



7 Data Lifecycle Management

7.1 Data Lifecycle Management = Information Lifecycle Management

Der strukturierte Umgang mit Daten über ihre gesamte Lebenszeit hinweg wird heute von den existierenden IT-Systemen nicht durchgehend unterstützt. Dennoch ist die Notwendigkeit gegeben, werden doch Unternehmensdaten immer mehr zur Grundlage eines Geschäfts.

Data Lifecycle Management / Information Lifecycle Management umfasst:

- *Im weiteren Sinn:* Die Bereitstellung von Policies (Taktiken / Richtlinien), Processes (Abläufen), Practises (Methoden), Services (Dienste) und Tools zum strukturierten Umgang mit Daten über ihre ganze Lebenszeit hinweg.
- *Im engeren Sinn:* ein Storage Management-Konzept, welches Informationsobjekte während der gesamten Lebenszeit aktiv verwaltet.

Die aktive Verwaltung der Daten über ihre ganze Lebenszeit hinweg umfasst die Umsetzung eines Regelwerkes bestehend aus Methoden, Prozessen und Technologie. Dabei stehen sich die Vorgaben aus den Geschäfts-Prozessen und die Notwendigkeit zur Optimierung der Speicherkosten gegenüber. Vorgaben sind beispielsweise der Wert eines Informationsobjektes, gesetzliche Rahmenbedingungen, Aufbewahrungszeiten, Zugriffsrechte und SLA's (Service Level Agreements). Informationsobjekte sind Logische Units (LUN), Volumes, Dateien, Verzeichnisse, Datenbanken, Records, Emails, etc. Data Lifecycle Management prüft nun Ereignisgesteuert oder in fixen Intervallen, ob sich die Bewertung eines Informationsobjektes geändert hat und stösst gegebenenfalls Aktionen an, um die Objekte optimal zu verwalten. Solche Aktionen sind ein Verschieben des Informationsobjektes innerhalb der Speicherhierarchie, ein Archivieren oder ein Löschen des Objektes.

7.1.1 Definitionen

Die Definition von Sun Microsystems [Wood 2005]:

Data Lifecycle Management (DLM) ist ein Verfahren zur Verwaltung von betrieblichen Informationsobjekten basierend auf ihrem Wertbeitrag während der gesamten Lebenszeit dieser Objekte.

Die Firma EMC, ein Hersteller von Speichersystemen, umfasst ILM genereller [Reiner et al. 2004]:

Information Lifecycle Management (ILM) ist eine business-orientierte Strategie zum proaktiven Management von Informationen anhand ihres Wertes.

Die wahrscheinlich älteste Definition stammt von der FEDSIM (Federal Systems Integration and Management Center) aus dem Jahr 1993:

Information Lifecycle Management: The integration of a variety of activities designed to manage information and information resources throughout their life cycle. Activities include planning, budgeting, organizing, directing, training, promoting, and controlling the information and information resources throughout the process of collecting, processing, transmitting, disseminating, and disposing of information.

Die umfassende Definition der SNIA (Storage Networking Industry Association) [SNIA 2004]:

Information Lifecycle Management umfasst die Richtlinien, Abläufe, Methoden, Dienste und Instrumente, die verwendet werden, um die angemessenste und kosteneffektivste Infrastruktur bereitzustellen, die Informationen entsprechend ihrem Unternehmenswert über den ganzen Lebenszyklus hinweg verwalten. Die Anforderungen eines Unternehmens werden als Management Richtlinien (Policies) und Service Levels, die sich auf Applikationen, Metadaten und Daten beziehen, abgebildet.

Die Abgrenzung zwischen DLM und ILM ist in der Fachwelt heute umstritten. Bestimmte Fachleute gehen davon aus, dass DLM eine Untermenge von ILM darstellt. Die zusätzlichen ILM Funktionen wären dann Data Mining, Prozessverarbeitung und Entdeckung von Geschäftstrends. Data Mining und Entdeckung von Geschäftstrends sind jedoch klassische Business Intelligence Funktionen, während Prozessverarbeitung ein Teilbereich des Business Process Engineering darstellt.

7.1.2 Der Lifecycle von Daten

Der Lebenszyklus von Daten kann in verschiedene Abschnitte eingeteilt werden, die als Grundlage für die Art und Weise, wie mit den Daten entsprechend ihrer Verwendung umgegangen wird, dient. Der klassische Data Lifecycle umfasst 6 Zustände, wie in **Abbildung 7.1** dargestellt.



Abbildung 7.1 Der Lifecycle von Daten

- *Create*: Daten werden erzeugt. Die jährliche Datenzunahme war gemäss der Studie der University of California in der letzten Dekade im Exabyte-Bereich [Lyman et al. 2003]. Heute liegt er bereits im Zettabyte-Bereich [Gantz et Al. 2013]
- *Transport*: Es werden 3.5-mal mehr Daten transportiert, als gespeichert werden.
- *Modify*: Lediglich 10 % aller Daten werden tatsächlich verändert.
- *Use / Store*: Ab den 30 Tag nach der Speicherung wird lediglich noch auf 20 % aller Daten zugegriffen.
- *Archive*: Die Archivierung von Daten hängt vor allem von den regulatorischen Vorgaben eines Unternehmens ab.
- *Shred*: Die gezielte und gesteuerte Vernichtung von Daten.

7.1.3 Gründe für ILM/DLM

Die wichtigsten Gründe für die Etablierung von Data Lifecycle Management sind:

- *Datenzuwachs*: Der jährliche Datenzuwachs betrifft sämtliche Branchen. Datenbanken wachsen im Schnitt zwischen 100% und 125% pro Jahr, während statische Daten um 60% wachsen. Digital Content (Filme, Musik, etc..) wächst um 50% pro Jahr. [DMG 2004].
- *Kosten für die Speicherverwaltung*: Während die Kosten für Hardware sinken, steigen die Kosten zur Verwaltung von Storage jährlich zwischen 25% und 30%.
- *Zugriffs-Anforderungen*: Die Anforderungen von Benützern steigen. Zugriffszeiten und Zugriffs-Sicherheit müssen definiert und garantiert werden können.
- *Komplexität*: Die Systemlandschaften der IT werden nicht einfacher. Die Einführung neuer Technologien, wie Beispielsweise iSCSI, SDS, etc. erhöhen die Anforderungen an die IT.
- *Regulatorien*: Neue Regelungen wie SOX (Sarbanes-Oxley Act), MIFID (Markets in Financial Instruments Directive – Richtlinie der EU zur Harmonisierung der Finanzmärkte), MIFID II oder BASEL II (Eigenkapitalanforderungen) und Basel III (Liquiditätsanforderungen) erfordern die spezifische Behandlung bestimmter Daten.

7.1.4 Datenzuwachs

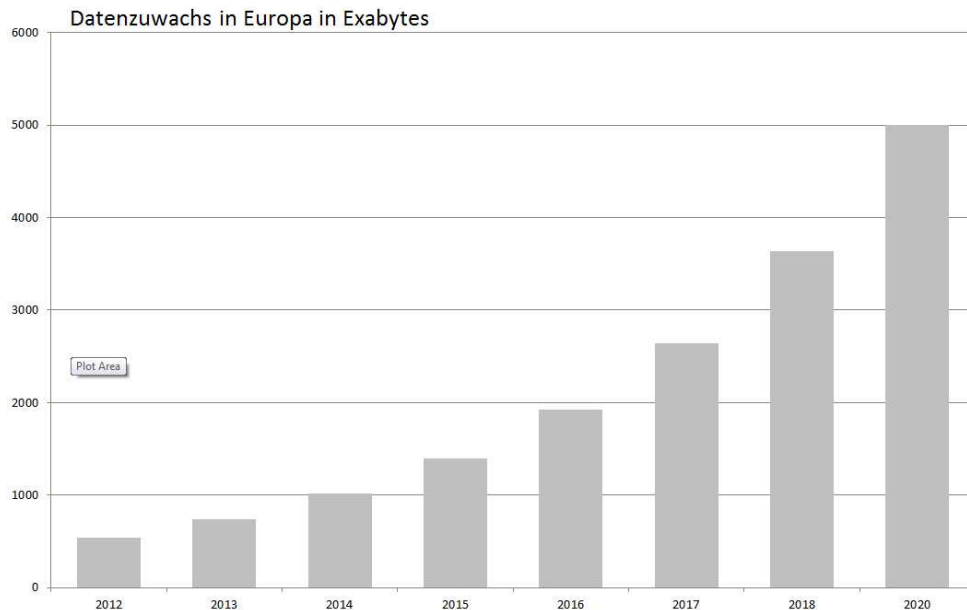


Abbildung 7.2 Datenzuwachs in Europa

Die Jährliche Zunahme der weltweit vorhandenen Informationsmenge wird von Analysten auf 30-40 % geschätzt. IDC geht sogar in der Studie „Digital Universe“ für die Firma EMC von einer Verdoppelung des Datenbestandes alle zwei Jahre in Europa aus. Im Jahr 2012 wurden in Europa etwas mehr als 500 Exabytes Daten produziert. Im Jahr 2020 werden es in Europa 5 Zettabytes sein – weltweit ca. 40 Zettabytes. Bemerkenswert ist die Tatsache, dass die Schätzungen früheren Studien bezüglich des Wachstums der Datenrate nachträglich immer wieder nach oben korrigiert werden mussten [Gantz et Al. 2007].

Informationen werden zum grössten Teil – 68% gemäss IDC - durch die Konsumenten erzeugt, Unternehmen haben jedoch Verpflichtungen und Verantwortung für ca. 80% aller Informationen, die im digitalen Universum erzeugt worden sind.

Eine Auswahl von Indikatoren und Faktoren für diese Zugangsrate sind:

- Der Anteil der durch Maschinen – Industriemaschinen, Fahrzeuge, Medizinische Geräte, Sicherheitsgeräte, Spielzeuge und andere - generierten Informationen nimmt stetig zu (von 11% im Jahr 2005 bis zu 42% im Jahr 2020).
- Das Internet der Dinge (IoT) – die Anzahl der Geräte, die über das Internet miteinander verbunden sind – und damit die Anzahl der Möglichkeiten, Informationen zu generieren steigt überproportional zur Weltbevölkerung. Es wird bis ins Jahr 2020 7 Geräte pro Erdbewohner geben, also ca. 50 Milliarden Geräte [Evans 2011].

- Der Globale Datenverkehr über das Netz wird in den nächsten Jahren pro Jahr um ca. 30 % wachsen [Cisco 2012].

