructured; internal and external; documents and data). The capability to analyze and then extrapolate results for future policies will exist in the not-too-distant future.

Current technological developments in IR, and specifically analysis capabilities within the field, focus on uses for present-day news supply, pushing alerts on keywords, and the "thermometer function" in an organization. These analyses allow for better support in policy development and decision making.

Eventually, searching will be characterized in ways similar to text and data mining and will occupy a more important position in the information supply. The value of future search solutions will tend increasingly toward the field of BI and, therefore, also to the core business functions of many organizations.

7.1.4 How Does the EIM Maturity Model Influence ESR?

Besides the growing maturity of products on the market, it is also interesting to look at the maturity levels that companies can attain with search systems and at their specific characteristics. Based on this information, it can be determined what the current position of an organization with respect to search and retrieval is or what it should be. On this basis, the organization can take the next steps to make improvements in its business model. The maturity model shows the next steps in development of ESR (Fig. 7.4).

7.1.5 Fail

An enterprise search has not been set up. Access to file stores will be limited to basic search functionality.

7.1.6 Run

The first search engines within systems make their appearance. In some cases this will allow employees to conduct personal searches on their own PC or laptop. Also, search modules are present in third-party applications (OEMs).

7.1.7 Achieve (and Comply)

The first enterprise search application that is capable of searching multiple unstructured sources is purchased and set up. The organization gets to know the power of searching on top of the known navigation possibilities. Search supports transparency and decision making. Also in this phase, for the first time, BI and ECM come together in a common search engine. This leads to clarity in information, which results in better decision making.

7.1.8 Outsmart

In this phase, for example, sentiment indexing is used to look at the expectations of manufacturers, buyers, and consumers concerning product development or the development of services that are supplied. The search strategy is used in such a way that signals are captured and lead to new initiatives. Search is used to support knowledge management. The search engine integrates most or all internal and external sources to support knowledge management.

In practice the typical organization is now in phase 3, between run and achieve. Among other things, the information flows that must be processed by organizations are growing at a tremendous rate due to constant improvements in communication. This has major consequences for the way information flows can be managed and monitored and concerns both internal and external information flows. External information is created outside of an organization and internal information is created by the organization itself. This, coupled with the fact that the storage of information is becoming cheaper and cheaper, is the reason that organizations are not able to retrieve information when they need it. Much time is wasted searching. Efficiency demands improvements, and in most cases the needed improvements are achievable.

7.2 Availability

As mentioned in the introduction of this chapter, we focus on search as a technical action of a search engine (bringing the content to the index). Information can be made available very easily by indexing one source in the search engine but the information can have a much wider meaning than reflected in the index. In this section, we learn what is behind a search box and what the meaning for each type of usage can be.

7.2.1 What Is a Search for a User?

ESR is, as was briefly mentioned in Sect. 7.1, different from a search on the Web or on your own PC. However, users still want to see the advantages of a Web search or a personal search in an ESR solution. Users expect the following advantages from a search:

- It is very easy to do (a simple search field),
- It is extremely fast,
- It locates the correct information.

In enterprise search some differences must be taken into account compared to Web searches:

- The enterprise content sometimes has hardly any content, texts, or references; communication is as short as possible to increase the effectiveness of messages, and often the information that is assumed to be known has a long history. Short sentences, lists, and graphs are frequently used;
- Employees must securely store and retrieve information; unlike information on the Internet, proprietary business information must be securely handled and should not be accessible to others or employees not authorized to see the (sometimes confidential) information;
- Employees need the best/correct information, not the most popular information, as on the Internet; the ranking algorithm must ensure this. This requires some preparation and possibly later adjustments.

7.2.2 Requirements

Requirements analysis is an important aspect in the process of purchasing an application. Different user groups have different requirements and aspirations, and, depending on the importance to the organization, all requirements or parts of them will be met. In almost all cases the basic functionalities of competing applications will not differ very much. The difference in product X, Y, or Z lies in very specific needs of an organization.

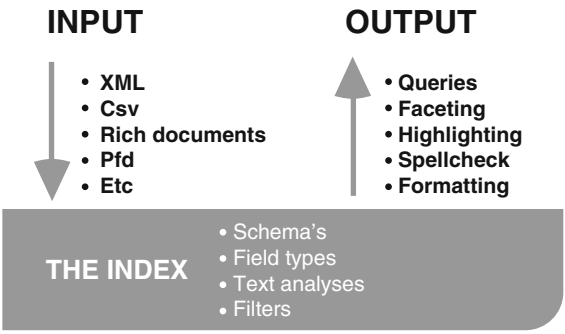


Fig. 7.5 Basic scheme for searching process

The preceding text shows that the decision to purchase a particular search solution cannot follow a one-size-fits-all approach. Each company must carefully weigh which sources should be unlocked and how the information from documents and objects should be retrieved and presented. A very precise description of the requirements will help to make the right choice for the organization.

Requirements should allow for the support of information workers on those points that are important for achieving the organization's goals. These requirements are important for selecting the right software applications or for being able to optimize the available software.

Policymakers are advised not to base one's selection of application solely on the desired future total solution. The risk factor for such a project is considered to be too great. It is important that all interested parties be aware of the desired situation and that the steps that are needed to arrive at this situation be well communicated. The sooner the first success is apparent to end users, the better the acceptance of the project within the organization.

7.2.3 Scope of Search

ESR solutions often need to be used in so-called cross-repository environments (Fig. 7.5). Thus, ESR must be able to regard the entire business environment as a source and should not be bound by a certain product repository or platform. Cross repository means that an index of data and documents is built within an ESR solution based on a large variety of sources: file systems (shared drives on internal servers), intranets, document management systems (DMSs/RMSs), e-mail, and databases/data warehouses. Based on the roles that need to be supported, the required sources can be selected regardless of their location or the format that is used to store content.

Broadly speaking, the process of establishing a search comes down to this: there are documents in a certain format (input) that are offered to a search engine, which makes an index of them (a simple task using analysis or filters) and then lets users ask questions in a specific form and produce results.

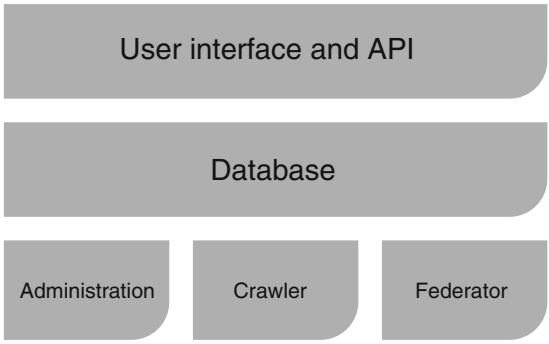


Fig. 7.6 Enterprise search architecture

7.2.4 Search Architecture

An enterprise search environment consists of five cooperating components (Fig. 7.6): a database, a crawler, administration, a user interface and application programming interface (API), and a federator.

Crawling is a technique used for the inventory of unstructured data. The crawler is a computer program that browses through sources in a methodical (systematic) and automated manner, like a robot. A source can be a collection of documents or texts on a company network (internal) or outside the organization (external). It may also be a Web site on the World Wide Web. In the latter case, we speak of a Web crawler or a spider. Further explanation of the term is limited to the activities within a company network. A crawler (also called a harvester for activities within an organization) starts with a list of URLs and locations that are to be explored. The crawler visits them one by one, and all documents that are found at the URL will be inventoried and mapped. This results in a list of terms (e.g., words, phrases) that occur in the documents. This list is called the index and is triggered when processing a search argument. Based on the match between the searched term and its presence in the index, the search engine decides whether or not the desired information has been found.

Administration makes it possible to configure the entire environment and is used to schedule the crawler (perform actions according to time settings), configure the environment(s) that the crawler visits, run reports, and so on.

The results of the crawler, the administrative components, and the federator are stored in a database. This forms the basis for the results a user will get when performing a search through a user interface (UI) and the API.

The federator makes it possible to integrate the results of other environments in the proprietary search results (see also federated search).

The UI and the application's links with other programs (API layer) result in a look and feel that can be adapted to the organization's usual interface.

The figures in the text show the parts of a search engine without details. Today's search engines have added several indexes that appear after being admitted by a parser (which divides them into machine-recognizable elements). The user asks a question (user query), and the parser passes the question on to the indexes and runs

a spell checker on it. Then results are found in the index and, based on preference rules for scores, they are translated in a certain order and presented to the user as a results page.

Behind the scenes there is considerable activity happening before a user gets to view the results. The more importance a company attaches to supplying relevant documents, the more focus can be put on the system's learning (MLR = machine learning ranking), so that the relevance of the results set can increase based on the internal needs of the company.

7.2.5 Search in All Shapes and Sizes

Depending on the capabilities of the search engine, the source you want to index, and the extent to which metadata are included in the sources, a certain method of searching and presenting is selected. Here are the most common methods:

- Federated Search
  A federated search is the simultaneous search on multiple online databases or Web sources. This kind of search is being used with increasing frequency on automated, Web-based library and information systems. The term *federated search* is also often used as a synonym for meta search (simultaneously searching multiple sources).
  A federated search sends a request to several other search engines or databases and, based on the individual results, composes a list of the information found by source and presents this information to the user. A federated search makes it possible to feed several search engines and put them to work simultaneously by entering search criteria only once.
  The challenge in federated searching is scalability. If the number of sources increases, the performance of the search decreases. Federated searches often work in large environments like the Web, but they are less applicable in a limited business environment. The limiting factor is the lack of a balanced method for determining the relevance of the results. Within each source, the relevance of a result can be different because the sources use different algorithms to determine the relevance (Fig. 7.7).

