An Analytical Model of Information Lifecycle Management

Lars Arne Turczyk, Oliver Heckmann, Rainer Berbner, Ralf Steinmetz
TU Darmstadt
KOM Multimedia Communications Lab
Merckstr. 25
64283 Darmstadt, Germany
Phone: +49 69 797 2778
Fax: +49 69 797 3437
e-mail: lars.turczyk@siemens.com

December 8, 2005

Abstract

In this paper we derive an analytical model for Information Lifecycle Management (ILM) systems and use that model to investigate the cost saving potential of ILM systems and to support decision finding. We show analytically that if ILM is employed correctly it can lead to significant storage costs savings in enterprises.

1 Introduction

ILM is based on the idea that in an enterprise there are different information with different values. The different information will be stored on different storage devices. A similar way of thinking is found in the area of operating systems at page swapping scenarios using virtual memory: RAM memory is more expensive than hard disc memory, therefore currently unused memory pages are swapped to the hard disc when memory becomes scarce [1]. The same principle is employed with ILM for storage systems.

ILM manages information according to its value. Valuable information is stored on systems with high Quality of Service (QoS). The value changes over time and therefore migration of information to cheaper storage systems with lower QoS is required. Automated migration makes ILM dynamic. By correctly establishing migration rules, the organization would see little to no delay in information access (keeping frequently accessed information or data requiring instant access regardless of age nearline), but would save significantly by conserving precious disk subsystem space and eliminating disk subsystem purchases to support growth.

In this paper we identify the cost factors and their influences. The paper is structured as follows: In section 2 the general definition of ILM is presented. Then section 3 introduces the analytical model whose implications are investigated in section 4. At the end, in section 5, we show the applicability of the model and how storage decision finding is supported.

2 SNIA's Definition of ILM

The generally accepted definition of ILM worked out by the Storage Networking Industry Association (SNIA) from this presents itself as follows [2]:

Definition 1 (ILM SNIA) *Information Lifecycle Management is compromised of the policies, processes, practices, and tools used to align the business value of information with the most appropriate and cost effective IT infrastructure from the time information is conceived through its final disposition. Information is aligned with business processes through management policies and service levels associated with applications, metadata, information and data.*

This definition forms the basis for an accurate occupation with ILM. Nevertheless this definition is too general to derive a mathematical model for ILM. Therefore we will generate a model to get results for ILM solutions. The model is derived canonically starting with 1-dimensional considerations which will be generalized by multidimensional considerations.

3 Analytical Model for ILM

In this chapter an analytical model based on analyst studies is derived. First we check if the effects of data growth are neutralized by price declining (or performance improvement) of storage components. Of course the well-known Moore's law [3] and similar predictions can be applied for deriving the model. Here analysts' statements are used to take the current market into account to provide more specific predictions.

The following table summarizes all used abbreviations in order to support the reading of the paper.

$g, g(t)$	growth rate of capacity demand
$g_i(t)$	growth rate of capacity demand in hierarchy i
$d, d(t)$	price decline per GB
$d_i(t)$	price decline per GB in hierarchy i
$c(t)$	cost
${}^n c(t)$	n-dimensional cost
$a(t)$	total amount of needed capacity
$a_i(t)$	amount of needed capacity in hierarchy i
$p(t)$	price of needed capacity
$p_i(t)$	price of needed capacity in hierarchy i
$ManC$	managing cost
α_i	hierarchy i's fraction of the total amount of needed capacity
β_i	price factor between hierarchy i and hierarchy 1 with $\beta_1 = 1$ always
${}^n r = \frac{{}^n c(t)}{{}^1 c(t)}$	ratio between n-dim.cost and 1-dim. cost