7.1.5 Ausnahme: WWW Zuwachs

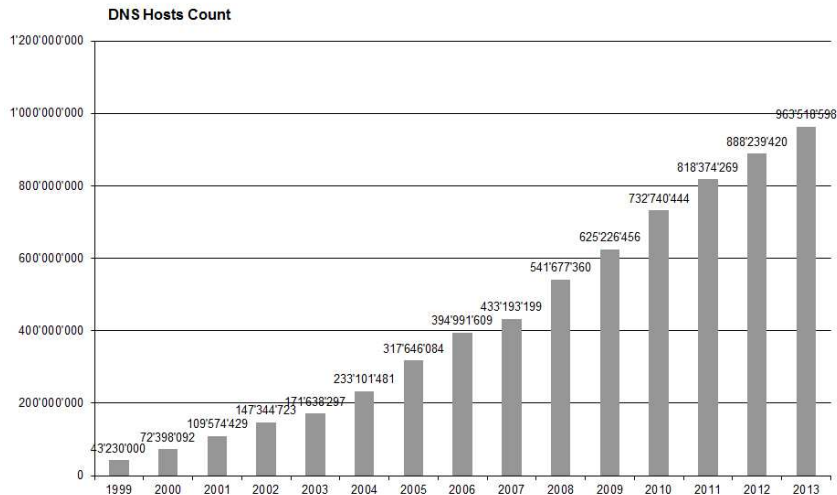


Abbildung 7.3 Zunahme der DNS Einträge (logarithmische Skalierung)

Die Annahme der Zunahme des Datentransports um 30% kann jedoch nicht auf das WWW übertragen werden, da alleine die Zunahme der registrierten DNS Einträge gemäss Internet System Consortium (www.isc.org) im Zeitraum von 1999 bis 2013 zwischen 67% und 8% jährlich beträgt (Abbildung 7.3).

Dieser Effekt multipliziert sich mit der Zunahme der gespeicherten Daten, da zu erwarten ist, dass auch der Umfang einer einzelnen Web Site jährlich um 30% steigt, auch wenn die Anzahl der aktiven Sites gemäss Netcraft (www.netcraft.com) nicht so stark wie die Anzahl DNS Einträge ansteigt.

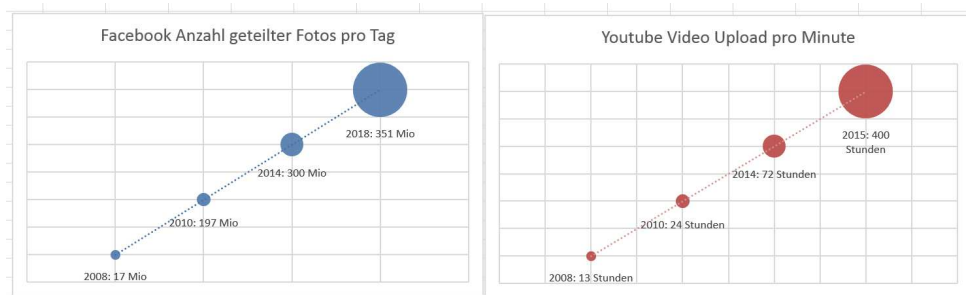


Abbildung 7.4 Andere Indikatoren: Youtube & Facebook

2011 benutzen ca 33 % der Weltbevölkerung das Internet (**Abbildung 7.4**). Der grösste relative Anteil der Bevölkerung hat Europa (74.4%) gefolgt von Nord- und Südamerika (56.3%) und CIS (Commonwealth of Independent States) (47.6%).

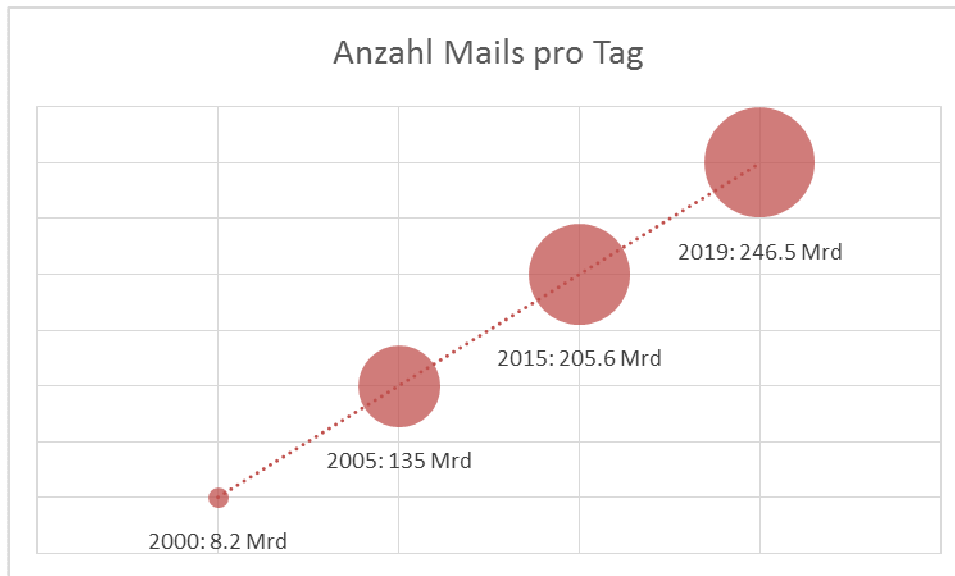


Abbildung 7.5 Entwicklung Anzahl Mails pro Tag

Der globale Internet-Traffic erreicht gemäss dem Cisco Visual Networking Index (VNI) im Jahr 2015 966 Exabytes – also knapp 1 Zettabyte.

Gemäss einer Studie der Research Firma Radicati Group nutzen heute mehr als 2 Milliarden Menschen E-Mail, es existieren jedoch 3'146 Mio E-Mail Accounts (2011). Der typische User, der E-Mail geschäftlich nutzt, empfängt täglich 75 Meldungen und versendet 37 Meldungen. Der tägliche Datentransfer ist gemäss dieser Studie 16 MB, hochgerechnet wären das 10 Petabytes täglich, eine Zahl die sehr unwahrscheinlich klingt.

7.2 Der Wert von Informationen

Die objektive Bewertung von Informationen ist ausserordentlich schwierig. Bis heute existiert noch kein international etablierter Bewertungsstandard. Dieses Kapitel fasst eine Reihe von Ansätzen zusammen:

- *Informations-Produktivität*: Ein Messwert über das gesamte Unternehmen hinweg.
- *Informationswert als Funktion*: Der Wert der Information wird als Funktion in Bezug auf die Information selbst, die Benutzung, den Verwendungszweck und das Resultat gemessen.

- *Der subjektive Wert der Information:* Das Verhältnis WTA (price willing to accept) und WTP (price willing to pay) im Vergleich zu anderen Waren.
- *Der taktische Wert von Informationen:* Die Bewertung von Information im militärischen Umfeld.
- *Information hat gar keinen Preis:* Die These, dass Information kostenlos werden wird.
- *Die Kosten fehlender Informationen:* Die Kosten für den Verlust von Daten entsprechen in dieser Betrachtung dem Wert der Information.

7.2.1 Informations-Produktivität

Paul Strassmann, der ehemalige CIO von XEROX, und Larry Dignan haben mehr als 500 amerikanische Unternehmen bewertet, indem sie deren Informations-Produktivität gemessen hat [Dignan 2004]. Der Wert soll aufzeigen, wie ein Unternehmen mit Information umgeht.



Informations-Produktivität

Information Productivity= Information Value-Added (output) divided by Transaction Costs (input)

- Information Value Added (IVA): Gewinn nach Steuern, jedoch vor Wertbereinigungen.
- Transaction Costs (TA): Kosten für Management, Planung, Rechnungswesen, Marketing und Forschung.

Seine These, dass diejenigen Firmen, die am Geschicktesten mit Informationen umgehen auch diejenigen sind, die am profitabelsten arbeiten, ist statistisch nicht belegt.

7.2.2 Informationswert als Funktion

Dr. Myron L. Cramer, Direktor der Windermere Information Technology Systems, stellte 1997 anlässlich einer Konferenz zum Thema Information Warfare eine Formel zur Messung des Wertes von Informationen vor [Cramer 1997]:

Wert der Information = Funktion (Information, Anwendung, Verwendungszweck, Aktionen, Resultat)

Dr. Cramer geht davon aus, dass der Wert der Information als Geldwert in einem wirtschaftlichen Kontext repräsentiert wird. Gemessen werden die Quantität, die Qualität und die Zeit.

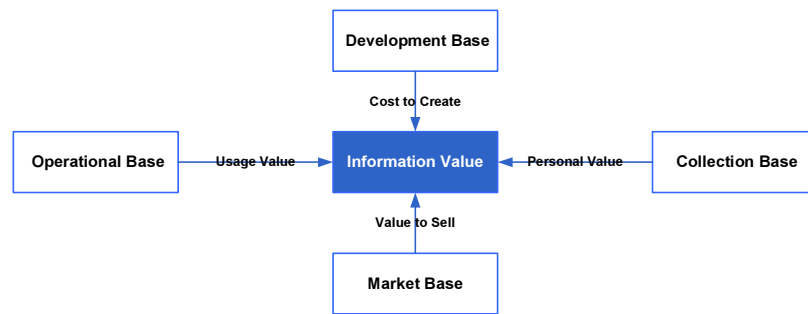


Abbildung 7.6 Der Wertefluss

Die Basis der Bewertung umfasst die vier Faktoren Aufwand zur Erstellung, den Wertbeitrag, Wiederverkaufswert und Sammelwert von Informationen, wie in **Abbildung 7.6** dargestellt.

- *Development Base*: Der Aufwand zur Erstellung einer Information.
- *Operational Base*: Der Wertbeitrag einer Information in einer bestimmten Situation.
- *Market Base*: Der situative Wiederverkaufswert einer Information und deren Aktualität.
- *Collection Base*: Der Sammelwert einer Information der über den Aufwand, den Wertbeitrag und den Wiederverkaufswert hinausgeht.

7.2.3 Informationsklassen

Informationen, die im Kontext eines Unternehmens (einer Organisation) eingesetzt werden, sind in verschiedene Klassen oder Arten aufgeteilt:

- *Produktgestaltung, Produktkonstruktion*: Das technische Design oder auch das Engineering eines Produktes. Diese Informationen werden oft als CAD / CAM Daten abgelegt.
- *Technische Produktdaten*: Diese Informationen sind in Benutzerhandbüchern oder Betriebshandbüchern beschrieben.
- *Betriebsführung*: Die Informationen der Betriebsführung umfassen Marketingpläne, Geschäftspläne, Geschäftsgrundsätze, geschäftspolitische Grundlagendokumente u.v.a.m.
- *Operative Datenbanken*: Sie enthalten Informationen über Produkte, Kunden, Aufträge, Inventar und die Finanzen.
- *Operative Abläufe*: Die Beschreibung von Techniken und Methoden der Geschäftsabläufe.
- *Technische Datenbanken*: Die technischen Basisdaten zur Produkterzeugung und zur Unterstützung der operativen Prozesse.