- Integrated Search
  Integrated search works differently. As with federated search, it is possible to query several source systems simultaneously. However, the search engine is configured by source system in a way that allows a particular search engine direct access to the source data. The search engine then indexes the indicated data of the source system in a central index (this can be very large, depending on the number of sources and the size of the collection per source). Once the query has been made, the search engine will be able to rapidly present search results to the user based on its own index. Depending on the user's preferences, the results will be sorted based on relevance, date, or source.

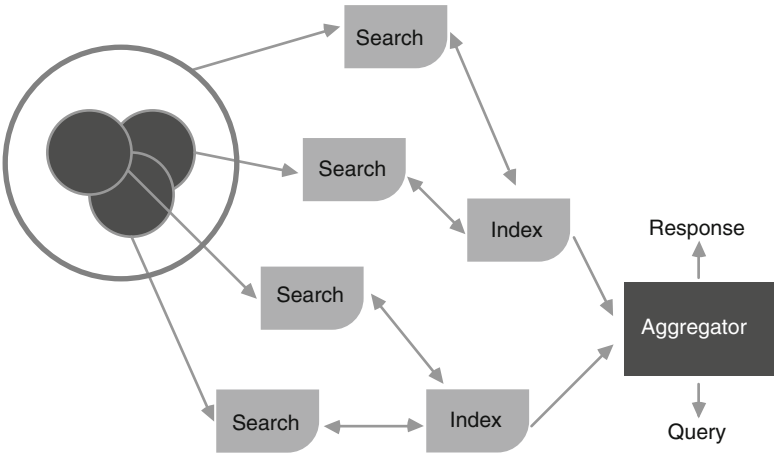


Fig. 7.7 Federated search

- Boolean Search
  A Boolean search uses a search in which documents that contain certain words are excluded or are specifically searched for. The words AND, NOT, and OR are often used for this, but also the plus symbol for AND (+), comma for OR (,), and minus for NOT (−). Not all search engines support Boolean search. Google, for example, automatically uses Boolean search behind the scenes and in advanced searches, so that you only have to use the minus symbol to exclude certain keywords. AND need not be used at all. Google automatically searches for all words in the search box.

- Concept Search
  When indexing, a link is created between one term and other terms that often go hand in hand with that term. A search will not only give search results but also other search options that fit the same concept. When someone searches for "holidays," for example, he will get, in addition to the standard results, the possibility to continue the search specifically for, for example, last minutes. Based on these concepts the search can quickly be refined to the results that are sought (Fig. 7.8).

- Natural Search
  A natural search is simply a search, often on the Internet, that only shows unpaid or organic search results. A synonym for natural search is therefore organic search; it constitutes a natural, realistic display of results, without manipulation of the search engine. Using a natural search, relevant results that are arranged using the algorithms of search engines are displayed. Pages can be listed higher in organic results by properly optimizing Web pages for search engines. This is also called a search engine result page (SERP). Unpaid or not paid in this context means that no fee is paid to the search engine to attain a higher ranking (known as, for example, Google's paid list).

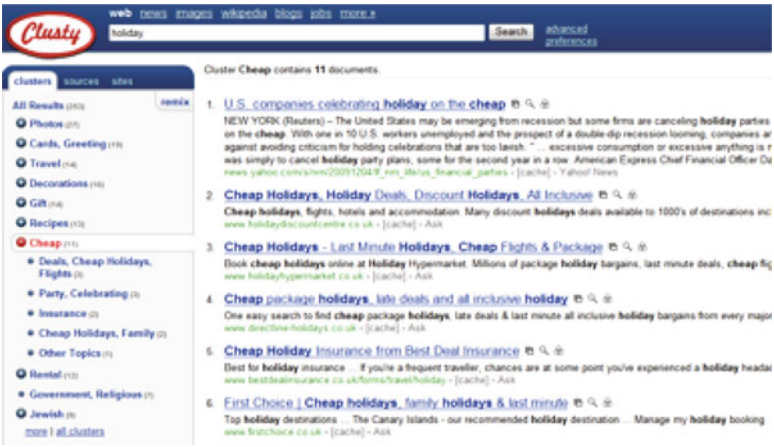


Fig. 7.8 Conceptual search on the Internet (left column) (source: www.clusty.com)

- Proximity Search
  A proximity search is also called a phrase search. It is a search in which users indicate that the results should contain pages in which the search words can be found together or in close proximity. In most search engines, it is possible to search by inserting the phrase between double quotation marks (for example: "enterprise information management"). This method makes it possible to quickly find the right combination of keywords in the results.

- Fuzzy Search
  A fuzzy search is characterized by its broad interpretation of what is searched. This means it also produces results for a search when words are misspelled or only partially fit the query or when the search phrase is partially entered. When someone searches for "applepie," the search engine will allow the user to search for "apple pie" instead. And with a compound search such as "enterprise information management" it will first show results for the compound term before showing results for the individual term "enterprise" or "information" or "management." With fuzzy search there seldom are no results to be shown. The disadvantage of this method is that it yields enormous amounts of results. Fuzzy search is not very suitable for organizations because too much time is spent on viewing irrelevant results.

- Faceted Search
  Faceted search is also called clustering. A cluster is a set of interrelated subjects, also called facets (and faceted search follows from this). Because of the links between subjects, one can use a searched item as a starting point to search for related matters. Searching for the word "wave" may result, as a cluster, in "sport," "science," "movie," and "water." By clicking on "sport," one gets results that have to do with sport (and not necessarily "wave"). Often, both clustering and concept search are used in the field. This is not illogical since the results are presented to the user in both cases. In faceted search, this is done based on a

predefined structure design (taxonomy, thesaurus; see more in Sect. 7.4), whereas in concept search this is done on the basis of concepts that are identified by the search engine according to the supply of the index, so without a subject structure. This means that in the case of concept search the presented classification may be different today than it is tomorrow depending on whether new information becomes available.

- Full-Text Search
  In full-text search the number of times a word is used in a text is counted. This is then placed in a hierarchy. A word that occurs often in a text will be considered an important word by the full-text engine. In a query, one or more keywords are entered. These are compared with values in the database. The document with the highest score will be the first search result shown onscreen.

- Semantic Search
  In semantic search indexing is based on the relationship between words. The engine builds a semantic database in which relationships between words are described. When a query is executed, a semantic search engine can offer a visitor alternatives to a search. This more natural way of searching leads to finding information faster.

- Natural-Language Search
  A natural language is a language that people use, both spoken and written, as a means of everyday conversation. English, Spanish, Catalonian, Chinese – all are natural languages. The concept of natural language is that a query can be made in a natural language, like a normal sentence that can be spoken to another person.

  An important challenge for information science is the development of computers that are able to understand natural language. In information science, the interaction between computers and humans based on (natural) languages is the field of natural language processing (NLP). Every language needs to be analyzed and divided into lexical (linguistic) functions after which the relationships between the parts are fixed in the index. As a result, combinations with the same or nearly the same meaning can be stored, and searching becomes possible.

  Thanks to NLP techniques, the first search engines to process searches through natural language are now available on the market and are used in, for example, Frequently Asked Questions collections.

  The results for use in an entire organization are still under discussion. For use in certain departments with limited or clear domain knowledge, NLP has resulted in positive experiences.

7.3 Accessibility

After deciding which source will be made available in the search and which group of users will have access to the search, there is always the question “it must be somewhere, I know it exists, but why can’t I find it?”. This chapter is not about

search systems and methods of search (Sect. 7.1) but about who gets access and how to improve search results on the input part of the system ("under the hood").

In many cases, organizations come to the conclusion that they do not get good results when searching for information or documents. How come? The cause can often be found in the storage of documents and the lack of applications that adjust the results for the organization. In other words, enterprise search alone, i.e., purchasing an application, is usually not enough. The organization has to invest in the conditions surrounding it. This brings us to the field of IR.

IR is the umbrella term used to cover a number of terms related to retrieving structured and unstructured information. The terms used are as follows:

- Unlocking methodologies: tools for looking up information;
- Retrieval performance: the extent to which software retrieves information;
- Content classification: ways to store information with more metadata;
- Document analysis: research into the degree to which information can be retrieved;
- Search engine applications: software to search for information;
- Content source integration: the extent to which various sources can be unlocked at the same time;
- Security levels: the extent to which users are allowed to perform actions with sources, systems, or content/documents.

In the IR process it is important to look closely at how documents are created, what data are added to the documents (metadata), how and where the documents are stored, and what tools are available to retrieve the documents.

Data added to documents (metadata) need to be managed and can be used to impose some order on the relevance of results. Metadata can be added through automatic extraction of characteristics of the documents or by assigning terms manually. To be able to add uniform terms and to use them in the documents, a structure must be determined for the set of terms. We come across sets of terms in different kind of shapes like taxonomy, standard lists of keywords (controlled vocabulary lists, or CVLs), folksonomy, thesaurus, or ontology.

