

**Fraunhofer-Institut
für Naturwissenschaftlich-Technische Trendanalysen (INT)**

Appelsgarten 2, 53879 Euskirchen, Telefon: 02251 18-0, Fax: 02251 18-277, E-Mail: info@int.fraunhofer.de

**„Data Management“
in der Fraunhofer-Gesellschaft**
-
**Empfehlungen für die
IT-Strategie-Roadmap**

Joachim Burbiel, Miloš Jovanović, Silke Römer

INT-Bericht 206

Euskirchen, Oktober 2008

Zusammenfassung

„Data Management“ beinhaltet die Gesamtheit aller Verfahren um mit der steigenden Datenflut sinnvoll umzugehen. Im Auftrag der IT-Strategiegruppe der Fraunhofer-Gesellschaft wurde eine Untersuchung zu diesem Thema durchgeführt. Diese Untersuchung ist die erste von drei Phasen eines Prozesses an dessen Ende eine übergreifende Strategie zum „Data Management“ für die Fraunhofer-Gesellschaft steht.

Im Rahmen der Studie wurden zwei Ansätze parallel verfolgt:

Zum einen wurden die technologischen Entwicklungen auf dem Gebiet „Data Management“ betrachtet. Als Einstieg hierfür wurde der so genannte *Gartner-Hype-Cycle* verwendet. Ausgehend von den dort genannten Konzepten und Technologien wurde versucht, einen Überblick, sowohl über das Themenfeld, als auch über aktuelle Entwicklungen zu erhalten.

Zum anderen wurden stichprobenartig Interviews mit Mitarbeitern der FhG durchgeführt, um festzustellen, welche Methoden und Praktiken des Data Management bereits angewandt werden. Außerdem sollten Erfahrungen der Mitarbeiter mit einbezogen werden, um Bedürfnisse und mögliche Probleme und Defizite im Bereich Data Management zu identifizieren.

Die dabei gewonnenen Erkenntnisse wurden zu einem „Ist“-Zustand der FhG zusammengefasst. Anschließend wurde eine Auswahl an Szenarien erstellt, um die Auswirkungen von verschiedenen Maßnahmen abzuschätzen. Abschließend wurden die Resultate der Szenarien bewertet, um zu Empfehlungen und Projektvorschlägen zu gelangen. Hierbei wurden wiederum die Erkenntnisse der aktuellen Entwicklung von Konzepten und Technologien berücksichtigt.

Im Kern der Empfehlungen stehen verstärkte Anstrengungen zentraler Stellen, lokal vorhandenes Wissen über Methoden des Data Management besser zu nutzen. Dadurch sollen Effizienzverluste, die derzeit durch Mehrfachbearbeitung von sehr ähnlichen Problemen entstehen, in Zukunft vermieden werden. Außerdem wird die Wichtigkeit von Schulungen hervorgehoben.

Inhalt

1	Einleitung	2
1.1	Themenfeld „Data Management“	2
1.2	Vorgehensweise	2
2	Fraunhofer-spezifische Anforderungsanalyse	4
3	Technologiebewertung	7
3.1	Grundansätze des Data Management	7
3.1.1	Föederal oder zentral?	7
3.1.2	Formatvielfalt oder Standardformat?	9
3.2	Erweiterung und Modifizierung des Technologiefeldes	11
3.2.1	Teilfeld 1: Übergeordnete Begriffe	13
3.2.2	Teilfeld 2: Data Integration und Data Quality	19
3.2.3	Teilfeld 3: Datenbank-Management und OSS	23
4	Erfassung der Anwenderperspektive: Befragung von Mitarbeitern verschiedener Abteilungen des INT	26
4.1	Frage 1: Begriffsdefinition/bekannte Methoden	26
4.2	Frage 2: Erfahrungen	27
4.3	Frage 3: Verbesserungsvorschläge	28
4.4	Zusammenfassung	29
5	Gezielte Horizonterweiterung: Befragung ausgewählter Mitarbeiter an anderen Fraunhofer-Instituten	31
6	Ist-Zustand: Data Management in der FhG	34
7	Bewertung anhand ausgewählter Szenarien	43
7.1	Szenario 1: „Status Quo“	43
7.2	Szenario 2: „Information“	44
7.3	Szenario 3: „Föderale Struktur“	45
7.4	Szenario 4: „Zentrale Struktur“	47
7.5	Vergleich der Szenarien	48
8	Zusammenfassung und Empfehlung	50
9	Anhänge	55
9.1	Fragebögen	55
9.2	Abkürzungen	58
9.3	Abbildungen	59
9.4	Kurzbiografien der Autoren	60

1 Einleitung

1.1 Themenfeld „Data Management“

Der Begriff „Data Management“ beschreibt ein großes Gebiet. Es umfasst alle technischen und organisatorischen Maßnahmen, um Daten strukturiert zu erfassen, zu verwalten und verfügbar zu machen. Es ist außerdem ein sehr dynamisches Gebiet, da es stark von Entwicklungen im Bereich der Informations- und Kommunikationstechnologie (ICT) beeinflusst wird.