- *Wissen der Mitarbeiter / Mitarbeiterinnen:* Diese Information umfasst die Ausbildung, das Training und die Erfahrung des Personals sowie das Markt- und Kundenverständnis der Firma.
- *Software:* Programme (nur ausführbare oder mit Quellcode) die die Gesamtheit der betrieblichen Informationssysteme eines Unternehmens umfassen.

7.2.4 Information und betriebliche Prozesse

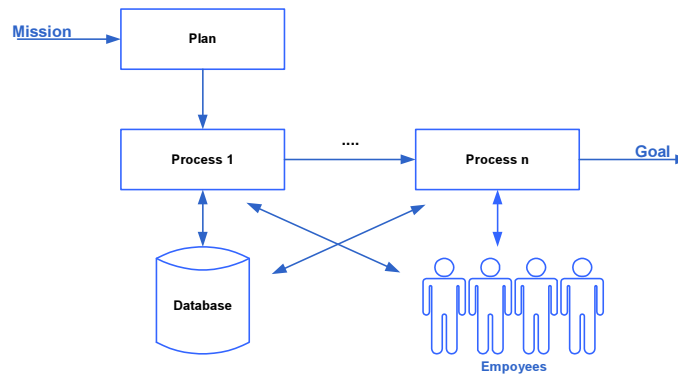


Abbildung 7.7 Die Rolle der Information im Geschäftsprozess

Die verschiedenen betrieblichen Informationsarten können dem generellen Prozessmodell zur Durchführung betrieblicher Abläufe zugeordnet werden, wie in **Abbildung 7.7** dargestellt.

Tabelle 7.1 Typisierung der Geschäftsabläufe und der entsprechenden Informationen

Typ	Verwendete Information	Beitrag und Einfluss	Risiko
Mission	Aufgabenbeschreibung	Führt Personal und definiert Prozesse	Verfügbarkeit, Integrität
Plan	Planresultate	Plant und kontrolliert Ressourcen zur Aufgabenerfüllung	Verfügbarkeit, Vertraulichkeit, Integrität
Prozesse	Technologie und Prozesse zur Produkterstellung	Beeinflusst andere Prozesse	Verfügbarkeit, Integrität
Datenbanken	Operationelle und technische Daten	Unterstützt alle Prozess-Schritte	Verfügbarkeit, Vertraulichkeit, Integrität
Personal	Wissen der Mitarbeitenden	Beeinflusst die Planung und die Ausführung der Prozesse	Verfügbarkeit

7.2.5 Der subjektive Wert der Information

Die beiden Wissenschaftler Sheizaf Rafaeli und Daphne R. Raban der Universität von Haifa untersuchten den subjektiven Wert der Information anhand eines einfachen Experimentes [Rafaeli, Raban 2003].

7.2.5.1 Der Limonadenstand-Versuch

Die Versuchsperson besitzt einen Limonadenstand und muss Gewinn durch den Verkauf von Limonade an Passanten erwirtschaften. Die Qualität des Produktes kann geändert werden (Zuckergehalt, Zitronen und Eis). Ausserdem kann die Versuchsperson entscheiden, wie gross das Lager (Zucker-, Zitronen- und Eismenge) ist und wie hoch der Preis für einen Becher Limonade ist. Ausserdem sind Wetterinformationen verfügbar, die jedoch eingekauft werden müssen. Die Wetterinformation und die Lagerbestände können gehandelt werden. Ein Verkaufstag verläuft vollautomatisch. Die Laufkundschaft gibt Kommentare über die Ware ab.

Die Resultate des Experimentes zeigten, dass die Differenz zwischen dem WTA (price willing to accept) und dem WTP (price willing to pay) um den Faktor 3 liegt. Dieser Unterschied, auch Besitztumseffekt genannt, entspricht der Norm bei Handelswaren. Information ist also, gemäss diesem Versuch, eine Handelsware wie jede andere auch.

Gemäss Fafaeli und Raban ist Information eine Ware, die teuer in der Herstellung, jedoch sehr billig in der Reproduktion ist. Weder die Produktionskosten noch der Wert der Information stehen in einer Relation zur Quantität des Produktes. Da Information lediglich ein Werkzeug zur Unterstützung von Entscheidungen darstellt, ist eine direkte Messung des Wertes nicht angemessen.

Aus theoretischer Sicht existieren drei Arten der Bewertung von Informationen:

- *Normative Bewertung*: Eine Information hat unabhängig von ihrem Verwendungszweck einen fixen Wert. Eine normative Bewertung der Information muss sich anhand objektiv messbaren Kriterien (Quantität, Zeit) festlegen lassen. Diese Bewertung erlaubt keinen Einbezug der wichtigsten Messgrösse einer Information, der Qualität.
- *Realistische Bewertung*: Die realistische Bewertung einer Information erfolgt nach der Benutzung der Information. Diese Bewertung ist demzufolge für die Evaluation von Inhalten nicht zu verwenden. Sie hilft also nicht bei der Entscheidung, welche Information in der konkreten Situation bezogen werden soll.
- *Subjektive Bewertung*: Die subjektive Bewertung der Information erfolgt durch das Individuum, welches die Information beziehen möchte.

Der subjektive Wert einer Information wird von einer Reihe von Faktoren beeinflusst.

Wir tendieren dazu, vorhandene Informationen wie Wahrscheinlichkeit und Stichprobenauswahl zu ignorieren. Nur so sind Investition mit Renditeversprechungen, die weit ausserhalb des Kapitalmarktdurchschnitts liegen, zu erklären. Anstatt eine Bewertung und damit auch einen Kaufentscheid auf vorhandenen Informationen abzustützen, entscheiden die meisten Menschen aufgrund vom repräsentativen Charakter (wie wahrscheinlich er-

scheint eine Information), Verfügbarkeit (wie und wann ist die Information verfügbar) und der Anpassung und Verankerung von Informationen (Überbewertung einer ersten Grundinformation oder falsches Schliessen aus dieser Information).

Rafaeli und Raban haben mit dem Limonadenstand-Versuch die Einflussfaktoren auf die subjektive Bewertung von Information untersucht und sind zu folgenden Feststellungen gekommen:

- *Die Reduzierung der Ungewissheit:* Im Moment der Suche nach einer bestimmten Information ist selten bekannt, welche Art von Information gefunden werden wird, wie die Qualität dieser Information ist und in welchem Ausmass sie tatsächlich hilft, die Ungewissheit zu reduzieren.
- *Indikatoren:* Einige a priori Indikatoren zur subjektiven Bewertung von Informationen, wie Filmtrailer, "getAbstract" Mechanismen oder die Amazon Empfehlungen sind zwar vorhanden, jedoch weder objektiv noch vollständig noch repräsentieren sie tatsächlich immer die Qualität der vollständigen Information.
- *Der Besitztumseffekt:* Dieser Effekt wird auch "endowment effect" genannt, beschreibt den Unterschied zwischen dem Preis, den wir für eine Information verlangen wenn wir sie besitzen, und dem Preis, den wir zu zahlen bereit sind, wenn wir sie noch nicht besitzen. Dieser Effekt hat zur Folge, dass der Handel mit Information gebremst wird.

7.2.6 Information hat gar keinen Preis

7.2.6.1 Die drei Gesetze der Cyber-Ökonomie

Michel Bauwens, ein belgischer Journalist, veröffentlichte 1996 die drei Gesetze der Cyber-Ökonomie [Bauwens 1996]:

- 1. Der Preis der Information wird gegen Null tendieren.
- 2. Der Preis der Kommunikation wird gegen Null tendieren.
- 3. Der Preis der Transaktion wird gegen Null tendieren.

Ausgangspunkt seiner Überlegungen war die Tatsache, dass mit der Verbreitung des Internets die Gestehungskosten (Kosten für die Herstellung) für die Produktion von Information relativ klein sind und eine Notwendigkeit zur kostenlosen Informationsverbreitung besteht. Beispiel für seine erste These ist die Tatsache dass Adobe seinen Acrobat Reader gratis abgibt, Netscape damals seinen Browser gratis verteilte, jedoch für den Web Server Geld verlangte und Sun Microsystems JAVA allen zur Verfügung stellte, jedoch für den professionellen Einsatz Lizenzgebühren verlangte.

7.2.6.2 Der Patentstreit

Der in den letzten Jahren geführte Streit um die Anwendung des Urheberrechtes und der Patente auf Software im Rahmen der Diskussion um den Richtlinienentwurf der EU-Kommission wurde und wird mit äusserster Heftigkeit geführt. Gemäss Dr. Melullis, Rich-

ter am Bundesgerichtshof, kann das Urheberrecht in seiner konkreten Ausgestaltung keinen adäquaten Schutz leisten, da dieses Recht vom "Werk" (Literatur und Kunst) ausgeht [Melullis 2000]. Ein Werk ist für bis zu 70 Jahren nach dem Tode des Verfassers / der Verfasserin geschützt.

Die Diskussionen um das Softwarepatent, insbesondere diejenige um den Vorschlag der EU-Kommission, ist gerade in diesem Jahr intensiv geführt worden. Kritiker der Patentierung von Software sind nicht nur in den Reihen der Open-Source Gemeinde zu finden. So kommentiert die Monopolkommission des Bundespräsidenten:

"Durch Softwarepatente werden insbesondere für kleine und mittlere Unternehmen erhebliche Marktbarrieren entstehen. Open-Source-Softwareprodukte stellen Ex Definitione den Programmcode allen Interessierten zur Verfügung und können Patentschutz deshalb generell nicht in Anspruch nehmen. Die mit dem Patentschutz verbundene vorübergehende Monopolstellung eines Unternehmens ist geeignet, die Konzentrationstendenzen auf dem Markt für Softwareprodukte weiter zu verstärken und den Wettbewerb zu behindern."
[Richter 2003]

Wie wenig übertragbar die Aussagen der Monopolkommission sind, lässt sich durch ein einfaches Umschreiben des Textes auf eine andere Industrie zeigen:

Durch Patente der Autoindustrie werden insbesondere für kleine und mittlere Unternehmen erhebliche Marktbarrieren entstehen. Open-Car-Produkte stellen Ex Definitione das Fahrzeug allen Interessierten zur Verfügung und können Patentschutz deshalb generell nicht in Anspruch nehmen. Die mit dem Patentschutz verbundene vorübergehende Monopolstellung eines Automobil-Unternehmens ist geeignet, die Konzentrationstendenzen auf dem Markt für Autos weiter zu verstärken und den Wettbewerb zu behindern.