Table 1: Abbreviations

3.1 Data Growth

The University of California in Berkeley concluded that in 2002 alone around 5 exabytes (10^{18} Bytes) of new "stored information" (paper, film, magnetic, and optical storage media) were produced. This is less than one third of the new information that is communicated through electronic information flows (telephone, radio and TV, Internet) which is about 17,7 exabytes [4].
Even more amounts of data will be produced over the next few years, several analysts report. They all speak of steadily growing capacity demand. The compound annual growth rate (CAGR) varies between 60% and 100%. Metagroup identified a growth rate of 60% [5]. In 2001 IDC estimated a CAGR of 76% over the years 2000-2004 [6]. A report created in 2003 by Horison speaks of a CAGR for data demands over the next few years of 60%-70% [7]. Although the exact value is not known the tendency is obvious. Effects of the increasing demand are seen in the business reports of storage supplier companies. In 2005 IDC detected a continued acceleration of the annual growth rate leading also to revenue growth at EMC, Dell and Network Appliances [8].

To conclude we assume the capacity demand grows by a factor $g \in [0.6; 1.0]$ per year.
This factor represents the overall demand and includes any demand reducing effects like deleting.

3.2 Price Decline of Hardware

As the tendency for demand is growing the tendency for storage prices is declining. Again analysts give a range of prognoses. Between 1998 and 2001 McKinsey determined for the price per-gigabyte (GB) a CAGR of -36% [6]. IDC took a look at the prices per GB between 2001 and 2003. In 2003 per-gigabyte external storage prices fell -33%, while in 2002 and 2001, they fell down -40% and -43% respectively. So the CAGR between 2001 and 2003 is -36% [8].

To conclude we assume the prices per GB decline realistically by a factor $d \in [-0.33; -0.36]$ per year. This factor only influences the hardware cost.

3.3 Managing Cost

The cost situation changes when considering the total cost of ownership (TCO). Both Gartner and IDC reported that an enterprise spends an average of \$3 managing storage for every \$1 spent on hardware [5]. Additionally Gartner Group speaks of \$3,5 being spent for managing each \$1 spent for storage hardware [9]. It is obvious that the effect of growing demand is not neutralized by the effect of price decline. This relation of approximately 3:1 between managing cost (ManC) and hardware cost has to be considered in the cost model, too. To derive the model step by step we first define hardware cost and then extend the model.

3.4 1-dimensional Cost Model

Hardware cost is a function of needed capacity and the price per GB. To consider the cost development over a period of time the integral of the function is calculated.

Definition 2 (Hardware Cost) *Let a(t) be the amount of needed storage capacity at time t, p(t) the price for storage capacity at time t and c(t) the cost of the storage realization at time t.*
Then c(t)=a(t)· p(t) and the storage cost over a period of time from t_0 *to t is:*

$$\int_t^{t_0} c(t)dt = \int_t^{t_0} a(t) \cdot p(t)dt$$

The following conclusion specifies the composition of $a(t)$ and $g(t)$.

Conclusion 1 *When g(t) is the function of capacity growth and d(t) is the function of price decline the cost is:*
$\int_t^{t_0} c(t)dt = \int_t^{t_0} a(t) \cdot p(t)dt$
$= \int_t^{t_0} a(t_0)(1 + g(t))^t \cdot p(t_0)(1 + d(t))^t dt$
with
$a(t) = a(t_0) \cdot (1 + g(t))^t \, [Bytes]$ *and*
$p(t) = p(t_0) \cdot (1 + d(t))^t [\frac{EUR}{Byte}]$

Now we add the administration cost mentioned already and get the TCO:

Definition 3 (TCO) *Let ManC be the managing cost needed to be spent when employing purchased hardware. Then the total cost of ownership is:*

$$c(t) = \int_t^{t_0} a(t_0)(1+g(t))^t \cdot p(t_0)(1+d(t))^t dt + ManC$$

ILM categorizes the demand and moves information items from higher hierarchies to lower ones in order to reduce the demand for new hardware in the highest hierarchy. ILM takes into account that in enterprises a lot of unused data is stored on high performance storage devices [10, 11, 12]. ILM assumes that the storage environment employs different hierarchies. Hence the cost model for ILM has to be extended to reflect the n-dimensional characteristic of an ILM solution.