In der vorliegenden Studie zum Themenfeld „Data Management“ sollen die aktuellen und zukünftigen Entwicklungen von Data Management-Konzepten und -Technologien, deren Abbildung auf Anforderungen der Fraunhofer-Gesellschaft (FhG) und Möglichkeiten zur Realisierung aufgearbeitet werden.

Die Autoren dieser Studie verstehen sich ausdrücklich nicht als IT-Fachleute (siehe auch Kurzbiographien in Kapitel 9.4). Dies führt eventuell dazu, dass in dieser Studie einigen technisch anspruchsvollen und organisatorisch komplexen Teilgebieten des „Data Management“ verhältnismäßig wenig Aufmerksamkeit zuteil wird. Dies ist durchaus gewollt, da die Erarbeitung von pragmatischen Vorschlägen, die den täglichen Bedürfnissen von „normalen“ Mitarbeitern der FhG entgegenkommen, im Fokus unserer Aufmerksamkeit steht.

1.2 Vorgehensweise

Es wurden zwei Ansätze parallel verfolgt:

Zum einen wurden die **technologischen Entwicklungen auf dem Gebiet „Data Management“** betrachtet. Als Einstieg hierfür wurde der so genannte *Gartner-Hype-Cycle* verwendet. Hierbei handelt es sich um eine Technologieeinschätzung der Beratungsfirma *Gartner*. Ausgehend von den dort genannten Konzepten und Technologien wurde versucht, einen Überblick, sowohl über das Themenfeld, als auch über aktuelle Entwicklungen zu erhalten. Hierfür waren sowohl Modifizierung als auch Erweiterung der Inhalte des *Gartner-Hype-Cycles* notwendig. Diese Aktivität wird im Kapitel 3 beschrieben. Spezifische Untersuchungen der Entwicklungen bei Dienstleistungsangeboten wurden nicht durchgeführt. Dies war nicht notwendig, da die Anwenderperspektive im Mittelpunkt stand.

Zum anderen wurden **stichprobenartig Interviews mit Mitarbeitern der FhG** durchgeführt, um festzustellen, welche Methoden und Praktiken des Data

Managements bereits angewandt werden. Außerdem war es uns wichtig, Erfahrungen der Mitarbeiter mit einzubeziehen, um Bedürfnisse und mögliche Probleme und Defizite im Bereich Data Management zu identifizieren. Dieser Teil der Studie ist in den Kapiteln 4 und 5 dokumentiert.

Die dabei gewonnenen Erkenntnisse wurden zu einem Ist-Zustand der FhG synthetisiert. Anschließend wurde eine Auswahl an Szenarien erstellt, um die Auswirkungen von verschiedenen Maßnahmen abzuschätzen. Die Resultate der Szenarien wurden bewertet, um zu Empfehlungen zu gelangen. Hierbei wurden wiederum die Erkenntnisse der aktuellen Entwicklung von Konzepten und Technologien berücksichtigt.

2 Fraunhofer-spezifische Anforderungsanalyse

In der Anfangsphase des Projekts wurde die Arbeitshypothese aufgestellt, dass es innerhalb der Fraunhofer-Gesellschaft drei Typen von Forschungseinrichtungen gibt, die sich im Hinblick auf ihren Umgang mit Daten grundsätzlich unterscheiden.