Ein solcher Kommentar durch eine deutsche Bundesbehörde ist wohl kaum denkbar.

7.2.7 Der taktische Wert von Informationen

Gemäss dem Handbuch FM 100-6 Information Operations der amerikanischen Armee ist Information anhand der Kriterien Richtigkeit, Relevanz, Rechtzeitigkeit, Brauchbarkeit, Vollständigkeit und Genauigkeit zu bewerten [Army 1996].

- *Richtigkeit*: Richtige Informationen sind solche, die der Sachlage entspricht.
- *Relevanz*: Diejenigen Informationen, die in Bezug auf die Mission, die Aufgabe oder die Situation relevant sind.
- *Rechtzeitigkeit*: Informationen, die genug Zeit zur Entscheidung lassen.
- *Brauchbarkeit*: Informationen, die in gebräuchlichen und leicht verständlichen Formaten und Darstellungen vorliegen.
- *Vollständigkeit*: Alle Informationen, die für die entscheidende Person notwendig sind, sind auch vorhanden.
- *Genauigkeit*: Die Information hat den notwendigen Detaillierungsgrad.

Am wertvollsten sind Informationen, die richtig und relevant sind. In zweiter Lesung sollten Informationen rechtzeitig und in brauchbarer Form vorhanden sein. Zusätzlich sollten sie vollständig und genau sein. Die Faustregel zur Bewertung lautet: Unvollständige oder ungenaue Informationen sind besser als gar keine; nicht rechtzeitige und nicht brauchbare sind keine Informationen; falsche oder unwichtige Informationen sind schlimmer als keine Informationen.

7.2.8 Die Kosten fehlender Informationen

Thomas Redman, der Präsident der Firma Navesik Consulting Group in New York, hat 1998 in einer Zusammenfassung einer Reihe von Untersuchungen zum Thema Datenqualität festgestellt, dass durchschnittlich 1-5 % aller Daten eines Unternehmens fehlerhaft sind [Redman 1998].

1. Kosten

- 8-12% des Einkommens eines Unternehmens (Durchschnittswerte aus verschiedenen Case-Studies).
- 40-60% der Ausgaben für Dienstleistungs-Unternehmen.

2. Andere Einflussfaktoren

- Langsamere und nicht akkurate Entscheidungen
- Schwierigere Strategie-Definition und –Umsetzung
- Zunehmendes Misstrauen dem Unternehmen gegenüber
- Ablenkung der Aufmerksamkeit der Unternehmensführung

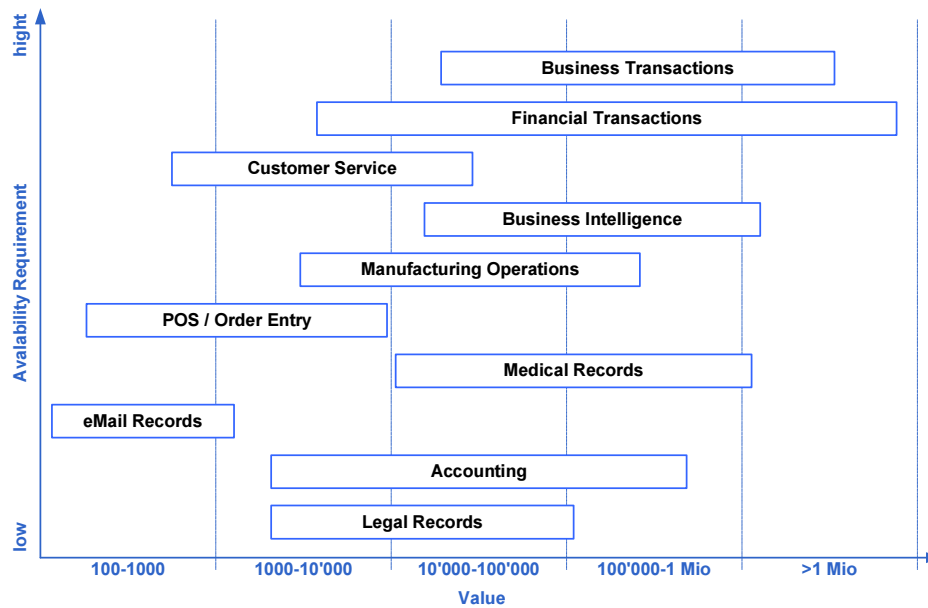


Abbildung 7.8 Informationskosten pro Daten-Kategorie

Gemäss der SNIA (Storage Networking Industry Association) existieren die in **Abbildung 7.8** dargestellten generellen Daten-Kategorien in einem Unternehmen [Peterson 2004]:

- *Business Transactions*: Transaktionen auf Firmenebene wie der Kauf oder Verkauf von Firmen oder Firmenbereichen, IPO's, Zusammenschlüsse.
- *Financial Transactions*: Finanztransaktionen wie Geldüberweisungen, Verwaltung von Aktien und Portfolios.
- *Customer Service*: Kundendaten eines Unternehmens, wie beispielsweise Kontaktinformationen, Bestellverhalten, etc.
- *Business Intelligence*: Daten, die zur Planung und Steuerung eines Unternehmens notwendig sind.
- *Manufacturing Operations*: Alle Daten, die für die Produktion von Gütern benötigt werden.
- *POS / Order Entry*: Daten betreffend dem Auftragseingang, respektive Daten aus dem Verkauf (POS: Point of Sales).
- *Medical Records*: Patienten-Daten sowie Daten über die Wirkung von Medikamenten.
- *eMail Records*: Alle eMails einer Firma mit den Attachments.
- *Accounting*: Buchungsdaten und Buchungsbelege sowie Rechnungsabschlüsse wie Bilanz und Erfolgsrechnung.
- *Legal Records*: Verträge und andere juristisch relevante Dokumente.

7.3 Die Klassierung von Datenbeständen

Voraussetzung für Data Lifecycle Management ist eine geeignete Klassierung von Daten, die als Grundlage zur Bewertung und zur Verwaltung dient.

Die Klassierung von Daten kann unter verschiedensten Aspekten erfolgen:

- Einfache pragmatische Klassierung aufgrund der Wichtigkeit der Daten für ein Unternehmen.
- Die Klassierung der Data Protection Initiative des SNIA (Storage Networking Industry Association).
- Eine Branchenorientierte Klassierung als Erweiterung der SNIA Klassierung.
- Eine Klassierung der Daten anhand von einzelnen Kriterien wie dem Format, des Alters oder der Änderungsrate der Daten.

7.3.1 Generelle Bewertungs-Schemas

7.3.1.1 Einfache Klassierung von Daten

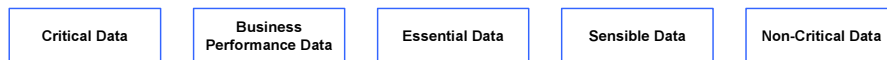


Abbildung 7.9 Eine mögliche Klassierung von Daten

Eine einfache Klassierung von Daten erfolgt anhand der Wichtigkeit der Daten für die Tätigkeit eines Unternehmens, wie in **Abbildung 7.9** skizziert.

- *Critical Data*: Daten, die für die wichtigen Geschäftsprozesse benötigt werden und deren Verlust zu einer operativen Katastrophe führen kann (die Leistungserstellung kann nicht mehr erfolgen). Sowie Daten, die aus rechtlichen Gründen aufbewahrt werden müssen.
- *Business Performance Data*: Daten, die für die Steuerung und die Planung eines Unternehmens relevant sind und deren Verlust zu einer unternehmerischen Katastrophe führen kann.
- *Essential Data*: Daten, die für das Daily Business verwendet werden und damit Teil des Business Know-hows seines Unternehmens sind.
- *Sensible Data*: Daten, die für das Daily Business verwendet werden, die jedoch entweder schnell wiederhergestellt werden können oder durch alternative Daten ersetzt werden können.
- *Non-Critical Data*: Daten, die mit geringen Kosten wiederhergestellt werden können oder aber Duplikate bestehender Daten.

7.3.1.2 Klassierung der Data Protection Initiative

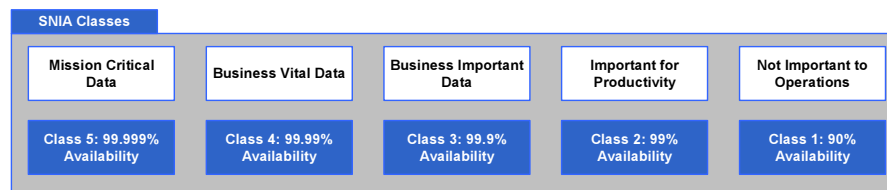


Abbildung 7.10 Die Klassierung von Daten der SNIA

Die SNIA (Storage Networking Industry Association) hat eine klar strukturierte Klassierung definiert (**Abbildung 7.10**). Diese Klassierung wird mit Vorschriften zur Behandlung der einzelnen Datenklassen ergänzt. Dabei werden folgende Bereiche umschrieben:

- **RPO (Recovery Point Objective):** Der Zeitraum zwischen zwei so genannten "Data Protection Events". Dies bedeutet, die maximale Menge der Daten, die verloren gehen dürfen.
- **RTO (Recovery Time Objective):** Wie viel Zeit darf vergehen, bis ein System wiederhergestellt ist.
- **DPW (Data Protection Window):** Der Zeitraum, der verwendet werden darf, um Daten zu sichern während das System nicht online ist.

Tabelle 7.2 Die Kennwerte der Datenklassen

Data Protection Class	Availability	RPO Dataloss	RTO Downtime	DPW Backuptime
1: Not Important	90 %	1 Week	1 Week	Days
2: Important for Productivity	99%	1 Day	1 Day	12 Hours
3: Business Important	99.9	2 Hours	2 Hours	10 Minutes
4: Business Vital	99.99%	10 Minutes	15 Minutes	None
5: Mission Critical	99.999%	1 Minute	1.5 Minutes	None

7.3.2 Die Datenklassierung für Branchen

Branchenklassierung

	Finance	Telco	Public Services	Industry	Retail
Business Transactions	5	3	1	1	1
Financial Transactions	5	3	2	3	3
Customer Service	3	5	4	2	4
Business Intelligence	4	4	2	3	4
Manufacturing Operations	1	1	1	5	4
POS / Order Entry	1	3	2	3	5
Medical Records	1	1	4	1	2
eMail Records	3	2	3	2	3
Accounting	4	3	3	2	2
Legal Records	4	3	4	2	2

Abbildung 7.11 Beispiel einer branchenorientierten Klassierung von Daten

Der Wert und damit auch die Klassierung von Daten sind für jede Branche verschieden. So sind für die Finanzbranche die Informationen über Finanztransaktionen und die Buchhaltungsdaten wesentlich wichtiger als für einen Automobilhersteller. Eine mögliche Klassierung ist in **Abbildung 7.11** und in **Abbildung 7.12** dargestellt.