Figure 7.9 indicates that in the realization of an integrated solution within EIM one must think of centralized storage of content and data, at one location in one index, so that results are obtained by means of searching and navigating.

7.3.1 Goals

Goals can help clarify who is searching, what is being searched for, and what is done with the found information. The starting point in ESR is that users may only find information they have access to (which may be found and accessed).

It is important to define which roles within an organization use enterprise search, in other words, which groups of knowledge workers exist in the organization. Each group will have its own approach to obtaining the desired results. This approach

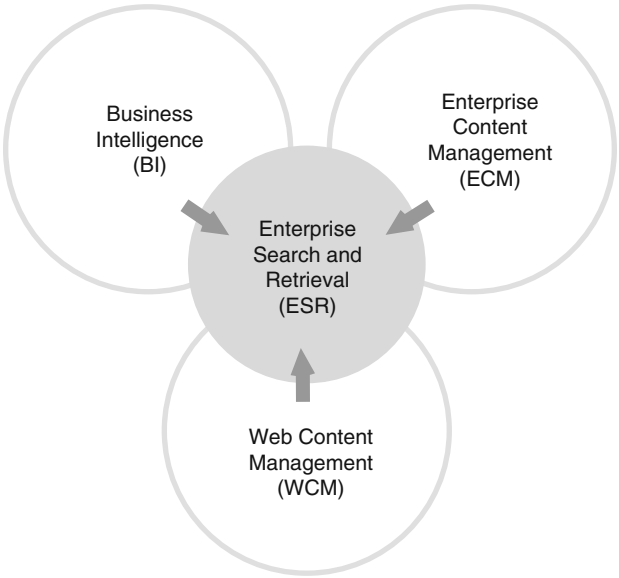


Fig. 7.9 Central structure of index from various information environments

must be respected and implemented in order to achieve optimal use of the ESR solution. This can often be realized by setting group profiles in the search environment according to the group's role, level, or function in the organization.

Dividing information into big chunks helps users to quickly determine the relevance of information. Project information, sales information, and product information are chunks that are recognizable within the organization and are therefore considered a meaningful format for a search application.

When the goal of the information is clear, the search can be adapted accordingly. The search can be extended with aspects connected to the goal. A search for annual figures of a company can be extended to the figures of a similar company. The intention is not to create an information overload situation in setting up an enterprise search environment but to be in line with the perceptions/needs of different user groups.

7.3.2 Enriching Content

Enterprise content is often characterized by a lack of context, texts, and references. Therefore, it is important to be able to enrich content in order to enhance search results. Enriching in this context means adding characteristics to the content. These characteristics acquire extra value in determining relevance within the ESR solution. As a result, this content is found easier and faster and is ranked higher in the

result set. There are some methods to enrich content, but in practice a mix of several methods is often used.

These methods are as follows.

1. Manual supplement in which an employee selects certain terms from a standard glossary or thesaurus/taxonomy and makes the necessary additions in fields belonging to the text. Although this method significantly increases accessibility, it cannot always be applied in extensive content environments for practical reasons: it is too expensive and time consuming. Moreover, it meets resistance by employees. Experience has shown that an additional three to four terms is often the maximum for the average user. In certain cases, professionals will be asked to add characteristics, but even then the size of the collection plays a part.
2. Metadata are generated using an automatic supplement based on the available thesaurus, taxonomy, or CVLs. Despite the high efficiency, incorrect results may occur and the organization must consider providing intensive support to optimally train the system and to learn how to deal with the organization's specific domain knowledge.
3. Tags that are included at the moment a search result is achieved is an intermediate method for enriching content. When this method is used, it is not necessary to enrich content as long as it is not part of a search. Once interest is shown, an employee adds metadata so that future users will benefit. The work is, so to speak, spread among all possible users. This method also includes forms such as social marking and the use of folksonomy. In both methods, ratings, terms, and qualifications are left to other users.

7.3.3 Access to Information

One of the most laborious aspects of creating an enterprise search engine is making sure that users only see those things that they are granted rights to in the source systems. In other words, the rights structure of the source system must be inherited. Often a complex security system is chosen for information in the source system. The more important a source is for the continuity of a business, the greater the importance that is attached to the issued rights. This must be maintained for all sources at all times.

The ESR machine must decide for each source and person whether the results of a query can be shown in the results list. When results are displayed, a choice can be made between showing all results and showing only those results that the user is allowed to open or view. Usually the latter option is chosen to prevent sensitive information from coming out just by showing the basic information about a document. In organizations where knowledge development is important, it is sometimes decided to show the basic information as part of content discovery. A contact person monitors the requests for actual access to the document. This so-called gatekeeper decides to allow an applicant access to the document or not, based on the function, level, and confidentiality of the document, or to send the document separately.

When saving document data one can choose to set the rights of employees such that one can check whether a user is allowed to view a document. This is called early binding because it is done early in the process. This could also be done as a final step before a document is included in a results list. This is called late binding. In recent years, search engine suppliers have often chosen a combination of both where the default rights are set but checked for changes just before information is delivered. Of course, all this is done to guarantee the good performance of a search engine, in the sense of fast delivery of results to the right person.

In enterprise search it is important that the available information can be found quickly, based on the search term(s). Speed is at odds with the number of sources, diversity of sources, size of information collections within the sources, and several indexes, filters, and lists that the information must be checked against before it can be disclosed. In many cases, speed represents a compromise between functionality (much can be done), technical facilities (with enough processors), and investments (enough money) made by an organization.

Cooperation is needed between the index that is built and the keyword structures or taxonomies that are important within the organization. A branch-specific taxonomy may be used in addition to a general (linguistic) taxonomy. The design focuses in particular on indexing the results of crawling and adding meaning to it. It is possible to involve several languages in this analysis using multilingual digital dictionaries. A common framework of reference (Latin or English) can be useful. All these enrichments, filters, and access controls take time, and as a result, enterprise search solutions often have lower performance than personal searches or Web searches where not all these steps are needed.

In this book, we have spoken several times about making words, terms, and numbers available in an index. It takes quite a bit of work to reach this point. The text analysis approach consists of four steps:

1. Prepare the basic components,
2. Organize the source data,
3. Specify the processing,
4. Analyze the results.

The preparation and organization of the source data will be handled here, whereas specification of the processing and analyzing the results will be covered in Sect. 7.4.

7.3.4 Preparation

The basic components that are necessary for analyzing a text are as follows:

- *Stop-word module*
  The stop-word module ensures that all stop words are excluded from the index. The list of stop words can be filled, modified, and updated (manually). Usually it starts with the words *a*, *an*, *the*, etc.

- *Synonym module*
  Defining synonyms is often a time-consuming task simply because many synonyms are used in the various documents and content. However, you have to start somewhere. "Good is good enough" is a sensible starting point. Starting small is perhaps even more sensible. You can always add more! Of crucial importance is to indicate what needs to be done with the synonyms when they occur in a text. The two most commonly used variants are making them equivalent and replacing one term with another. In the latter case, the reader is referred at the nonpreferred term to the preferred term (for example: *bike*: see *bicycle*).
  To improve the index, along with the search result, it is possible to create links between terms in the index. The link can be between words that go together (bike = bicycle; synonyms), hierarchical links (bike, moped, bus = means of transport), preferred terms of reference (*AH* see *Albert Heijn*), or language alternatives (*bike*: see also *velo*, see also *fiets*).

- *Homonym module*
  The homonym list consists of three elements: first, the homonym itself (the abbreviation of a certain term), second, a description to replace the homonym, and third, some content indicators under which the replacements can take place. For instance, "S.I." may be replaced by "stroke index" when the context is "golf" and "course specification." Or it might be replaced with "Systeme International d'Unités" when the context is "units of measurement" and "standards." In taxonomy terms, this is called creating scope notes, or defining which term is used in what situation.

- *Thesaurus module*
  Creating categories or a classification structure based on thesaurus terms enables indirect search. By including :putter," "driver," "iron," and "wood" in the category "clubs" (synonym for golf sticks) one gets a broad and more complete result than without the category. This is creating hierarchies in terms, in which "clubs" is the broader term for the other terms, which are called narrower terms. More on classification systems can be found in Sect. 4.4 .

- *Origin module*
  In the origin module, words are linked to the Latin or Greek root to which they belong. This enables multilanguage search, provided that the words belong to the same language group (examples are language families like the Germanic languages or the Romance languages). Reducing terms to its basic forms, also called lemmas and morphological transformations, can only work when syntactic dictionaries and grammar rules are included in the search engine and can be applied systematically.

  Text analysis is done by searching an unstructured document based on an argument or search criterion. A hit means that the search criterion has been found in at least one location. Besides searching based on arguments, a search engine can also recognize patterns if they are included in its task for indexing. Examples are a telephone number and Social Security numbers. One could also include texts that follow fixed texts or fixed positions such as name, address, or invoice number.

When additional characteristics in an index are given to words, numbers, or features in a text based on this information and the results of a certain form can be presented more quickly, this is called entity extraction. The ability to identify certain structures or patterns in a document with entity extraction, without additional metadata, has been an important development for many organizations. This allows them to bring structure to a document later on (after it has been in a file system for years) if they feel it is useful (for example, production numbers, organizational structure of departments, and project numbers).