3.5 n-dimensional Cost Model

When information are migrated between different hierarchies the cost per each hierarchy has to be considered.

Definition 4 (Multidimensional Cost) *The TCO for an n-dimensional ILM solution is:*

$$^n c(t) := \sum_{i=1}^{n} \int_{t_0}^{t} a_i(t) \cdot p_i(t)dt + ManC$$

with
$a_i(t) = (1 + g_i(t))^t$
the amount of needed storage capacity in hierarchy level i at time t,
$p_i(t) = (1 + d_i(t))^t$
the price for storage capacity in hierarchy level i at time t.

The SNIA End User Council (EUC) Top Ten Pain Points survey showed that end-users ranked their pain points in the following order: Costs (price and TCO) have the highest priority before "the challenge of managing growth and meeting capacity needs" [13].
This shows that cost is the driving factor for decisions concerning storage. Therefore the cost effects of ILM have to be investigated.
In chapters 3.1, 3.2 and 3.3 we have shown that there are useful ways to attach value to each single cost factor. In the next chapter we focus on the hardware cost and show that there are positive effects when employing ILM.

4 Analysis of ILM systems

In ILM the storage hierarchies have different costs. Assuming hierarchy 1 is more expensive than hierarchy 2, etc., then there is a $\beta_i < 1$ with
$p_i(t) = \beta_i \cdot p_{i-1}(t)$ for all $i \leq n$.
Further the amount of needed storage capacity $a_i(t)$ for hierarchy i is a fraction of a(t):
$a_i(t) = \alpha_i \cdot a(t)$ with $\alpha_i < 1$ and
$\sum_{i=1}^{n} \alpha_i = 1$

β_i defines the price relation to the highest and most expensive hierarchy. α_i defines the portion of the overall volume stored on hierarchy i.

We will show that α_i and β_i are effective parameters for deciding on an ILM solution. Of course $a(t)$ and $p(t)$ also have effects. But $a(t)$ is determined by the starting situation of the relevant enterprise and develops according to the day after day business. Neither the starting value is changeable nor is it the intention of ILM to influence the business. Furthermore $p(t)$ is influenced by the global markets and not changeable.
Thus α_i and β_i are the controllable parameters. β_i is determined by the choice of employed technology. Choosing an appropriate technology is one task of ILM concept. α_i is determined by specific requirements of the related enterprise, ie. after conducting an assessment the different QoS requirements are determined and the related volumes α_i are assigned.

In the next section we define some assumptions and derive results which are directly connected to the α_i's and β_i's.

4.1 Assumptions

There are three assumptions to be made:
1. $d_i(t)$ are the same for all hierarchies.
2. $g_i(t)$ are the same for all hierarchies.
3. $d(t)$ and $g(t)$ are constant.

As shown in chapters 3.1 and 3.2 these assumptions are allowed and analysts give advice for their general conditions.

The assumptions simplify the model. With the simplified model the effects of price per capacity and amount of capacity needed are derived. The effects are still there when the assumptions are neglected. Therefore the assumptions are not necessarily needed to apply the results.

Now we look at the multidimensional hardware cost. First we look at the ratio between 1-dimensional cost and 2-dimensional cost and afterwards we increase the dimensions.

4.2 2-dimensional Cost

In order to get the ratio between 1-dimensional and 2-dimensional cost we divide ${}^2c(t)$ by ${}^1c(t)$:

$$\frac{{}^2c(t)}{{}^1c(t)} =: {}^2r(t)$$

$$= \frac{\int_{t_0}^{t} a_1(t) \cdot p_1(t)dt + \int_{t_0}^{t} a_2(t) \cdot p_2(t)dt}{{}^1c(t)}$$

$$= \frac{\int_{t_0}^{t} \alpha_1 a_1(t) \cdot \beta_1 p_1(t)dt + \int_{t_0}^{t} \alpha_2 a_1(t) \cdot \beta_2 p_1(t)dt}{{}^1c(t)}$$

$$= \frac{\alpha_1\beta_1 \int_{t_0}^{t} a_1(t) \cdot p_1(t)dt + \alpha_2\beta_2 \int_{t_0}^{t} a_2(t) \cdot p_2(t)dt}{\int_{t_0}^{t} a_1(t) \cdot p_1(t)dt}$$

$$\overset{\sum \alpha_i=1, \beta_1=1, \beta_{i\neq 1}<1}{\Rightarrow} \quad {}^2r < 1 = \frac{{}^1c(t)}{{}^1c(t)} = {}^1r$$