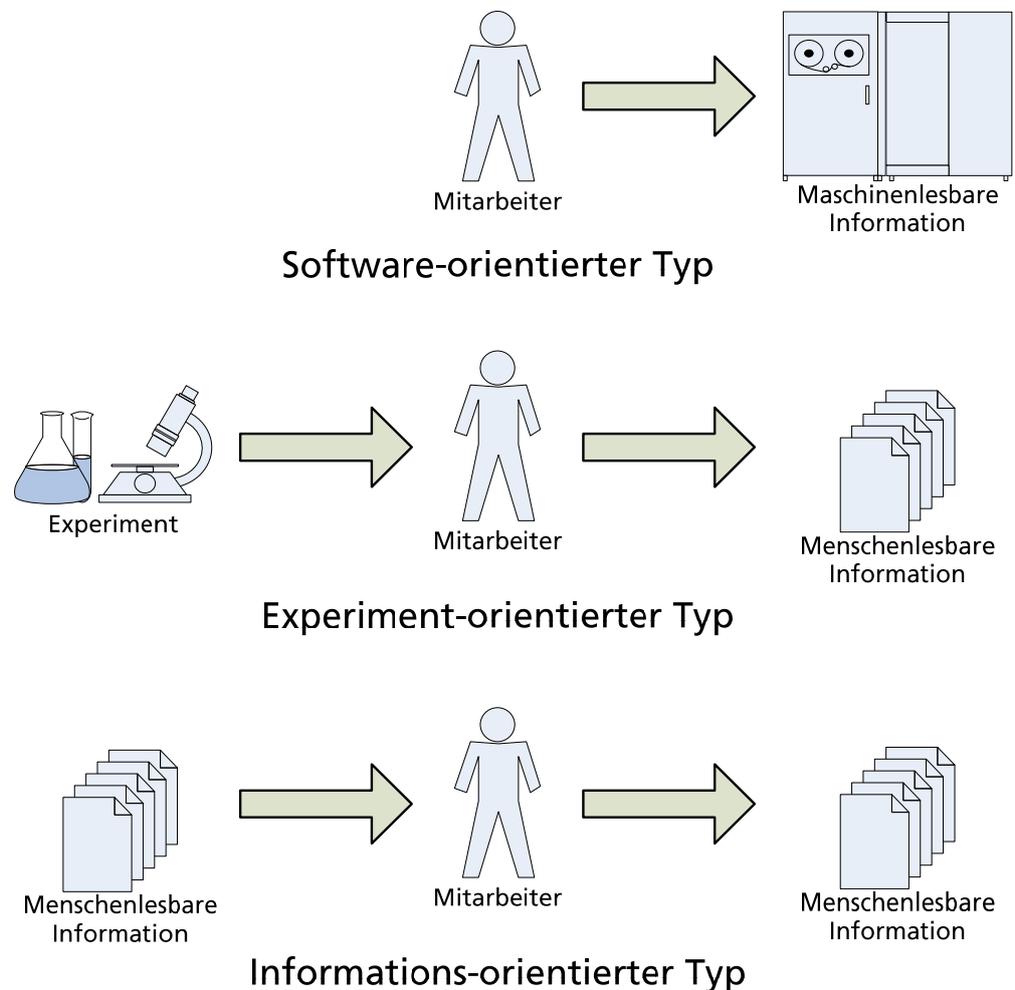


Abbildung 1: Schematische Darstellung der drei Typen wissenschaftlicher Arbeit

Den ersten Typus haben wir **Software-orientiert** genannt. Hier kommen Daten vor allem als Programme, Routinen und Module vor. Das Besondere an diesen Daten ist, dass sie durch eine geistige Leistung erzeugt wurden und in ihrer Endform selbst das Produkt darstellen. Die Herausforderung in Bezug auf Data Management besteht für solche Bereiche vor allem darin, kollaborative Arbeit zu organisieren und Versionskonflikte zu minimieren. Die Daten fallen dabei naturgemäß in maschinen-freundlicher Form (z. B. Quellcode) an.

Beim zweiten Typus, dem **Experiment-orientierten**, wird vor allem mit maschinell erzeugten Daten umgegangen, wie sie etwa als Rohdaten in einem Experiment anfallen. In Hinblick auf die Anforderungen an das Data Management ist es dabei irrelevant, ob die Daten physischen Sensoren oder virtuellen Anordnungen bzw. Simulationen entstammen. Die Daten müssen in jedem Fall aufgearbeitet und interpretiert werden, was einen großen Teil der geistigen Leistung in Bezug auf Daten ausmacht. Andere geistige Leistungen, wie Versuchsdesign und -aufbau, gehen meist mit geringeren Datenmengen einher. Die Herausforderung in Bezug auf Data Management besteht für solche Bereiche vor allem darin, Daten, die häufig in proprietären Formaten anfallen, sinnvoll zu archivieren und zu vereinheitlichen um sie miteinander in Beziehung setzen zu können. Außerdem ist bei vielen experimentellen Anordnungen die schiere Masse an gemessenen Daten ein Problem.