Basically all words in a document are included in this index. Some categories of words can be excluded, for example, articles and adverbs. A search will use the index and users will therefore obtain results extremely quickly. Only when the search zooms in on a hit, when the hit needs to be opened in the results list, is the real source (e.g., document, photo, e-mail) consulted.

7.3.5 Organizing the Source Data

Source data must be supplied in an electronic format. If the source does not store the information in the right format, it has to be converted to the proper format first. Most search engines can cope with virtually all major formats. At the beginning of the project, when creating the order in which sources are indexed, it is smart to mainly choose sources that the search engine can process well.

For enterprise search solutions, not every document or page that is found is important. Users attach more importance to one source than to the other. Therefore, weighting factors of certain sources can be determined. The same goes for certain types of documents. In most cases an authorized document from the management will have a higher value than the company's restaurant menu.

Also the location of terms within a document is important to determine the value. A term that is found as a keyword (especially in manual enrichment, so added by means of human intelligence) should have a higher value than the same term found in the body of the document (Fig. 7.10).

7.3.6 Types of Subject Structures

We use metadata to bring structure to unstructured information. This metadata can consist of several parts. One of these metadata is frequently used: keyword.

In order to bring structure to a collection of documents or objects, based on the subject content, they are categorized into classes. This process is called classification. The different kinds of structures are called classification systems (numeric or alphabetic). These can be very simple but also very extensive and complicated. The types of classification systems are discussed in the upcoming pages.

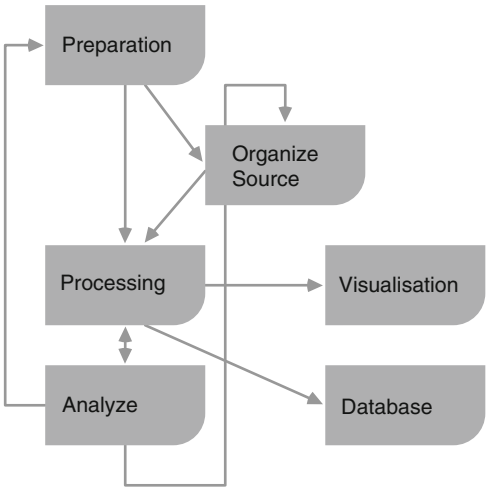


Fig. 7.10 Methodology for text analysis

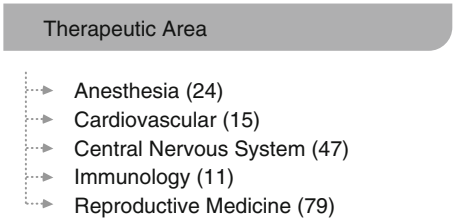


Fig. 7.11 Structure search results in therapeutic areas

A keyword is a term that best describes the content of a document. The term can be singular (car) or compound (hybrid car). The term can be freely entered or chosen from a list. When it can be chosen from a list, the simplest form is an unstructured alphabetical list with chosen terms. This list is called controlled vocabulary list (CVL). When layers are starting to form in the list (for example two layers in which the first layer only says 'departments' or 'locations' and the second layer contains the names of the departments and place-names) then this forms the basis for a structured CVL.

Such a list can be used in the search engine as a filter for the entity extraction. After the information has passed the filter, the structure of the terms can serve as navigation in the result screen and indicate how many documents are identified which have this term somewhere in the text (or a field yet to be identified). This can then lead to a navigation structure for, for example, the therapeutic research areas of an organization, which is presented in the search engine as displayed in the example below (Fig. 7.11).

Within many organizations it is possible to order the subjects, themes or processes that are executed. Sometimes this only exists inside the employee's mind (implicit) or it is set out in a structure of the file store, filing cabinets and databases

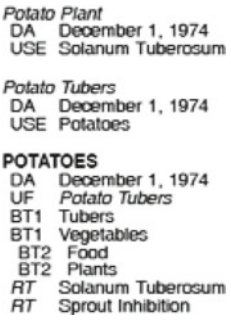


Fig. 7.12 Search structure (www.etde.org/edb/etdesuth.pdf)

(explicit in a topic structure, thesaurus, taxonomy). When there is a good structure in the links between terms, the correct information is found more easily.

The moment the structure becomes more complicated or it is important to indicate the relationship between different terms, we use the term thesaurus. Thesauri have been used in libraries to place books on the same topic together (either physically or in a catalog).

The terms of a language are categorized and compared to related terms. A thesaurus uses, apart from the terms themselves, broader terms or parents (BTs), narrower terms or children (NTs), related terms (RTs), USE (see relation; whereby at a nonpreferred term the reader is referred to the preferred term; bike USE bicycle) and UF (used for; see relation; indicated at the preferred term, where from the nonpreferred term a reference is made to the preferred term; bicycle UF bike). A structure may look like this (basis) (Fig. 7.12):

Each term and each level in the thesaurus can be given a certain value that affects the score, which determines the relevance of a result. In other words, when searching for term "x" all documents that contain that term and its synonyms get a value (weighting factor) on the basis of which they are presented in order of value (score in terms of relevance). The document with the highest score (whose content most closely matches the requested term) is presented first in the results list.

The creation of a thesaurus requires a lot of domain knowledge and is a long and intensive process. The result is more or less static. It is used for many years and rarely changed (additions may be made) (source: http://nl.wikipedia.org/wiki/Taxonomie).

Taxonomy can be helpful when a certain structure is desired. A taxonomy is a standard description and classification system. It contains the terminology (homonyms, synonyms) and jargon of the working field or field of study, and these terms are properly classified and grouped. Often a tree or network structure is used. Based on the actual situation found in the document/information to be searched for, a taxonomy can be created from different perspectives and be changed easily.

A taxonomy is a different kind of classification system. A taxonomy refers to the classification of things as well as the underlying method for this classification. Almost anything can be classified taxonomically. In recent years, within the fields of computer science and artificial intelligence attempts have been made to have

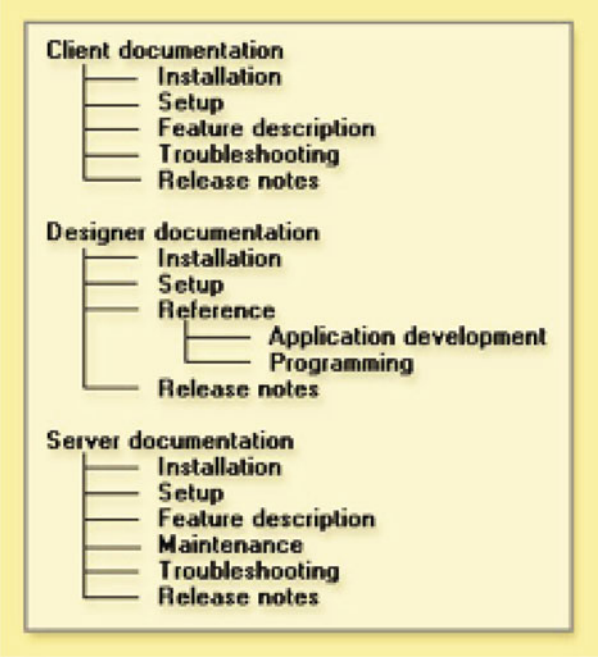


Fig. 7.13 Elaborated taxonomy of IBM for classifying documents (source: http://www.ibm.com/developerworks/lotus/library/ls-Kmap_tax/index.html)

machines automatically create a taxonomy of a collection of objects. An example is the automatic classification of a group of documents for digital libraries.

In this field, a distinction is made between taxonomy and typology. The main difference is the way a classification is created (in computer science this is called the classification algorithm). In a taxonomy one starts with a group of exemplary objects that one is trying to subdivide. Next, the characteristics of the objects are examined and the taxonomy begins to take shape.

In typology, a different approach is used. First, one thinks of the distinctive characteristics potential objects would normally have, and then the actual objects are sorted based on these rules. In that respect, typology seems more like a thesaurus. One could say that taxonomies have an empirical (inductive) origin, whereas typologies and thesauri have a more conceptual origin (deductive).

Before a classification system can be created, some thought must be put into what aspects will be displayed, that is, a scheme should be created first and then the classification system. In Fig. 7.13, the taxonomy consists of only three layers, and the starting point is product lines. Other classifications are also possible if another aspect is chosen.

In practice, the terms thesaurus and taxonomy are used interchangeably. Although they are not exactly the same thing, they are both used to classify unstructured information and can be of great value within an (enterprise) search solution.

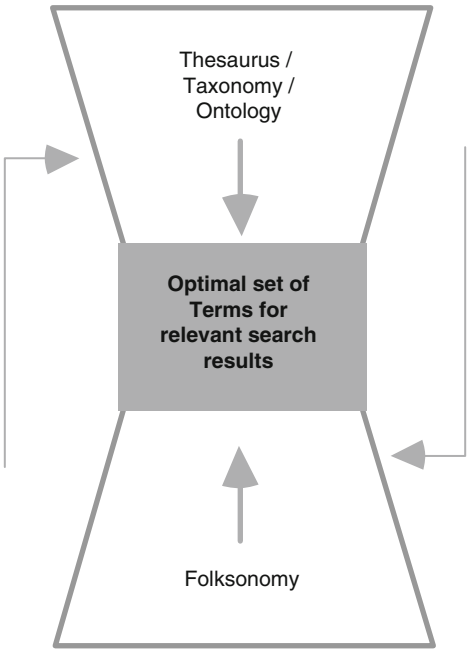


Fig. 7.14 Optimal set of terms for access and relevant search results

If a thesaurus or taxonomy is available, then it can be used to add metadata to unstructured information, which makes it possible to display the results in a navigational structure.