Class
Class 1: Not Important
Class 2: Important for Productivity
Class 3: Business Important
Class 4: Business Vital
Class 5: Mission Critical

Abbildung 7.12 Die verschiedenen Klassen und ihr Farbcode

Für die Finanzindustrie sind Daten, die sich auf Transaktionen beziehen absolut kritisch, während für die Telekommunikationsbranche die Kundendaten wichtig sind. Der Auftragseingang ist die Basis für das Retail Geschäft und die Daten der Produktionssteuerung sind essentiell für die industrielle Fertigung.

7.3.3 Klassierung in Kombination mit dem Wert der Daten

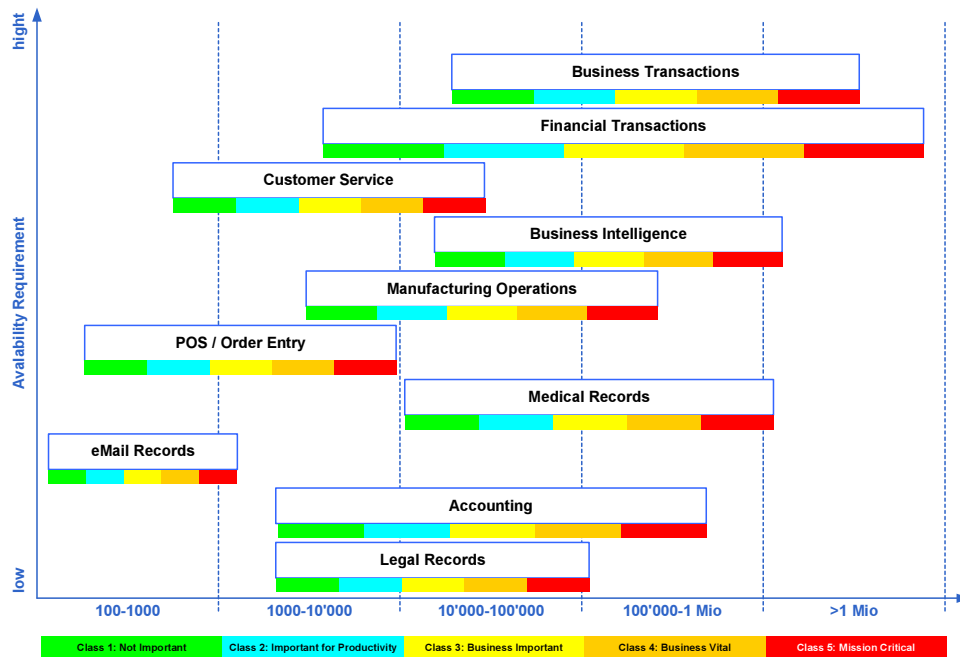


Abbildung 7.13 Die Bewertung von Daten kombiniert mit der Klassierung

Die Bewertung der Daten wird mit der fünfstufigen SNIA Klassierung hinterlegt. Daraus ergibt sich eine Abstufung und eine stufenweise Bewertung der Daten wie in **Abbildung 7.13** dargestellt. Unter der Annahme, dass die Daten entsprechend der Klassierung bewertet werden ergibt sich die Bewertungsskala der **Abbildung 7.14**.

Bewertung der Daten	Min	Max	1	2	3	4	5
Business Transactions	25'000	5'000'000	50'000	100'000	400'000	900'000	4'000'000
Financial Transactions	5'000	9'000'000	9'000	60'000	250'000	900'000	5'500'000
Customer Service	700	50'000	900	3'000	6'000	9'500	40'000
Business Intelligence	20'000	1'500'000	40'000	80'000	450'000	850'000	2'500'000
Manufacturing Operations	5'000	500'000	6'000	18'000	55'000	80'000	300'000
POS / Order Entry	200	10'000	400	800	2'000	4'000	8'000
Medical Records	10'000	1'000'000	20'000	60'000	150'000	450'000	850'000
eMail Records	150	1'100	200	400	600	800	1'000
Accounting	1'300	700'000	6'000	15'000	50'000	100'000	450'000
Legal Records	1'300	110'000	5'000	8'000	20'000	50'000	90'000

Abbildung 7.14 Bewertungsskala für Daten

Daraus ergeben sich folgende Faktoren für die Werte der Daten pro Industrie (Basierend auf der branchenorientierten Klassierung), wie in **Abbildung 7.15** aufgelistet. Dabei ist zu beachten, dass es sich um abgebildete Schätzwerte handelt. Die Größenordnung und nicht der konkrete Wert ist entscheidend.

Calculation	Finance	Telco	Public Services	Industry	Retail
Business Transactions	4'000'000	400'000	50'000	50'000	50'000
Financial Transactions	5'500'000	250'000	60'000	250'000	250'000
Customer Service	6'000	40'000	9'500	3'000	9'500
Business Intelligence	850'000	850'000	80'000	450'000	850'000
Manufacturing Operations	6'000	6'000	6'000	300'000	80'000
POS / Order Entry	400	2'000	800	2'000	8'000
Medical Records	20'000	20'000	450'000	20'000	60'000
eMail Records	600	400	600	400	600
Accounting	100'000	50'000	50'000	15'000	15'000
Legal Records	50'000	20'000	50'000	8'000	8'000
SUM	10'533'000	1'638'400	756'900	1'098'400	1'331'100

Abbildung 7.15 Faktoren für die Bewertung pro Branche

Die Kosten beziehen sich auf die durchschnittlichen Kosten pro Incident, falls die entsprechenden Daten verloren gehen, respektive nicht im definierten Zeitraum (RPO) wiederhergestellt werden können.

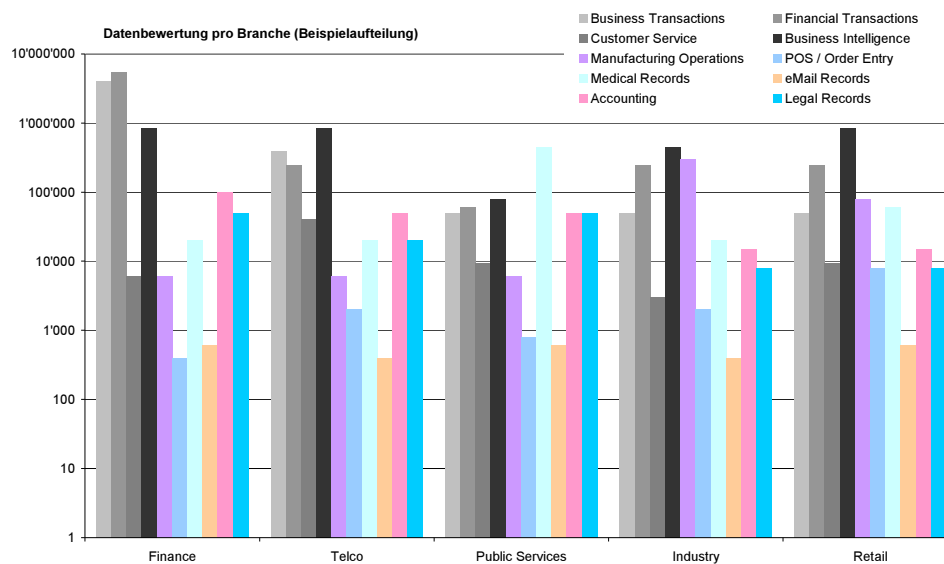


Abbildung 7.16 Datenbewertung pro Branche und Datenkategorie

Die dargestellten Werte sind eine Schätzung und betreffen die Kosten eines Datenausfalles pro Case (Data Protection Event). Werden nun die verschiedenen Bewertungen, wie in **Abbildung 7.11** auf die Datenklassen abgebildet, so ergibt sich der Wert der Daten pro Branche wie in **Abbildung 7.16** dargestellt

7.3.4 Beispiel: Daten-Bewertung der Finanzbranche

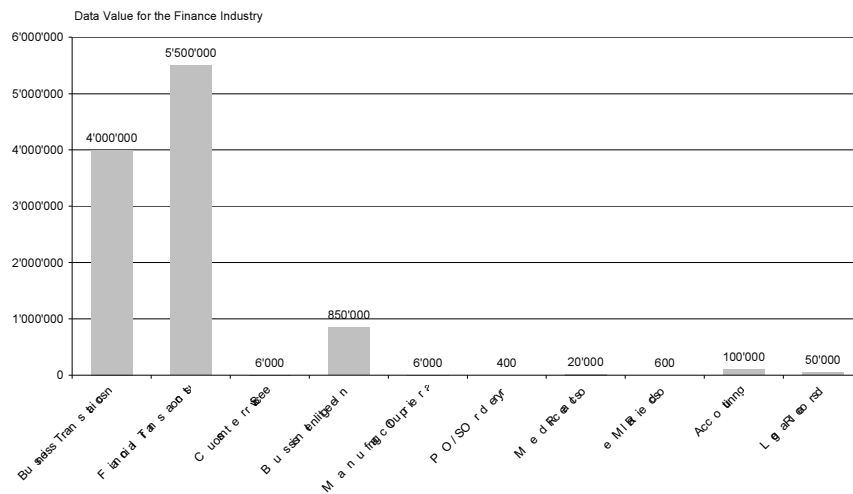


Abbildung 7.17 Der Wert der Daten in der Finanzindustrie

Die dargestellten Werte sind eine Schätzung und betreffen die Kosten eines Datenausfalles pro Case (Data Protection Event).