It is shown that ILM influences hardware cost. A cost reduction is achieved by categorizing the storage demand into hierarchies. The main characteristics of different hierarchies are different costs resulting from different QoS. QoS summarizes in particular characteristics like security, backup frequency or access speed [14]. Thus ILM reduces the demand for new hardware in the most expensive information class. This leads to a reduction in the hardware cost.
The effect shown with n=2 can be generalized for $n > 2$. Each new hierarchy i, $i > 2$, reduces the cost if there is an amount of data $a_i(t) = \alpha_i \cdot a(t)$, to be categorized in the hierarchy.

It is obvious that theoretically the number of information classes can be high. In reality the storage environment becomes quite complex with many hierarchies and managing cost increase. Thus the TCO for each new hierarchy in the case $n > 2$ has to be considered.

4.3 Effects of α and β on ${}^2r$

As shown before, ILM has positive effects on cost. To analyze how influential the employment of ILM is α and β have to be examined. In case of ${}^2r$, α_2 and β_2 have to be examined[1].

$$ {}^nr = \sum_{i=1}^{n} \alpha_i \beta_i \cdot \int_t^{t_0} a_1(t) \cdot p_1(t) dt $$

To analyse the effects of α_i and β_i on ${}^nr$ the sum $\sum_{i=1}^{n} \alpha_i \beta_i$ is investigated.

Apparently there are degenerated values like $\alpha_2 = 0$, $\alpha_2 = 1$, $\beta_2 = 1$ or $\beta_2 = 0$. These values do not represent a 2-dimensional ILM solution. Therefore they are excluded.

For non-degenerated α_2 and β_2 the following graphic shows the effects on ${}^2r$:

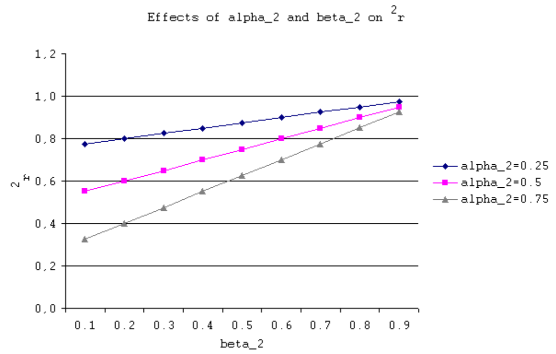

Figure 1: Effects of α_2 and β_2 on ${}^2r$

It is shown that α_2 and β_2 each have impact on ${}^2r$. If β_2 is close to 1 (=β_1) the effect is almost 0, in fact irrespective of α_2.
The effects on ${}^2r$ depend on the distance between α_1 and α_2 and the distance between β_1 and β_2.
This means that for ILM scenarios the number of hierarchies is limited by the effects of cost. Therefore not every hierarchy is worth being established.
In the next sub-chapter we show that the adding of a hierarchy can lead to neutral and even negative effects.

4.4 Effects of α_i and β_i on ${}^nr$

In the multidimensional case with n>2 similar effects like those for ${}^2r$ can be expected. As seen for n=2 the effects depend on the distance between the α_i and the difference between the β_i. Since $\sum_{i=1}^{n} \alpha_i = 1$ there is a constraint influencing the distances between α_i.

We show two cases, where the effects of α_i and β_i are considered. In each case four different dimensions (n=2, n=3, n=4 and n=5) are investigated. Case 1 reflects a situation where the portion of hierarchy 1 is fix 50%. Although there is potential for all hierarchies n and the prices are well arranged, the influences of α_i and β_i cancel out each other.