7.3.7 Folksonomy

Thesauri and taxonomies attempt to bring structure to subjects in a certain field, as much as possible from a theoretical, professional point of view. The terms from the structure are added to the documents as metadata before being added to the index in order to improve the relevance of search results. This is a top-down approach (a structure is created and then becomes the sole structure authorized for use) (Fig. 7.14).

A folksonomy does the exact opposite, that is, the end users are offered the opportunity to come up with terms with which they found a document or had expected to find a document. These terms are added to the document as search terms (tags) in the index and the document is thereby improved and made more accessible for future use. These terms can also lead to changes and additions in the thesaurus with synonyms, references, or new terms.

The added value of a folksonomy is that it provides insight into user search behavior and thinking about a concept. A term from a tax declaration is a good

example: the tax authorities use “tax partner” as a thesaurus term, whereas a user of the site will search for “husband,” “wife,” “spouse,” or “roommate.” Making good use of the actual terms that are used in queries will improve the quality of the results and enhance users’ experiences with the site.

7.4 Relevance

Relevance most commonly refers to topical relevance or aboutness, that is, the extent to which the topic of a result matches the topic of the query or information need.

Relevance can also be interpreted more broadly to refer to generally how “good” a retrieved result is with regard to the information need. The latter definition of relevance, sometimes referred to as user relevance, encompasses topical relevance and possibly other concerns of the user such as timeliness, authority, or novelty of the result (http://en.wikipedia.org/wiki).

What is relevant for a user depends on the situation, and whether the information is relevant for a user depends on the time delivered, the number of documents or amount of information delivered, and the form in which the information is delivered. Only if all elements suit the users will they be satisfied with the results of their search.

In this chapter we look at users and their behavior with regard to searches and results; in addition, we will have a further look into the more mathematical definition of relevance as a recall and precision calculation.

7.4.1 Defining Search Results for the User

Because it is impossible to predict who will search for what and where, it is important to acknowledge logical search types. An analysis of current search actions (usually recorded in log files) often provides important clues as to the type of users within an organization. The analysis provides groups of users that search for:

- Other employees who have a particular expertise. This type of search provides an alphabetical list of names in which the expertise is categorized (much knowledge, average, no knowledge). Contact information is often also relevant in this context;
- Information related to business processes. These employees search for terms that occur in a taxonomy that serve as a starting point or basis for finding information about the desired process;
- Business objects based on their keywords, with the result that all the sources that have information on the subject can be combined.

Every search type is linked to a default preference for presenting results. The “person searcher” wants a table with information, preferably presented with links to e-mail addresses or telephone numbers. The “process-related searcher” wants to see

what subprocesses are defined in order to reduce its search to precisely the thing he is looking for. The "keyword searcher" wants the most relevant document as quickly as possible and would like to influence the relevancy of the results as well. Again it is clear that, based on the size of the groups of the search characterization, different forms of presentation and results must be taken into account; once again the approach is not one size fits all.

As previously mentioned, it is important to use goals to clarify who is searching, what is being searched for, and what is will be done with the information that is produced. It is important that the information being sought be made available quickly. The idea is that users may only find information that they are allowed access to (documents may be found and opened/viewed). Users are not all the same, in other words, there are different types of search behavior and different preferences for the way results are presented – through a search box or by navigating within a structure – and what each group will call relevant information.

7.4.2 Behavior

Research has shown that users and user groups have preferences when it comes to information retrieval. The type of search is related to the purpose of the search. There are roughly four types of search behavior and associated search methods. Each type of search behavior has its own characteristics and requires specific applications of the search engine, some functionality, to support it.

Most users will apply various types of search behavior, depending on the situation.

1. *A known time*

   The search for a known fact is the easiest to understand. These users (1) know what they want (100% relevant results), know what words to use to describe it, and have a reasonably good understanding of where to start.

   In addition, users may be happy with the first answer they find (though not always), and the task may not change significantly during the process of looking for the answer.

   The following items would be very helpful in searches for users of this type:

   - Search box: this is a very good solution. People can articulate what they are looking for and are able to type it in the search box. As long as the search engine shows the results of a word in context or shows a clear description of the results, users are likely to recognize suitable pages from the search results.
   - A-Z indexes: these are important for supporting users who are able to specify the word they are looking for. As long as the A-Z index contains the user's word, all she has to do is read down the list and spot the right item. One way to make sure the list of terms in an A-Z index matches the words users think of is to look at the terms that the users use during the research or to search in logs (log files save all queries and give an idea of the terms that are used within the organization).

- Quick links: links to frequently used items allow easy access to them. Again, the terms in the list must match the users' terms; when they are recognized, this adds value.
- Navigation through a subject structure: guidance by a structure can support the activation of the memory. For these users it is effective to choose one term from a range of terms that fits the image they have in mind.

2. Explorative
   In an explorative task, users only have some idea of what they need to know. They cannot exactly put it into words, and if they can, they are unable to use the right words/terms. They do not know where to start looking/searching. Most of the time they know when they have found the right answer, but they do not know whether or not the information is sufficient or if what they have is all there is to be found. Relevance is based on aboutness.

   The information needs of these users change rapidly when the gap between their vague idea and the found information closes.

   These users are assisted in their search by providing:

   - Navigation; this allows users to try to explore a subject themselves by following a path, discovering and learning as they go. While navigating, users may go deeper or broader into a hierarchy or to related information;
   - Related information: related links may be created from a list of related topics, a manually created list of relevant pages, or lists based on items purchased or recommended by other users. Contextual links in the text may also help to give more direction to what is sought;
   - A search box; a search can be useful for explorative tasks but can be problematic due to users' inability to articulate what they are after. An initial search can help users learn about the domain and get some ideas for keywords. It can also be useful to provide synonyms for the search term as they may help users to better articulate their query.

   For this mode, it is critical that there always be avenues for exploration and that the visitor never reaches a dead end.

3. *The user does not know what he needs to know*
   The key concept behind this mode is that people often don't know exactly what they need to know. They may think they need one thing, but in fact they need something else to be informed about what it is they actually need to know.

   This mode of seeking information occurs in a number of situations when users are searching for information:

   - About complex domains such as law, politics, or finance. For example, a staff member may want to know how many weeks' maternity leave they are entitled to, but they might need to know the conditions surrounding that leave. Another example is studying terms and conditions of new products and services. There may be important restrictions, but they are too often buried in legal garble, which the layman may have a hard time understanding;

- To promote certain information. The average user will not use the same jargon and terms for a certain product or service as an expert would. It is impossible to find information when key terms are missing;
- About unknown domains, for example, when someone is told by friends that he or she should check out a new service, product, or Web site but does not yet know why he or she would want to know about it. Much of the available information will not be meaningful to the user at first because his or her "receiver" is not switched on, so to speak;
- To remain up to date. People often want to make sure they are up to date with what is happening within an industry or topic, but they are not looking for answers to specific questions.

The challenge for all searches is to provide results that allow users to get the right information out of the results list. That means the user must be enabled to find out about a particular topic. This can be achieved by:

- Straightforward answers and results. Simple, concise answers allow people to have their initial information need met. For example, in the four situations described previously, users would be helped by a summary of maternity leave benefits, key issues of concern in the terms and conditions, an outline of the benefits of a new Web site or service, or a list of the latest releases respectively;
- More detailed information. After simple information has been given, a need might arise for more detailed information. When this information is easily available, for example, in the form of related links or contextual links in the body of the content, it is quite easy to gather more information on a subject.

The challenge for this group of users is to satisfy them with answers and then give them the opportunity to find additional information before they give up. Whether the information is relevant will be found out over time.

4. *Refinding*
This mode is relatively straightforward because people are looking for things they have already seen; the location is more of a challenge. It is possible users will remember exactly where the information they seek was located, that is, remember what site it was on, or they may have forgotten where it was. For many users, searching for information they have already seen is very time consuming.

To make it easier to refind information, one could come up with reminders such as lists or favorites. Users create such reminders when they think the information might be useful in the future. In most other cases, users trust their memory.

Many Web sites help people by creating such reminders: task lists, "save for later," and favorites. Besides users' conscious actions, some information systems (generally Web sites) also offer solutions to make things easier for users when they return to the site (passive solutions). When found, the information is recognized and 100% relevant.

7.4.3 Mathematical Relevance

Optimally, a search will provide only relevant results. This is never achieved because relevance is subjective and can be defined differently for each user and for each search (especially in an enterprise environment).

Based on the alignment of the value of the information in the sources, a mathematical formula is created that determines which of the findings are presented first and which last.

Relevance is an important way for presenting found results. However, in some cases it is more convenient to sort by title or to get a chronological overview by searching based on date.

A relevance of 100 % means that the found document contains the term that was sought. Usually, the relevance is determined based on the number of times the term occurs compared to the total number of words in a text. The appearance of a word in a short text gives a higher relevance than in a long text. To many users relevance is the most important criterion for presentation (other criteria are date of publication, source, and author).