7.3.5 Beispiel: Daten-Bewertung der Telekommunikations-Branche

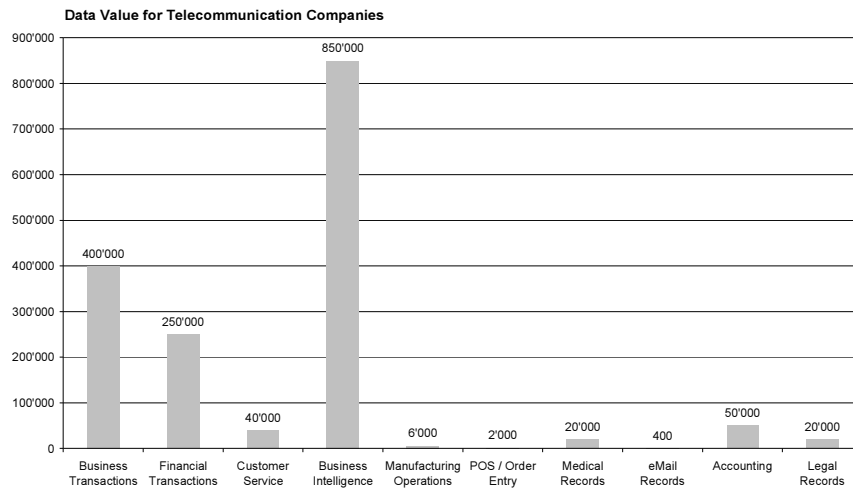


Abbildung 7.18 Der Wert der Daten in der Telekommunikationsindustrie

7.3.6 Beispiel: Daten-Bewertung der Industrie

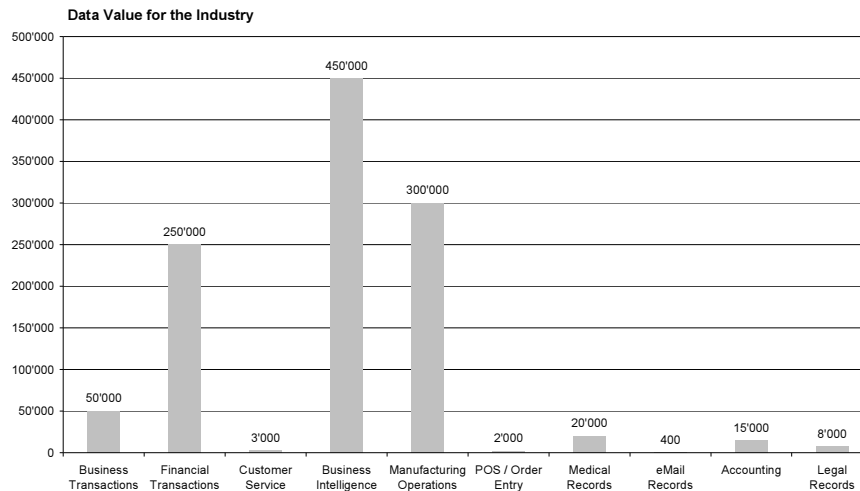


Abbildung 7.19 Der Wert der Daten in der Fertigung

7.3.7 Die Klassifikation von Branchen

Wirtschaftsbranchen werden durch verschiedene Klassifikationen aufgeteilt. Die verbreiteten Branchenklassifikationen sind SIC, NAICS und NACE:

- *Standard Industrial Classification (SIC)*: Die SIC wurde 1939 in den Vereinigten Staaten für amtliche Statistiken kreiert. SIC umfasst 10 Hauptklassen.
- *North American Industry Classification System (NAICS)*: Erweitert das SIC um die Dienstleistungsbranchen und erlaubt auch die Klassifikation der modernen Branchen.
- *Nomenclature général des activités économiques dans les Communautés Européens (NACE)*: Durch Verordnung der Europäischen Union (vom Oktober 1990) wird zum 1. Januar 1993 die NACE Rev. 1 als verbindliche Klassifikation der Wirtschaftszweige zum Zwecke der amtlichen Statistik bei Eurostat und bei den statistischen Ämtern der EU-Mitgliedsländer vorgeschrieben.

7.3.8 Klassierung anhand von einzelnen Kriterien

7.3.8.1 Klassierung anhand vom Datenformat

- *Strukturierte Daten:* Strukturierte Daten liegen in den meisten Fällen in numerischer oder alphanumerischer Form oder einem anderen einfachen Format vor. OLTP und Data Warehouse Systeme sowie relationale Datenbanken verwenden strukturierte Daten.
- *Unstrukturierte Daten:* Meist liegen unstrukturierte Daten als grosse Datenmenge ohne interne Gliederung (z.b. eine Firma) oder als sehr komplexe Kombination einfacher Datentypen (CAD Daten) vor. Web Grafiken sind unstrukturierte Daten. XML Datenbanken oder auch in seltenen Fällen Objektorientierte Datenbanken werden für unstrukturierte Daten eingesetzt.
- *Semi-strukturierte Daten:* Das klassische Beispiel semi-strukturierter Daten ist Text. Viele Textverarbeitungsprogramme verwenden eine Mischung aus strukturierten und unstrukturierten Daten, alles gespeichert in einer einzigen Datei.

7.3.8.2 Klassierung anhand des Alters der Daten

- *Current Data:* Auf aktuelle Daten muss oft sehr schnell zugegriffen werden. Diese Daten müssen bei Ausfall so schnell wie möglich wiederhergestellt werden. Sie werden so nahe wie möglich am Hauptspeicher gehalten.
- *Aging Data:* Diese Datenart ist mehr als drei Tage und weniger als ein Jahr alt. Diese Daten werden oft nur im Rahmen eines grösseren Zugriffs benötigt. Einzelne Zugriffe sind selten. Diese Daten werden sehr oft auf Disk gehalten.
- *Archived Data:* Archivierte Daten sind normalerweise älter als ein Jahr. Sie werden meist nur aufgrund von Anforderung bei der Prüfung auf Governance Standards oder auf staatliche Regulatorien (SOX, Basel II) benötigt.

7.3.8.3 Klassierung aufgrund der Änderungsrate der Daten

- *Dynamic Data:* Dynamische Daten ändern sehr oft (Beispiel: Kundendaten).
- *Static Data:* Statische Daten werden einmal gespeichert und ändern anschliessend nicht mehr (Beispiel: Logfiles, Eventrecords).

7.4 DLM Konzepte

Die Entwicklung eines Data Lifecycle Management, respektive eines Information Lifecycle Management Konzeptes hängt von einer Reihe von externen Requirements und der Erbringung von Vorleistungen ab. Die Vorgaben sind die Bewertung der Information durch das Unternehmen, die vereinbarten Service Level Agreements im Betrieb, sowie die Defi-

nition der Zugriffsrechte und der Datenintegrität. Als Vorleistungen gelten heute die Konsolidierung und Virtualisierung der Storage Infrastruktur, sowie die Einführung eines Storage Resource Managements.

Das Konzept selbst ist eine Kombination aus Technologie und Prozessen. In diesem Kapitel werden drei mögliche Lösungsansätze skizziert, das Information Lifecycle Management Konzept des BITKOM, der Information Lifecycle Management Ansatz der Firma StorageTek und das Vorgehen der SNIA.

7.4.1 Rahmenbedingungen

Die Klassierung von Informationsobjekten findet einerseits auf der Ebene der Bewertungskriterien aus Unternehmenssicht statt, andererseits ist eine technische Abbildung der unternehmerischen Vorgaben in einzelne Klassen zu realisieren. Einzelnen Informationsobjekten werden nicht mehr individuelle Attribute zugeordnet, sie werden lediglich verschiedenen technischen Storage Klassen zugewiesen. Alle Operationen eines Data Lifecycle Management beziehen sich dann auf die einzelnen Storage Klassen, die dann ihrerseits eine Menge von zugeordneten Informationsobjekten umfassen.

7.4.2 System Managed Storage

Eine technische Umsetzung zur Verwaltung von Informationsobjekten ist in dem von IBM definierten System Managed Storage (SMS) realisiert worden [Gelb 1999]. Dabei werden 3 Klassen definiert und je nach Bedarf mit den entsprechenden Attributen instanziiert.

Tabelle 7.3 Die drei wichtigsten SMS Klassen

Klasse	Standartattribute
Data-Class	Platzbedarf, Dateiorganisation
Storage-Class	Zugriffszeit, Verfügbarkeit
Management-Class	Datensicherung, Migration und Aufbewahrungsfrist

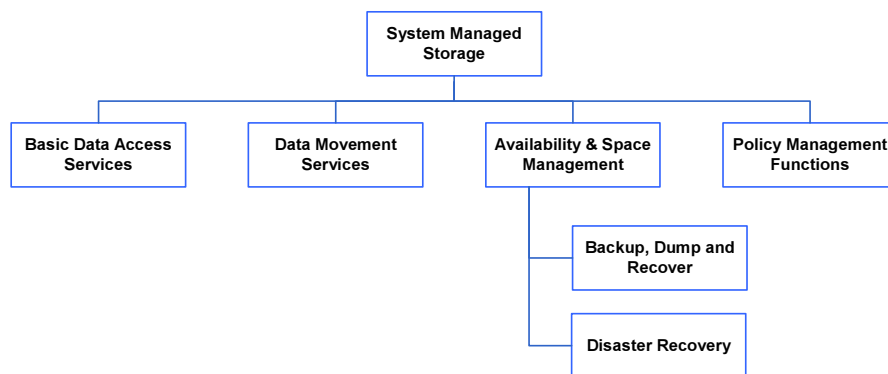


Abbildung 7.20 System Managed Storage

Der Funktionsumfang von System Managed Storage [Yamazaki et al. 1999] umfasst, eine Reihe von Services, wie in **Abbildung 7.20** dargestellt.

- *Basic Data Access Services*: Dienste zur Bereitstellung von Zugriffsmethoden, Dateisystemen und die Unterstützung verschiedener Devices.
- *Data Movement Services*: Services für das schnelle Verschieben von Daten, wie das schnelle kopieren ganzer Datenbestände, dump/restore Funktionalität sowie die Konvertierung oder auch das Printing von Datenbeständen.
- *Availability & Space Management*: Backup / Restore und mehrstufige Disaster Recovery Funktionalität.
- *Policy Management Functions*: Die Bereitstellung von System-Management Routinen für Volumes und Data Sets.