Den dritten Typus bezeichnen wir als **Informations-orientiert**. Damit ist gemeint, dass Informationen von externen Quellen bezogen, interpretiert und in einen neuen Zusammenhang gebracht werden. Beispiele für solche Bereiche sind das Fraunhofer ISI oder die nicht-experimentellen Bereiche des Fraunhofer INT, aber auch die Arbeit in den beiden oben genannten Typen haben Informations-orientierte Anteile, z.B. das Lesen eines Buches über Software-entwicklung oder das Studium von Patenten bezüglich experimenteller Aufbauten. Die bearbeiteten Daten liegen zumeist in maschinen-unfreundlicher Form vor (Bilder, Texte). Die Herausforderung in Bezug auf Data Management besteht für solche Bereiche vor allem darin, die Mensch-Maschine-Schranke zu überwinden, damit die Leistung von Computern genutzt werden kann, um menschliche Schwächen (z.B. Vergesslichkeit oder unscharfe Datenspeicherung) auszugleichen.

Auch wenn gelegentlich ganze Institute einem einzelnen Typ zuzuordnen sind, kommen doch häufig Mischformen vor. So können z.B. Arbeitsbereiche, die mit Simulationen umgehen, einem der beiden ersten Typen zugeordnet werden, je nachdem, ob sie primär Simulationsprogramme erstellen oder solche benutzen. Wir sind aber der Meinung, dass das Schema so angelegt ist, dass sich jede Arbeitsgruppe oder Abteilung einem der drei Bereiche zuordnen kann.

Wir haben im Laufe dieser Untersuchung Personen aus allen drei Institutstypen befragt, wobei wir uns den Umstand zu Nutze gemacht haben, dass am

Fraunhofer INT alle drei Bereiche vertreten sind: Ein kleiner Teil des Instituts arbeitet Software-orientiert (z.B. bei der Programmierung von Datenbanken oder bei der Durchführung von Simulationen), ein Teil der Abteilung „Nukleare und Elektromagnetische Effekte“ arbeitet experimentell, die zwei Abteilungen „Technologieanalysen und –vorausschau“ und „Übergreifende Analysen und Planungsunterstützung“ gehen Informations-orientiert vor. Um die Unterrepräsentation beim Software-orientierten Typ auszugleichen, wurden bei der Befragung von Mitarbeitern weiterer Fraunhofer-Institute bewusst solche aus diesem Bereich ausgewählt.

Der **Bereich Verwaltung** unterscheidet sich grundsätzlich von den bisher genannten. Hier erfolgt Datenverarbeitung in weit stärker schematisierter Form: Sowohl die Datenerhebung (über Formulare, Standardbriefe), als auch die Datenverarbeitung (mittels des Programmpakets SIGMA) und die Ausgabe (standardisierte Weitergabe) laufen wenig flexibel ab, was sicherlich auch sinnvoll ist. Auch das Bibliothekswesen hat einen Sonderstatus. Wir werden am Rande auf die spezielle Situation dieser Bereiche eingehen.

3 Technologiebewertung

3.1 Grundansätze des Data Management

Innerhalb des Data Management existieren zwei Grundansätze zum Umgang mit Daten, die wir im Folgenden mit den Begriffen *Data Federation* und *Data Integration* bezeichnen werden (in der Fachliteratur finden sich für die zwei Ansätze sehr unterschiedliche Bezeichnungen, wie überhaupt die Nomenklatur in diesem Bereich extrem unscharf ist). In diesem Fall geht es um zentrale oder dezentrale Speicherung von Daten und die Art des Zugriffs auf die Daten durch verschiedene Nutzer.

In diesem Zusammenhang gibt es zwei verschiedene Ansätze, mit der tendenziell wachsenden Zahl an Datenformaten umzugehen: einerseits Konzepte und Technologien, um mit einer hohen Variabilität von (Daten-) Formaten umzugehen, andererseits Bemühungen, in vielen Bereichen Standardformate zu etablieren, die im Idealfall die Anzahl an benötigten Formaten reduzieren sollen.

3.1.1 Föderal oder zentral?

Unter **Data Federation** (Abb. 2) verstehen wir alle Ansätze, bei denen ein Tool (häufig als Portal bezeichnet) genutzt wird um eine einheitliche Darstellung der Daten zu gewährleisten, obwohl die zugrundeliegenden Daten an unterschiedlichen Speicherorten, und möglicherweise auch in verschiedenen Formaten, existieren. Dabei werden die ursprünglichen Daten nicht angetastet, sondern nur abgebildet und der Besitzer dieser Daten entscheidet, wer wann auf "seine" Daten zugreifen kann.