The ultimate goal of information retrieval is to find the correct information in a minimum amount of time by navigating or searching. Correct information is defined as information that meets "the expectations of the end user." This means that, for example, in every environment expectations may differ significantly from, for instance, the average Internet usage. In practice, we see that searches with search engines result in a collection of documents (called catch/result/recall) graphically represented as A/(A and B), but the user expects a different result set (relevance/precision) represented as A/(A and D).

It is clear that recall and precision are at odds and often cannot be optimally realized simultaneously.

As a field of study, IR searches for possibilities to improve the ratio between recall and precision (Fig. 7.15).

Figure 7.16 is a graphical representation of the ratio between recall and precision. The challenge within the field of IR is therefore to find methods and techniques in the design of systems that allow end users to find the desired information.

In the universe of search engine systems, the products and functionalities become more complicated and more extensive in order to get closer to the maximum result: supplying only relevant results to each individual user.

Figure 7.17 shows the functionality, grouped by solution type, that must be present to achieve the optimum results. Today's simple search engines are able to produce approximately 60 % of relevant results, and the leading search engines get very close to the upper limit of 100 %.

Lastly, some thesaurus terms can be given higher values than others. The whole process of organizing the source and content largely determines the relevance that can be achieved in presenting the results. Based on the initial experience with

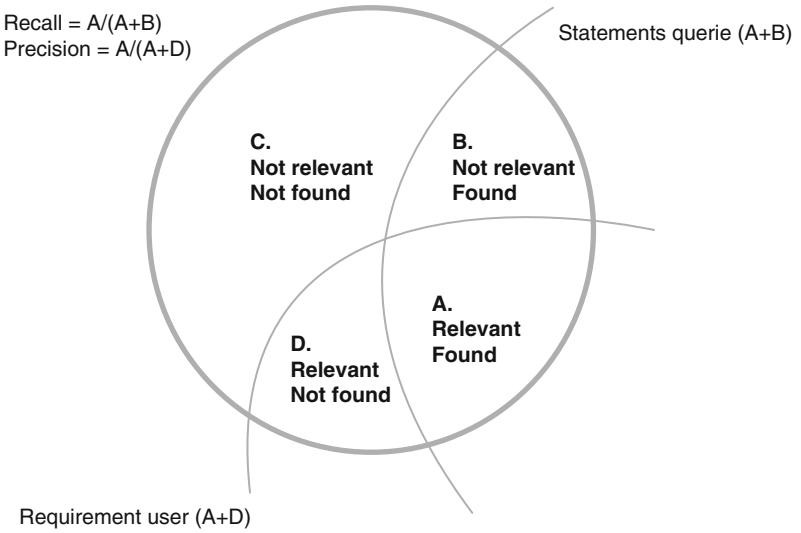


Fig. 7.15 Ratio recall and precision

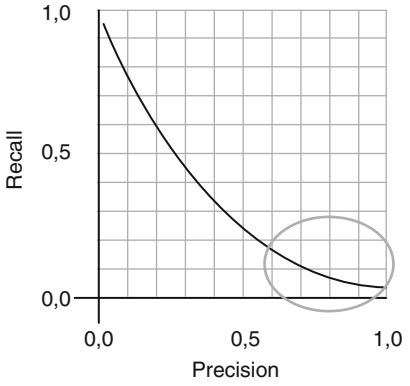


Fig. 7.16 Graph of a typical relationship between recall and precision. Circle: preferred result

scoring the results, together with the relevance and order of presentation of the results by users, the first set of these values is further adapted to a specific user group. This is called query cooking.

The ultimate effort to get as close as possible to the 100% relevance is query cooking. Based on the first results, the previous steps are repeated. Synonym and homonym lists and stop words can be expanded and categories perfected. The whole process is repeated until the end result is satisfactory. This, in combination with the

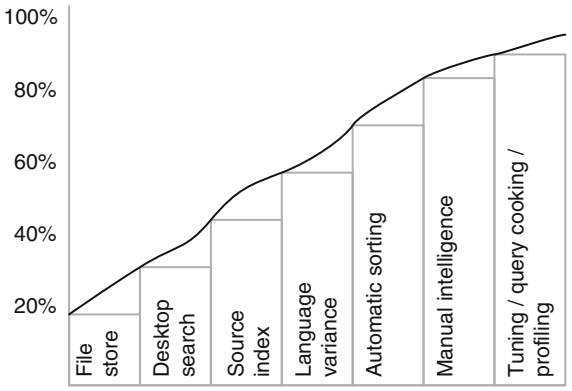


Fig. 7.17 Structure of relevant results sets

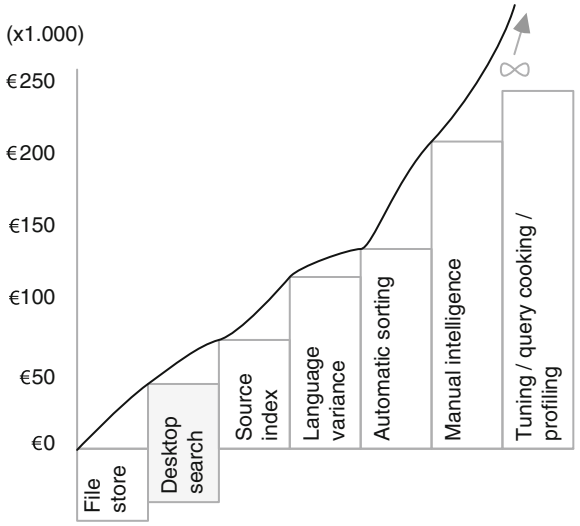


Fig. 7.18 Cost functionality of search engines

fact that most search engines are used by multiple people over multiple sources and very different situations, means that query cooking can only be done in very exceptional cases where information is crucial for the user.

Then where is the optimal balance between investing in search solutions and efficiency improvement? Of course, there is a financial component to the search for the maximum number of relevant search results. This is shown in Fig. 7.18. The figure shows that an organization can invest endlessly in an attempt to squeeze out every last bit of return on search results relevance. Every organization must decide for itself whether the investment is worthwhile. In some cases it is essential to get to the bottom of things in a minimum amount of time. In that case, money hardly plays a role (think of compliance issues within the banking, insurance, and pharmaceutical industries).

7.4.4 Useful Functionality in Search Engines

Search engines come in all shapes and sizes, but they do not all come with the same functionalities. Some handy features that in some cases improve the relevance of results should be mentioned.

- Wildcards (asterisk or *) abbreviate a word on a certain character and thereby shows variants of that word that occur in the index. For example "tele*" shows, for example, television, telephone, and telegraph.
- Stemming entails parsing words to their stem, linking them to the variants, and then showing them in the result. For example: walk, walking, walked but also glass and glasses or cup and cups. Most search engines have a stemming module for the major world languages like English, French, German, and Spanish.

  Other languages are often not available, and in Dutch for example stemming is hard to do because of the irregular verbs and plurals. Compiling a word list with the variants of words that are important for an organization can be helpful (vocabulary list).
- Did you mean? Sometimes you make a typing error or you do not exactly know the right spelling of a word. In those cases, it is useful when the search engine searches the index for words that are very similar to the required word or phrase and gives suggestions for other possible relevant search terms based on this.
- Case-sensitive search means that it is possible to indicate (within the query) what the spelling of a word in the result should be in order to be relevant. This includes the use of capital letters. Words should occur in the results list with the exact same spelling. For example, "Subway" often refers to the restaurant chain that sells submarine (sub) sandwiches and other items, but written without a capital, it refers to a means of transportation. Given this distinction, a relevant result set is likely to be found more quickly.

7.4.5 Presenting Relevant Results

The results of the found information must be offered to the user. Users want to be able to navigate to the information they need or desire to see quickly whether or not the results are relevant. This can be done in several ways. The following ways are recognized.

1. *Sorting*
   First, there is the option to sort the information. This can be done based on a characteristic, often the relevance. It may be easier to sort based on date or origin (or any other characteristic). Of course, these items can also be combined. This way, the most relevant information can be put first in the list, making it accessible to the user as quickly as possible.

Fig. 7.19 Clustering I

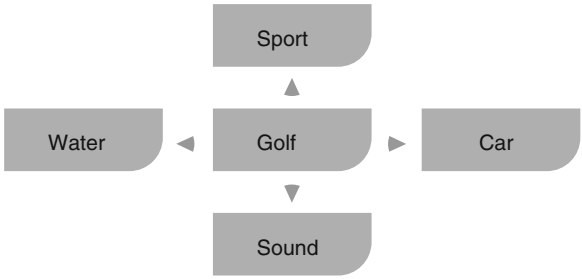
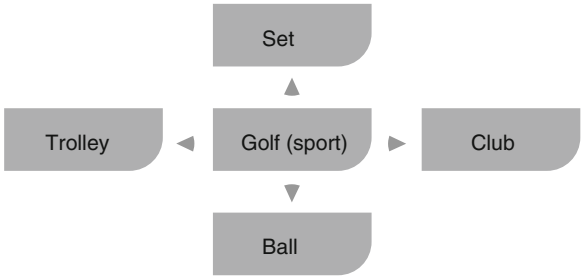


Fig. 7.20 Clustering II



2. *Clustering*
   In addition to sorting, a clustering technique could also be used. Clustering means that a certain search term is placed under an umbrella term. A term may now occur in multiple concepts. This forms the basis for the technique used in file stores in which folders are placed under a main term.