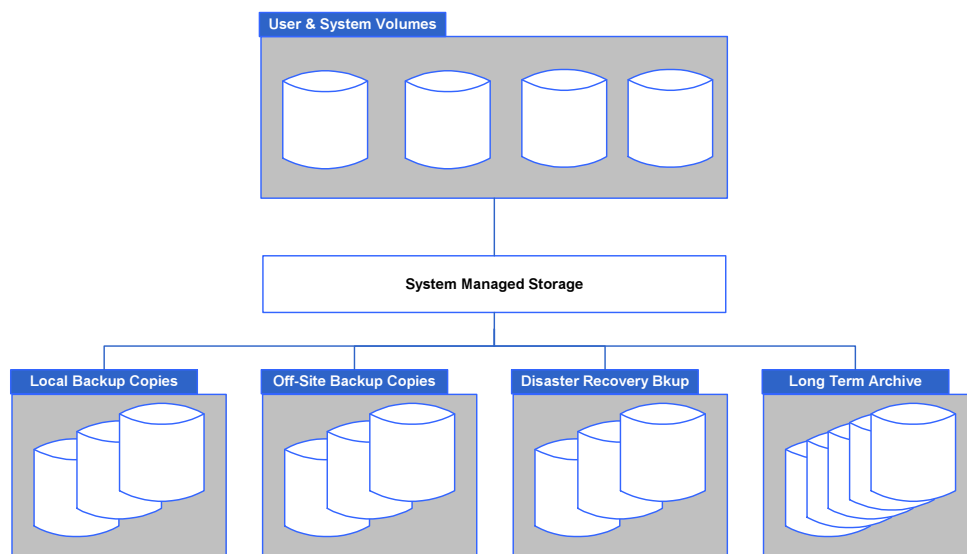


Abbildung 7.21 DLM Funktionen im SMS

Die System Managed Storage Facility von IBM kann bereits eine Reihe von DLM Funktionalität abbilden (**Abbildung 7.21**). Leider ist diese Funktionalität in einem heterogenen und dezentralen Umfeld noch nicht verfügbar. Bemerkenswert ist insbesondere die Disaster Recovery Backup Funktionalität, die IBM ABARS (Aggregate backup and recovery support) nennt.

Basis von ABARS sind so genannte Aggregate Groups, die wiederum einzelne Aggregate enthält. Ein Aggregate beschreibt ein Informationsobjekt mit folgenden Daten:

- *Selection Data Set*: Die Auflistung aller Daten, die im Aggregate enthalten sind.

- *Instruction Data Set*: Die genauen Instruktionen, als die Auflistung aller Kommands, die bei einem Recovery durchzuführen sind.
- *Output Data Set*: Der Output bestehend aus den konkreten Daten im Recovery-Fall und dem Control-File zur Verifikation eines Recoveries.
- *Management Class*: Enthält die Information, welches Ablaufdatum ein Aggregate hat.
- *Count*: Die Anzahl der gleichzeitig zu erstellenden Kopien, die generiert werden müssen.

7.4.3 Storage Ressource Management

Storage Resource Management (SRM) umfasst die Steuerung und die Überwachung der Speicherressourcen und Datenbestände eines Unternehmens durch definierte Managementregeln. Es erlaubt die Verteilung der Unternehmensdaten durch gezielte Automatismen auf alle im Unternehmen vorhandenen Speicher, unabhängig von der konkreten Hardware und Software, die zur Speicherung eingesetzt wird.

Die Merkmale eines SRM sind:

- Unternehmensweite Verwaltung (Überwachung, Planung und Reporting) aller logischen Speicherressourcen.
- Trend- und Backupplanung, basierend auf den täglich geänderten Datenmengen.
- Durchgängiges Management aller Speicherinformationen.
- Interoperabilität zwischen Speichersystemen.

Alle grossen Hersteller von Storage Systemen haben SRM Software in ihrem Portfolio [Thurnhofer 2003].

7.4.4 Hierarchical Storage Management

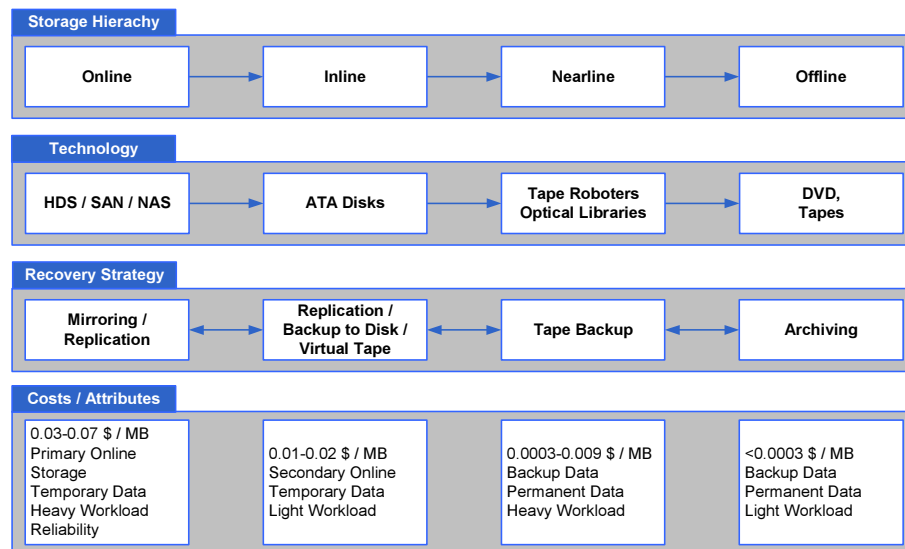


Abbildung 7.22 Hierarchical Storage Management

Hierarchical Storage Management (HSM) ist ein Mechanismus, der inaktive Dateien, auf die seit längerem nicht zugegriffen wurde, markiert und auf die nächste Stufe der Hierarchie transferiert (**Abbildung 7.22**). Das betroffene Informationsobjekt wird verschoben und durch einen Platzhalter repräsentiert. Beim Aufruf einer ausgelagerten Datei, respektive des entsprechenden Platzhalters kopiert das Hierarchical Storage Management die ausgelagerte Datei in die Online Storage oder in die definierte Storage Hierarchiestufe und setzt danach den Bearbeitungsprozess fort.

Bereits werden 5-Stufige Speicherhierarchien in der Storage Industrie diskutiert.

7.4.5 Die Zeit als Faktor

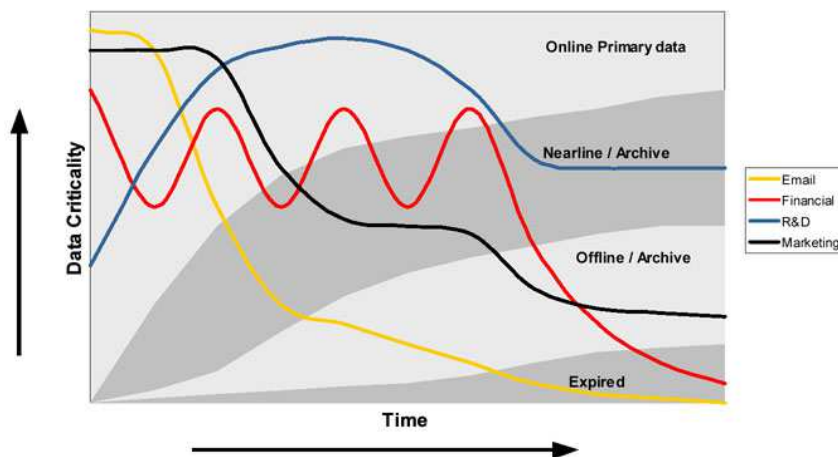


Abbildung 7.23 Wichtigkeit der Daten über die Zeit

Der Wert der Daten ändert sich im Laufe der Zeit, wie in **Abbildung 7.23** dargestellt [Lundberg 2003]. So können dieselben Daten anhand von zeitlichen Vorgaben in verschiedenen Speicherbereichen gehalten werden. Emails verlieren sehr schnell an Aktualität und müssen nur noch sporadisch zugegriffen werden, während beispielsweise Daten aus der Forschung und Entwicklung mit der Zeit sogar an Relevanz für ein Unternehmen zunehmen. Eine Besonderheit stellen die finanziellen Daten eines Unternehmens dar, die periodisch (Quartalsabschluss, Jahresbilanz oder Revision und Steuerprüfung) sehr aktuell sind, in der Zwischenzeit jedoch nicht verändert werden.

7.4.6 ILM des BITKOM

Der Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. (BITKOM) stellt in seinem Leitfaden sein ILM Konzept vor [Born et al. 2004]. Das Konzept basiert auf einer Reihe von Vorleistungen, die unbedingt durchgeführt werden müssen, bevor ILM überhaupt umgesetzt werden kann.

7.4.6.1 Vorleistungen

- *Konsolidierung und SAN / NAS Infrastruktur:* Sämtliche Daten werden in intelligenten Speichernetzwerken gehalten, deren wichtigste Eigenschaft es ist, dass im Prinzip jeder Server auf jeden Speicherbereich Zugriff hat. Dies schafft gemäss ILM die Voraussetzung um eine globale Erstplatzierung durchzuführen.
- *Virtualisierung:* Ein effektives ILM geht von einem problemlosen Zugriff auf Ressourcen aus. Die Virtualisierung von Speicher bedeutet die Abstraktion des verfügbaren Speichers von der unterliegenden Technologie.

- *Storage Ressource Management*: Die Erfassung der Kennzahlen der vorhandenen Speichersysteme, physikalischen Platten, der logischen und virtuellen Volumen, Dateisysteme, Dateien und Datenbanken erfasst und in einer zentralen Datenbank zur langfristigen Verwaltung abgelegt. Diesen Daten werden die entsprechenden Speicherklassen, respektive die entsprechende Klassierung der Informationsobjekte zugeordnet.

7.4.6.2 ILM Prozesse

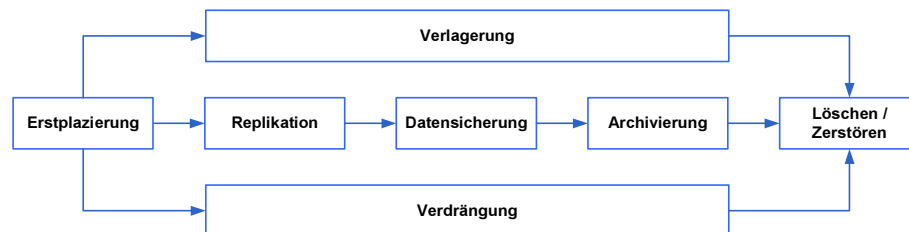


Abbildung 7.24 Information Lifecycle Management Konzept des BITKOM

Die ILM-Prozesse sind in **Abbildung 7.24** dargestellt. Die möglichen Übergänge von einem Prozess in einen Anderen sind in **Abbildung 7.25** skizziert.