Case 1:

n=2	$\alpha_1 = 1/2$	$\alpha_2 = 1/2$			
n=3	$\alpha_1 = 1/2$	$\alpha_2 = 1/4$	$\alpha_3 = 1/4$		
n=4	$\alpha_1 = 1/2$	$\alpha_2 = 1/6$	$\alpha_3 = 1/6$	$\alpha_4 = 1/6$	
n=5	$\alpha_1 = 1/2$	$\alpha_2 = 1/8$	$\alpha_3 = 1/8$	$\alpha_4 = 1/8$	$\alpha_5 = 1/8$

n=2	$\beta_1 = 1$	$\boldsymbol{\alpha_2 = 1/2}$			
n=3	$\beta_1 = 1$	$\beta_2 = 2/3$	$\boldsymbol{\beta_3 = 1/3}$		
n=4	$\beta_1 = 1$	$\beta_2 = 3/4$	$\beta_3 = 2/4$	$\boldsymbol{\beta_4 = 1/4}$	
n=5	$\beta_1 = 1$	$\beta_2 = 4/5$	$\beta_3 = 3/5$	$\beta_4 = 2/5$	$\boldsymbol{\beta_5 = 1/5}$

Now the relating ${}^nr$ for case 1 are considered.
${}^nr = \sum_{i=1}^{n} \alpha_i \beta_i \int_t^{t_0} a_1(t) \cdot p_1(t) dt$
Calculation of the term $\sum_{i=1}^{n} (\alpha_i \cdot \beta_i)$:

n=2:
$\sum_{i=1}^{2} (\alpha_i \cdot \beta_i) = 1/2 \cdot 1 + 1/2 \cdot 1/2 = 0.75$

n=3:
$\sum_{i=1}^{3} (\alpha_i \cdot \beta_i) = 1/2 \cdot 1 + 1/4 \cdot 2/3 + 1/4 \cdot 1/3 = 0.75$

n=4:
$\sum_{i=1}^{4} (\alpha_i \cdot \beta_i) = 1/2 \cdot 1 + 1/6 \cdot 3/4 + 1/6 \cdot 2/4 + 1/6 \cdot 1/4 = 0.75$

n=5:
$\sum_{i=1}^{5} (\alpha_i \cdot \beta_i) = 1/2 \cdot 1 + 1/8 \cdot 4/5 + 1/8 \cdot 3/5 + 1/8 \cdot 2/5 + 1/8 \cdot 1/5 = 0.75$

Although the added hierarchy is cheaper than all existing hierarchies, the positive effect of smaller β_n is compensated by smaller α_i.
The next case shows that even when the amount of stored data in hierarchy 1 is reduced by adding a new hierarchy, the effect becomes more and more marginal.

Case 2:

n=2	$\alpha_1 = 1/2$	$\alpha_2 = 1/2$			
n=3	$\alpha_1 = 1/3$	$\alpha_2 = 1/3$	$\alpha_3 = 1/3$		
n=4	$\alpha_1 = 1/4$	$\alpha_2 = 1/4$	$\alpha_3 = 1/4$	$\alpha_4 = 1/4$	
n=5	$\alpha_1 = 1/5$	$\alpha_2 = 1/5$	$\alpha_3 = 1/5$	$\alpha_4 = 1/5$	$\alpha_5 = 1/5$

n=2	$\beta_1 = 1$	$\boldsymbol{\beta_2 = 1/2}$			
n=3	$\beta_1 = 1$	$\beta_2 = 2/3$	$\boldsymbol{\beta_3 = 1/3}$		
n=4	$\beta_1 = 1$	$\beta_2 = 3/4$	$\beta_3 = 2/4$	$\boldsymbol{\beta_4 = 1/4}$	
n=5	$\beta_1 = 1$	$\beta_2 = 4/5$	$\beta_3 = 3/5$	$\beta_4 = 2/5$	$\boldsymbol{\beta_5 = 1/5}$

Now the relating ${}^nr$ for case 2 are considered.
${}^nr = \sum_{i=1}^{n} \alpha_i \beta_i \int_t^{t_0} a_1(t) \cdot p_1(t) dt$

[1] Due to determination of α_2, α_1 is determined, too. $\beta_1 = 1$ always.