   The user does not specifically search for a term on the lower level (deepest layer) but limits the set of folders in which a document can be found with every choice that is made (drill down).

   Take, for example, the search term ”lub.” Umbrella terms may be “Sport,” “Golf,” “Music,” “Association,” etc. (Fig. 7.19).

   If you choose “Golf,” then the information that has to do with “Golf” and “Club” will become available (Fig. 7.20). Within this set, clusters can be formed like “Wood,” “Set,” “Ball,” “Putter,” etc.

   In this way, the information that is made available is more and more relevant to the end user and can be found quickly. This type of navigation is often displayed graphically with vectors, and the distance between two terms represents the degree of relationship between the terms (examples can be found in products like Aquabrowser and Inxight).

3. *Relevance feedback*
   The quality of the result sets of a query can be improved by adding search terms that have been added by the user (folksonomy). In this connection, it would also be interesting to find out whether users can contribute to quality improvement in any other way. One way is relevance feedback. Users are consciously or

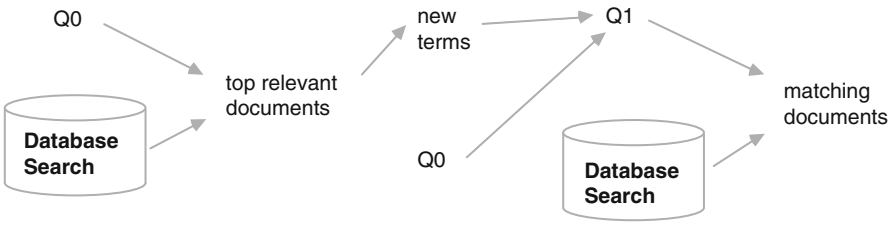


Fig. 7.21 Schematic representation of functioning of relevance feedback (source: http://www.ir.iit.edu/~dagr/cs529/files/handouts/07Feedback.pdf)

unconsciously used to expressing their views about the quality of the result set. Feedback can be described in three ways: explicit, implicit, and blind (also called "pseudo") feedback.

With explicit relevance feedback, users are asked whether, in their eyes, an outcome is considered relevant to the question they posed. Users can also be asked to compile a top ten list of the presented results based on relevance. In this way, more weight is given to relevant documents and new terms from these relevant documents are added to the query to get better search results. Based on relevance feedback, higher recall and precision are obtained (Fig. 7.21).

Implicit relevance feedback looks at user behavior based on the presented results and then interprets that behavior in terms of relevance. When a document is presented as most relevant but not opened or viewed by the user, then eventually this document will no longer be presented as most relevant. Google, among other search engines, uses this principle in its Web search engine.

Blind relevance feedback is based on the idea that the top X documents in a results list contain the most relevant information and selects the most important terms from these documents. These are added to the query to get even better results. The user does not notice and selection is done based on a ranking, not user preferences (in other words, blindly).

7.5 Interpretability

Interpretability of information as result of a search activity means that the information was not only relevant (useful) but also allows the user to do something with that information. The information offers an explanation of ideas of the author or explains the meaning of other facts or conceives the significance of other facts.

The final phase of a search is that the information found as a result is processed by the user or that the person who requested the information is informed. The information serves as the basis for an answer to a question, for a new idea, or for a decision that needs to be made. In all cases, something can be done with the knowledge gained.

This sometimes means that someone is needed to translate the relevant information into a language understood by the searcher or that someone is needed to help explain the results to the searcher.

The trend is to supply information in as clear a form as possible, so that the information can be processed, analyzed, integrated in other applications, or shared with others. Until a few years ago, a reference to the location of information was sufficient as the result of a search. Nowadays, users expect answers to their questions in the form of text, graphs, or a combination of the two. The search engine is supposed to help the user process the results to reduce the time spent on structuring the information in the results set.

An example of this is sentiment analysis. Search engines can give an impression of the way the media write about a certain company, product, or development based on a positive or negative connotation of words stored in the index. Sentiment analysis is often applied to blogs, tweets, messages, and social media like Hyves and Facebook, but when organizations develop 2.0 facilities, sentiment analysis also becomes more important for decision making or adjusting policy.

The best example in this context is Google. Ask the average user of search systems in an organization what his wishes are regarding searches and very often the answer will be: "I would like to be able to search like I can with Google." Users are mainly concerned with the ease with which a query can be made and a result obtained.

What do you do when you need information from the Internet? You open a browser and start Google. You enter a query in the search box, and in less than a second, references to the required information are presented as a list of documents and addresses. The rule of thumb is that the list should be drawn up in such a way that the information you are seeking can be found on the first or second page. The need to go through pages and pages in search of relevant information is avoided. The phenomenon is known as Googlification.

Why don't we use this technique within an organization? Why do we still worry about which drive and folder we used to store information? Do we still think in terms of "knowledge is power"? The introduction of new search capabilities requires a different attitude within an organization. The introduction of Googlification within an organization is called enterprise search. To make enterprise search as successful as Googlification, a thorough preparation is required, on both the business side and the IT side.

A company should know which information it needs and in what situation. Otherwise, people will be looking for a needle in a haystack. Taxonomies enable companies to obtain better results in their searches.

For example, Google would recognize the query "1z1234567890123456" as being the tracking code for a UPS package (though this number is not a valid tracking code). It would provide the user with the link to the UPS site. This technique could also be used in an organization. By creating standard codes for searching, one could search for business objects. For example, a search can be started for the sales figures of a product by entering a product key. Another query about a customer's total order amount starts by entering the customer number. Apart from these

structured data, the search engine gives all documents related to the search argument (customer or product) that are available on an intranet or in some other location in the information environment. The company should also know which people are allowed to view the information. This means that there must be a basis for trust. Only people who are authorized to see certain information actually have access to it. Confidence is essential when it comes to sharing information. Sharing information is necessary for establishing cooperation.

Just as in ECM, queries can be forwarded to colleagues or other interested parties by e-mail, a search profile can be stored so that the right form of presentation is used for each user, and queries can be stored for reuse or shared with others.

It is also possible to subscribe to an RSS feed that gathers new or modified results of a certain query (alerts) and forwards these to subscribers.

The internal and external worlds of organizations are becoming increasingly transparent. The outside world expects to find everything it wants to know about a company within seconds on the Internet. Organizations want to know who they are dealing with and who is tweeting about them, and they base their actions on this information. In searches, we see a nice blend of internal and external sources grouped in dashboards for personal use or role-based use. The blend is arranged for the tasks, decisions, and actions that are requested by people in certain roles. In the near future there will probably no longer be a difference between inside or outside information but only information that needs to be discovered in order to outsmart the competition.

About the Authors

Paul Baan

Paul is a familiar face in the world of enterprise content management and Director at Incentro. This organization is engaged in consultancy and implementation in the fields of enterprise information, business intelligence, content management, and search. Paul is a strong conceptual thinker and is a driving force behind the concept of enterprise information management as the integration of both structured and unstructured information.

Robbert Homburg

Robbert likes to collaborate with a team of coworkers and client employees on projects to enhance the value of information. As a manager he focuses on energy and responsibility. We have to deliver energy all day to add value, and our energy and drive are our most valuable assets. This starts with physical energy by eating sensibly and ends with a mission in life and a clear sense of your added value. Responsibility gives people the space to make decisions and stick to them. Apart from his work as a consultant and manager Robbert spends time on research. His main focus is on the brain and its information-processing power, the last hurdle in information productivity. He thinks we can share information smarter to make better use of our brain's capabilities. Awareness about how information processing works in our brain is a very good first step! Specific knowledge areas: knowledge management, information management, neurosciences, business intelligence, big data, enterprise and Web content management, project management, change management, enterprise2.0, the new world of work, and digitization.

Anja van der Lans

Clients call Anja when they have a problem with information or knowledge in their organization. She consults with them to determine exactly what the problem is, whether she (or a colleague) can solve it, and the best approach to solving it. Depending on the nature of the problem, the assignments Anja gets from clients

P. Baan (ed.), *Enterprise Information Management: When Information Becomes Inspiration*, Management for Professionals 2, DOI 10.1007/978-1-4614-5236-2, © Incentro 2013

vary in size from several days to several months. She is the coauthor of the Dutch version of this book and has written many articles on the subject of search and information retrieval. Specific knowledge areas: business consultancy EIM, change management, workshops, interviews, knowledge management, requirements analysis, business analysis, information analysis, process analyses for information retrieval, search solutions, Web sites, and process-oriented solutions.

John Septer

John specializes in the field of enterprise information management (EIM) and business and ICT alignment as advisor, business analyst, and information architect. To align business and ICT, John uses a varied, organization-focused approach to design or improve information productivity. To create support John focuses on: the involvement of stakeholders, bringing forward recognizable arguments, the removal of uncertainty, and building expertise and skills among the employees involved. Specific knowledge areas: EIM, business and ICT alignment, information management, business analysis, information analysis, enterprise content management.