- *Erstplatzierung der Daten*: Die Erstplatzierung legt fest, in welcher Speicherklasse ein Informationsobjekt gespeichert werden soll. Die Definition der Klassen ist unternehmensspezifisch.
- *Replikation*: Die Replikation von Daten erfolgt entweder über Mirroring oder über PIT (Point in Time) Mechanismen. Dabei werden sämtliche Daten möglichst aktuell repliziert, respektive nachgeführt.
- *Datensicherung*: Die Datensicherung basiert auf einem Versionsschema. Informationsobjekte werden in regelmässigen Abständen kopiert, ohne dass das Original zerstört wird. So können auch auf ältere Versionen desselben Speicherobjektes zugegriffen werden.
- *Archivierung*: Die Archivierung ist eine gezielte Speicherung von Informationsobjekten in eine andere Speicherklasse. Dabei wird definiert, wie lange ein Informationsobjekt nicht gelöscht werden darf (Vorhaltefrist – Retention).
- *Verdrängung (Hierarchical Storage Management)*: Die Verdrängung verschiebt ein Informationsobjekt von einer Speicherklasse in eine andere ohne Veränderung der Zugriffs-Syntax.
- *Verlagerung*: Die Verlagerung eines Informationsobjektes ist ein irreversibler Vorgang der vor allem im Rahmen vom Einsatz neuer Hardware zum Zug kommt.
- *Löschen / Zerstören*: Beim Löschen wird ein Informationsobjekt lediglich logisch freigegeben, während das Zerstören eines Objektes eine Garantie beinhaltet, dass ein Informationsobjekt nie mehr gelesen oder rekonstruiert werden kann.

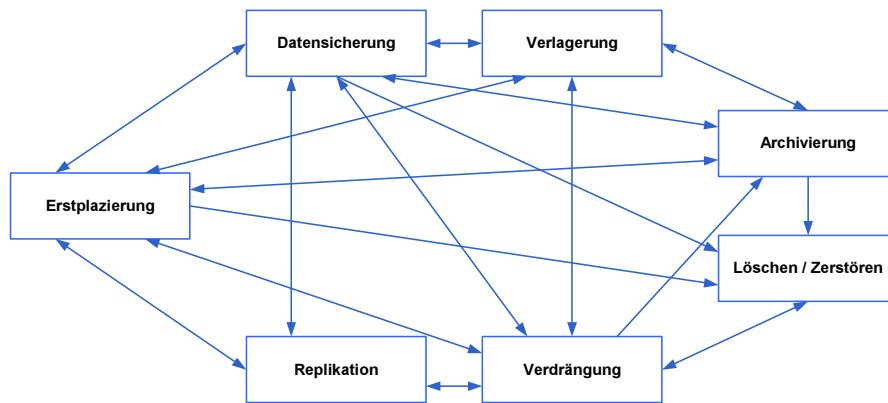


Abbildung 7.25 ILM Prozesse und Übergänge

7.4.7 ILM von StorageTek

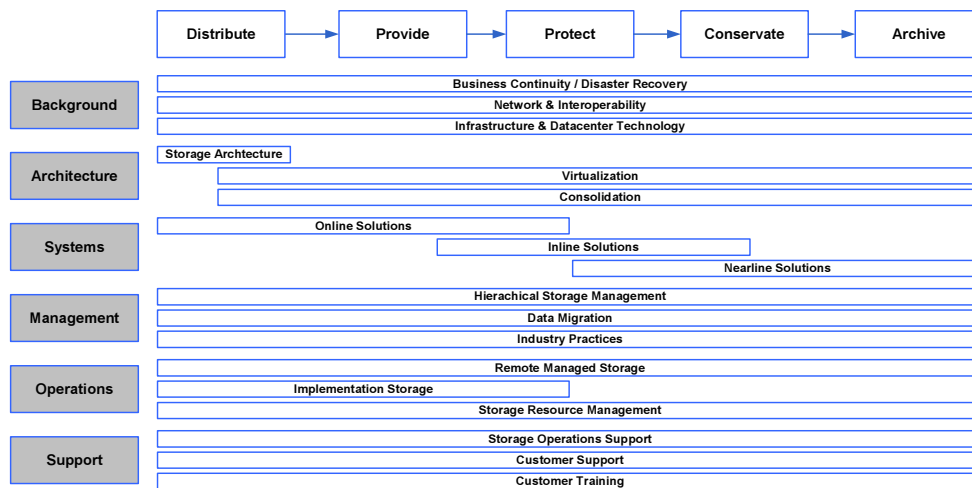


Abbildung 7.26 ILM von der Firma StorageTek

Das Information Lifecycle Management Konzept der Firma StorageTek [Giesselbach 2003] umfasst folgende Kernprozesse:

- *Distribute (Verteilen)*: Der Verteilung der Daten auf den Online Systemen. Dieser Prozess umfasst auch die spezifischen Online-Backup sowie die Replication / Mirroring Mechanismen.
- *Provide (Bereitstellen)*: Die Bereitstellung aller Daten im Online und Inline Bereich. Dieser Prozess umfasst das Inline-Backup.

- *Protect (Schützen)*: Definierte Datenbereiche (Informationsobjekte) werden aufgrund geschäftspolitischer Policies gesichert und mittelfristig bereitgestellt.
- *Conserve (Konservieren)*: Aufgrund regulatorischer Rahmenbedingungen und Business-Überlegungen werden Informationsobjekte konserviert.
- *Archive (Archivieren)*: Die mittel- und langfristige Archivierung von Daten im Nearline- und Offline-Bereich.

7.4.8 Das ILM Framework der SNIA

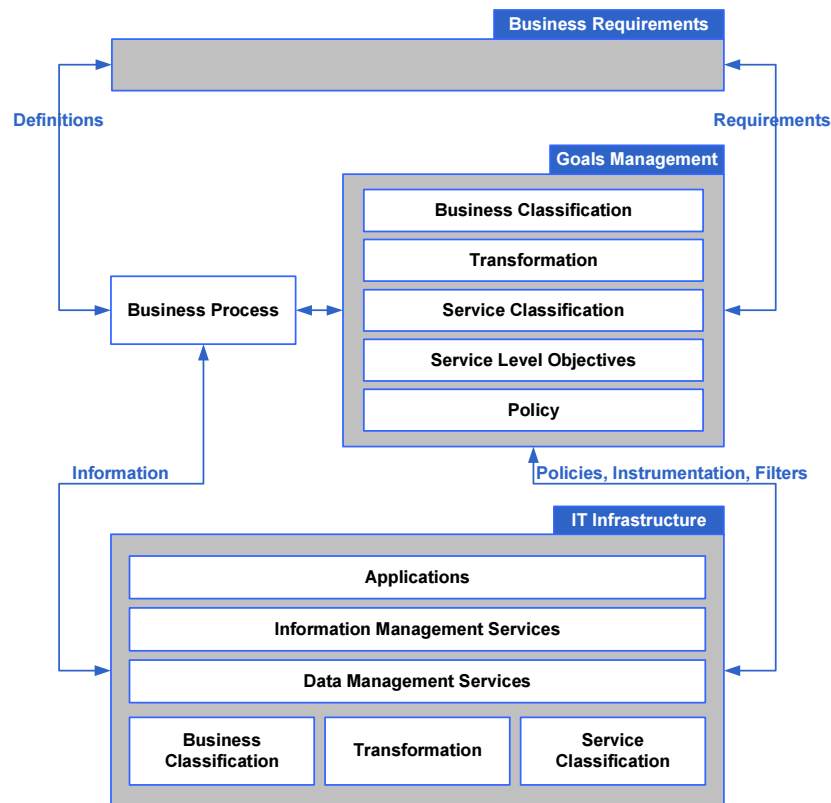


Abbildung 7.27 Das ILM Framework der SNIA

Das ILM Framework besteht aus den drei Bestandteilen Business Requirements, Goals Management und IT Infrastructure und einer Reihe von Prozessen und technischen Hilfsmitteln wie in **Abbildung 7.27** dargestellt [SNIA 2004]:

- *Business Requirements*: Die Geschäftsanforderungen an Information Lifecycle Management werden in Abläufe (Business Process) und die Verwaltung von Umsetzungszielen (Goal Management) abgebildet.

- *Goal Management*: Die Verwaltung der Umsetzung der Geschäftsanforderungen erfolgt im Goal Management durch eine Abstraktion der Anforderungen, die als Policies der IT Infrastruktur zur Umsetzung übergeben werden.
- *IT Infrastructure*: Die Infrastruktur besteht aus allen physischen und logischen Komponenten der betrieblichen Informatik und setzt die Anforderungen um.

Zentrales Element des ILM Framework ist das Goal Management welches in mehreren Schritten erfolgt. Die *Business Classification* erfasst Anforderungen wie beispielsweise „Patientendaten müssen gesetzeskonform aufbewahrt und gesichert werden“ in geordneter Form. Die *Transformation* kombiniert die erfassten Anforderungen mit den vom *Business Process* kommenden kundenspezifischen Abläufen zu generalisierten Anforderungen. Diese Anforderungen werden wiederum mittels einer abstrakten *Service Classification* auf Dienstklassen abgebildet. Aus diesen Dienst-Klassen werden Service Level Objectives abgeleitet und als RPO / RTO / DPW Bedingungen definiert

7.5 Fragen zum Kapitel

Nr	Frage
1	Welche Bewertung von Informationen finden Sie am vernünftigsten? Begründen Sie Ihre Auswahl.
2	In welchem Bereich der betriebliche Informatik spielen die Kennwerte Recovery Point Objective und Recovery Time Objective eine Rolle?
3	Was ist der Besitzumseffekt und bei welcher Informationsbewertung spielt er eine Rolle?
4	Was ist genau Verfügbarkeit?
5	Was ist der Unterschied zwischen SMS und HSM?
6	Welche Rechnertechnologie arbeitet mit System Managed Storage und was ist der Vorteil dieser Technik?
7	Wenn Sie die Daten Ihres Unternehmen klassieren müssten, welche Art von Klassierung würden Sie vorziehen? Begründen Sie Ihre Auswahl.
8	Die gesetzeskonforme Archivierung von eMail Daten ist ein grosses Problem. Erklären Sie warum?
9	Ein Unternehmen führt eine Information Logistics Governance ein. Welches Vorgehen zur Realisierung der ILM Anforderungen würden Sie empfehlen?
10	Welche Bedingungen sollten erfüllt sein, dass ein ILM durchgeführt werden kann?

7.6 Übung zum Kapitel

Ein Hersteller von Heizungskörpern möchte die Indikatoren für den Wert seiner Unternehmensinformationen aufgezeigt haben. Wählen Sie die Klassierung und berechnen Sie den Indikator für einen Data Protection Event. Der Hersteller hat weltweit 30'000 Angestellte. Was schätzen Sie kostet ein Data Protection Event?