Calculation of the term $\sum_{i=1}^{n}(\alpha_i \cdot \beta_i)$:

n=2:
$\sum_{i=1}^{2}(\alpha_i \cdot \beta_i) = 1/2 \cdot 1 + 1/2 \cdot 1/2 = 0.75$

n=3:
$\sum_{i=1}^{3}(\alpha_i \cdot \beta_i) = 1/3 \cdot 1 + 1/3 \cdot 2/3 + 1/3 \cdot 1/3 = 0.667$

n=4:
$$\sum_{i=1}^{4}(\alpha_i \cdot \beta_i) = 1/4 \cdot 1 + 1/4 \cdot 3/4 + 1/4 \cdot 2/4 + 1/4 \cdot 1/4 = 0.625$$

n=5:
$$\sum_{i=1}^{5}(\alpha_i \cdot \beta_i) = 1/5 \cdot 1 + 1/5 \cdot 4/5 + 1/5 \cdot 3/5 + 1/5 \cdot 2/5 + 1/5 \cdot 1/5 = 0.6$$

In case 2 each adding of a new information class creates a positive effect, of course. But it is shown that the advantage of adding a hierarchy becomes more and more marginal (0.75 vs. 0.667 vs. 0.625 vs. 0.6).

Summary Cases 1 and 2:
It is shown that α_i and β_i are the only individually adjustable parameters in an ILM scenario. In order to get the highest gain realizing ILM scenarios they have to be determined sensitively.
Now, for a specific case we show, how decisions for ILM can be made.

5 Application of the multidimensional cost model

In chapter 4 we showed that α_i and β_i are effective factors for cost calculations and therefore for purchase decisions. The question is "How to determine these factors?". For the price factors β_i the answer is given by technology. The different storage technologies have different prices. The price relations are given by market analysts. The price difference between enterprise disk (FC, SCSI, FICON, ESCON[2]) and midrange disk (SCSI, FC) is 2.5:1. The price difference between midrange disk and low cost disk (S-ATA[3]) is 3:1. The price difference between low cost disk and automated tape is 6:1 [7].

These are quite useful results and complete the estimations given in chapters 3.1, 3.2, 3.3. What is missing are the potential factors α_i. How many gigabytes are needed in the different hierarchies? Is there enough potential for ILM? To get these figures we conducted a case study in spring 2005. In the case study we investigated a database of a german DAX-30 company. The access patterns of 150,000 files were investigated. A random sample of 1,000 files was taken and all accesses were logged. The logging of all accesses since creation of each file provided the following results [12]:
There were more than 150,000 files on the system and 89 percent of them were not accessed 90 days after their creation.

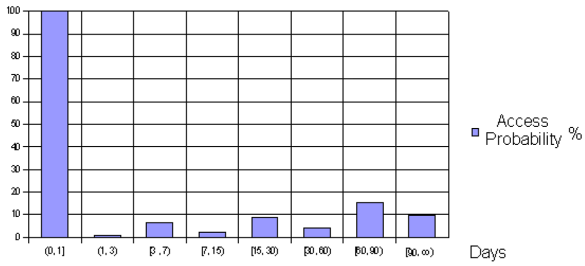

Figure 2: Access Probabilities

Therefore in this case the potential factors are: $\alpha_1 = 0.11$ and $\sum_{i \neq 1}^{n} \alpha_i = 0.89$.
Looking the results of the case study we consider necessary to create an ILM concept, because 89 percent of the files could be migrated. The question is "How many hierarchies?". Since currently there is one hierarchy only, the answer is 2 or 3. It is not advisable to create too much hierarchies. As we showed in chapter 4 the effects become more and more marginal. If the distinction in two hierarchies is enough, 2-dimensional ILM should be applied.
Assuming there are requirements for more than two hierarchies the potential of 0.89 has to be distributed over hierarchies 2 and 3. Let be $\alpha_2 = 0.3$ and $\alpha_3 = 0.59$ ($\alpha_1 = 0.11$). Then a 3 dimensional-ILM solution of a 1.3 TeraByte database would, for example, look like:

Total Volume: 1.3 TeraBytes
Storage hierarchy 1:
Enterprise Disk (FC), required capacity: 143 GB
Storage hierarchy 2:
Low Cost Disk (S-ATA),
required capacity: 390 GB
Storage hierarchy 3:

[2]Fibre Channel, Small Computer System Interface, Fiber Connectivity, Enterprise Systems Connection
[3]Serial Advanced Technology Attachment

Automated Tape, required capacity: 767 GB

The price relation is:
FC : S-ATA : Automated Tape =
$1 : 1/7.5 : 1/45 = \beta_1 : \beta_2 : \beta_3$

By determination of these key figures the ILM cost effects are characterised sufficiently, and IT-managers are supported to find an informed cost decision.

6 Conclusion and Outlook

In this paper we presented an analytical model for Information Lifecycle Management (ILM). Based on cost considerations we have shown that decisions about ILM solutions depend on enterprise external and enterprise internal factors. External factors like market prices for hardware are the same for all ILM solutions. The internal factors can be influenced individually by the enterprise and need to be taken into account separately. These are, in particular, the α_i and β_i. In detail the conclusions are:

1. (n-dimensional) ILM has proven positive effects on cost.
2. The change from no ILM to 2-dimensional ILM has the biggest gain.
3. The change 2-dimensional ILM to n-dimensional ILM does not guarantee gains.
4. For each new ILM-dimension the potential α_i and the price factor β_i have to be considered.

If we sum up the detailed conclusions, ILM systems can offer significant cost savings for enterprises if they are used correctly.
Our next step will be the simulation of actual ILM systems in order to get reliable statements for migrating information. Furthermore the case study will be extended to derive distribution functions for file accesses.

References

[1] A. S. Tanenbaum, *Modern Operating Systems*, Prentice Hall, 2nd Edition, February 2001

[2] M. Peterson, *ILM Definition and Scope - An ILM Framework*, SNIA Data Management Forum, Version 2.3, July 2004

[3] G.E. Moore, *Cramming more components onto integrated circuits*, Electronics, volume 38, number 8, 1965.

[4] A. Lyman, B. Varian, *How Much Information? 2003*, University of California, October 2003

[5] D. Fletcher, K. Elliott, *Benchmark and Trend Analysis*, Division of Information Technology Services, State of Utah, October 2003

[6] T. Kraemer, J. Berlino, *The Storage Report - The customer Perspectives & Industry Evolution*, McKinsey & Company, June 2001

[7] F. Moore, *Storage - New Game New Rules*, Horison Information Strategies, 2003

[8] H. Nguyen, *IDC's Worldwide Disk Storage Systems Quarterly Tracker*, March 2005

[9] R. Paquet, *Why You Need a Storage Department*, Gartner Research, June 2004

[10] M. Satyanarayanan, *A Study of File sizes and Functional Lifetimes*, Proceedings of the 8th Symposium on Operating Systems Principles, Association of Computing Machinery, 1981

[11] S. Strange, *Analysis of Long-term Unix File Access Patterns for Application to Automatic File Migration Strategies*, University of Berkeley, 1992

[12] R. Gostner, L. Turczyk, R. Berbner, O. Heckmann, R. Steinmetz, *Analyse von Datei-Zugriffen zur Potentialermittlung fuer Information Lifecycle Management*, TU Darmstadt KOM Technical Report 01/2005

[13] R. Bauer, *Top Ten Pain Points Survey Results Report*, SNIA End User Council, May 2004

[14] J. B. Schmitt, *Heterogeneous Network Quality of Service Systems*, Kluwer Academic Publishers 2001