Peter van Til

Peter is a highly experienced business intelligence consultant. He is the author of several books on business intelligence. What is so special about Peter is that he not only has considerable conceptual knowledge (architecture, modeling, business intelligence competence centers) but also a lot of hands-on experience. Specific knowledge areas: business intelligence, advising, architecture, coaching, information analysis, design, training, business intelligence and data warehouse development.

Bibliography

AIIM (Association for Information and Image Management) (2010) [Online]. www.AIIM.org

Andrews K (1971) The concept of corpoarte strategy. Irwin, Homewood

Ariely D (2009) Predictably irrational, revised and expanded edition: the hidden forces that shape our decisions. HarperCollins, New York

Atkinson R (1999) Project management: cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria. Int J Proj Manage 17(6):337–342

Baan P, van Til P, van der Lans A (2010) Enterprise information management: de fusie tussen business intelligence, content management en enterprise search. VLC, Lulu. ISBN 978-1-4457-2910-7

Bonnati B (2009) Architettura di Enterprise Information Management [Online]

Brynjolfsson E, Hitt L, Kim H (2011) Strength in numbers: how does data-driven decisionmaking affect firm performance? Available at SSRN: http://ssrn.com/abstract=1819486 or http://dx.doi.org/10.2139/ssrn.1819486

Camerer C, Loewenstein G, Prelec D (2005) Neuroeconomics: how neuroscience can inform economics. J Econ Lit. http://www.hss.caltech.edu/~camerer/JELfinal.pdf

Chater N, Lambert K (2001) Big decisions that go wrong – psychologists ask why ignore the research? http://www2.warwick.ac.uk/newsandevents/pressreleases/ne1000000081492/

Choo CW (2006) The knowing organization: how organizations use information to construct meaning, create knowledge, and make decisions. Oxford University Press, New York

Collis D, Montgomery CA (1995) Competing on resources: strategy in the 1990s. Harvard Bus Rev 73(July–August):118–128

Davenport TH (2001) Competing on analytics. Harvard Business School Press, Boston

Davenport T (2005) Thinking for a living: how to get better performances and results from knowledge workers. Harvard Business School Press, Boston

Dietz J (1996) Introductie tot DEMO: van informatietechnologie naar organisatietechnologie. Samsom BedrijfsInformatie, Alphen aan den Rijn

Dijksterhuis A (2007) Het slimme onbewuste. Bert Bakker, Amsterdam

Forrester (2009) The state of workforce technology adoption, US benchmark [Report]. [s.l.]: Forrester, October 2009

Freeman M, Beale P (1992) Measuring project success. Proj Manage J 23(1):8–17

Gartner (2007) Hype cycle for application development (ID nr. G00147982) [Online]. www.gartner.com

Gladwell M, Levitt S (2005) Blink, the power of thinking without thinking. Little, Brown and Company, New York

Hamel G (2007) The future of management. Harvard Business School Press, Boston

Hammer M (1990) Reengineering work: don't automate, obliterate. Harvard Bus Rev 68 (4, July–August):104–112

P. Baan (ed.), *Enterprise Information Management: When Information Becomes Inspiration*, Management for Professionals 2, DOI 10.1007/978-1-4614-5236-2, © Incentro 2013

Han J (2006) Data mining, concepts and techniques. Morgan Kaufmann, San Francisco
Hinssen P (2010) The new normal. Lannoo, Tielt
Inmon WH (2001) The corporate information factory. Wiley, New York
Inmon WH, Nesavich A (2008) Tapping into unstructured data. Prentice Hall, Upper Saddle River
Inmon William H (2009) The elusive virtual datawarehouse, www.b-eye-network.cn/view/9956
Jeston J, Nelis J (2006) Business process management: practical guidelines to successful implementations. Butterworth-Heinemann, Burlington
Kahneman D (2012) Thinking, fast and slow. Farrar, Straus and Giroux, New York
Kearns GS, Lederer AL (2004) The impact of industry contextual factors on IT focus and the use of IT for competitive advantage. Inf Manage J 41:899–919
Kimball R (1997) A dimensional modeling manifesto. www.dbmsmag.com/9708d15.html
Kimball R (1998) Bringing up supermarts. www.dbmsmag.com/9801d14.html
Kirby J (2005) Towards a theory of high-performance. Harvard Bus Rev 83(7, July–August): 30–39
Kooij J van der (2009) Soc Bus Intell [Journal]. Database Magazine. pp 22–25
Kotter J (1995) Leading change. Harvard Business School Press, Boston
Kotter J (1996) Leading change. Harvard Business School Press, Boston
Kurweil R (2006) The singularity is near: when humans transcend biology. The Viking Press, New York
Langseth J (2004) Real-time data warehousing: challenges and solutions. http://dssresources.com/papers/features/langseth/langseth02082004.html
Lans R van der (2009) The flaws of the classic *datawarehouse* architecture [Online]. b-eye-network. April 2009. http://www.b-eye-network.com/view/9960
Larsen KB et al (2011) Factors differentiation concerning information productivity. RSM, Rotterdam
van der Lek H (2006) Sterren en dimensies. Array Publications, Alphen aan den Rijn
Leong L, Jarmoszko AT (2010) Analysing capabilities and enterprise strategy: a value proposition framework. Int J Manage Inf Syst 14(1):53–59
Lim CS, Mohamed MZ (1999) Criteria of project success: an exploratory re examination. Int J Proj Manage 17(4):243–8
Logan D, Bill H (2009) Enterprise information management; a requirement for enterprise-scale business intelligence and performance management initiatives [Report]. [s.l.]: Gartner G00169152, 2 July
Lohr S (2011) When there's no such thing as too much information. The New York Times. http://www.nytimes.com/2011/04/24/business/24unboxed.html?_r=1
Manning CD, Raghavan P, Schütze H (2008) Introduction to information retrieval. Cambridge University Press, New York. pp 1–11
Manyika J, Roberts R, Sprague K (2007) Eight business technology trends to watch, McKinsey Quarterly
Mehrjerdi YZ (2010) Enterprise resource planning: risk and benefit analysis. Bus Strateg Ser 11(5):308–324
Moore G (1965) Cramming more components onto integrated circuits. Electronics Magazine, p 4
Nesbitt J (1990) Megatrends. New York: Harper & Row
Open Text Corporation (2005) ECM methods: what you need to know, vol 3. Open Text Corporation, Ontario
Pascale R (1981) The art of Japanese management. Graduate School of Business, Stanford University
Peslak A (2003) A firm level study of information technology productivity by industry using financial and market based measures. J Inform Techno Impact 3(2):77–90
Peters T, Waterman R (1982) In search of excellence. HarperCollins, New York
Pijpers (2011) Het informatie Paradijs Haystack, Zaltbommel
Pinto JK, Slevin DP (1988) Critical success factors across the project life cycle. Proj Manage J 19(3):67

Porter ME (1985) Competitive advantage: creating and sustaining superior performance. The Free Press, New York
Porter M (1996) What is strategy? Harvard Bus Rev 76(November–December):61–78
Porter M (2008) The five competitive forces that shape strategy. Harvard Buss Rev 54(January): 1–18
Propper I, Litjens B, Weststeijn E (2004a) Lokale regie uit macht of onmacht? http://www.partnersenpropper.com/upload/artikelen/da7d8457995a3a9df7d358cbedd00a17.pdf. Accessed 2 Apr 2012
Pröpper I, Litjens B, Weststeijn E (2004b) Lokale regie uit macht of onmacht? Onderzoek naar de optimalisering van de gemeentelijke regiefunctie. Partners+Propper, Vught
Ranjan J (2008) Business justification with business intelligence. J Inf Knowl Manage Syst 38(4):461–475
Roberts D (2010) Behavior change causes changes in beliefs, not vice versa. http://grist.org/politics/2010-11-23-behavior-change-causes-changes-in-beliefs-not-vice-versa/
Ross M (2009) The 10 essential rules of dimensional modeling. Intelligent Enterprise. www.intelligententerprise.com/showarticle.jhtml?articleID=217700810
Smiley D, Pugh E (2009) Solr 1.4 enterprise search. Server Book Packt Publishing, Birmingham
Strassman P (1999) Information productivity: assessing information management costs of U. S. Corporations. Information Economics Press, New Canaan
Strassman P (2006) 5 steps to improve your information productivity. Baseline Magazine. http://www.baselinemag.com/c/a/Projects-Management/5-Steps-to-Improve-Your-Information-Productivity/
Strassmann PA (2004) Defining and measuring information productivity. The Information Economic Press, version 2.0, January 2004
Strassmann PA (2010) Strassmann, INC. Biography. Accessed 07 Dec 2011 from http://www.strassmann.com/bio.php
Tracey M, Wiersema F (1995) De discipline van marktleiders: Kies uw klanten, verklein uw focus en domineer uw markt. Scriptum Books, Schiedam
Wateridge J (1998) How can IS/IT projects be measured for success. Int J Proj Manage 16(1):59–63
Weggeman M (2000) Kennismanagement: de praktijk. Scriptum, Schiedam
Zimmermann M (1989) The nervous system in the context of information theory. In: Schmidt RF, Thews G (eds) Human physiology. Springer, Berlin, pp 166–173

Index

P. Baan (ed.), *Enterprise Information Management: When Information Becomes Inspiration*, Management for Professionals 2, DOI 10.1007/978-1-4614-5236-2, © Incentro 2013