Ein einfaches Beispiel ist ein Desktop-Suchtool, das es ermöglicht über eine Eingabemaske in einer Vielzahl von Ordnern und Formaten zu suchen. Im Allgemeinen greift ein Portal auf Daten zu, die bereits in irgendeiner Weise verwaltet werden. Dies wird in der Regel so funktionieren, dass das Portal, durch eine Anfrage aktiviert, die Anfrage an die einzelnen Datenbanken weitergibt (in einer für die einzelnen Datenbanken verständlichen Form). Die zurück erhaltenen Antworten werden so aufbereitet, dass sich eine „virtuelle Gesamtschau“ ergibt.

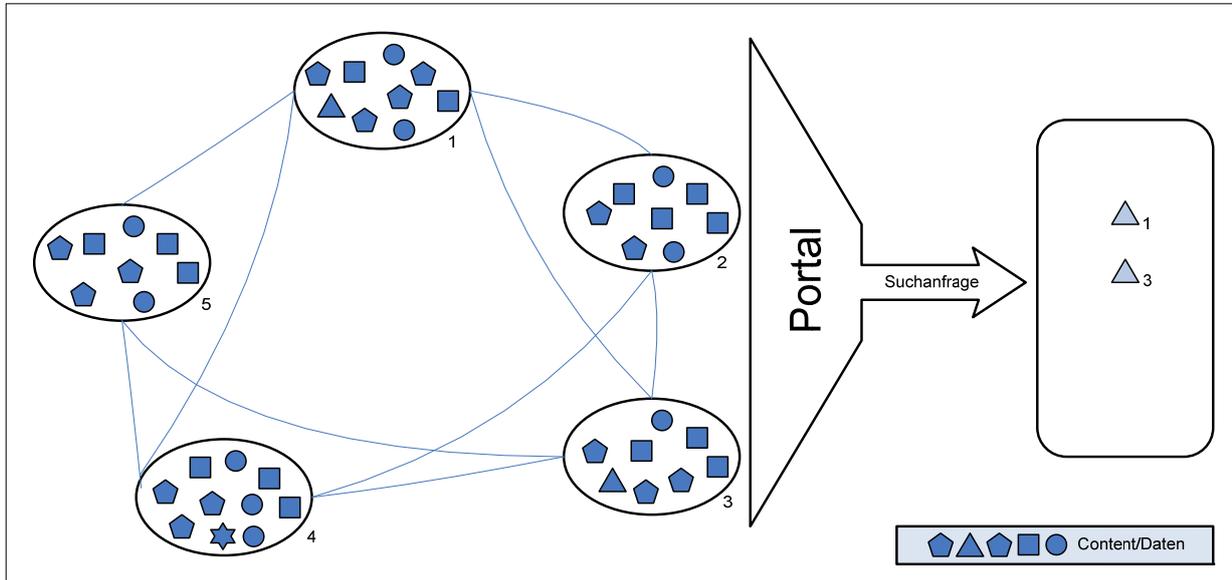


Abbildung 2: Schema zur Data Federation; Daten werden dezentral gespeichert und können via Portal vereinheitlicht angezeigt oder – wie hier dargestellt – durchsucht werden.

Dieses Konzept toleriert auch die Einführung eines neuen Datenverarbeitungssystems in einem Teilbereich, ohne dass hierfür die Architektur des Gesamtkonzepts modifiziert werden muss. Damit dies funktioniert, muss das Portal ausreichend flexibel sein. Durch die „virtuelle Gesamtschau“ wird der Nutzer dann idealerweise gar nicht bemerken, dass ein Teilbereich verändert wurde.

Ein Nachteil der Data Federation ist, dass die Qualität der eingeholten Daten schwer überprüfbar ist. Wenn verschiedene Quellen unterschiedliche Antworten auf die gleiche Frage liefern, ist es für den Nutzer nicht eindeutig, welche Informationen autoritativ sind (das klassische Google-Problem). Dies macht die Anwendung eines solchen Ansatzes für kritische Daten zweifelhaft.

Unter **Data Integration** (Abb.3) verstehen wir alle Ansätze, bei denen der Versuch unternommen wird, alle Daten in einer autoritativen Datenbank abzulegen, auf die der Nutzer dann zugreifen kann. Hierbei können zwei Unteransätze unterschieden werden: Der weiche, bei dem die ursprünglichen Daten unverändert bleiben, und der harte Ansatz, bei denen die einzelnen "Unterdatenbanken" mit der zentralen Datenbank abgeglichen werden. Der Nachteil der weichen Variante ist, dass sich eventuell Widersprüche einschleichen und unterschiedliche Geschäftsbereiche mit unterschiedlichen

Daten arbeiten. Der Nachteil der harten Variante ist, dass falsche Informationen schlechter erkannt werden können, da sie, wenn sie einmal in die autoritative Datenbank gekommen sind, plötzlich überall als maßgeblich auftauchen.

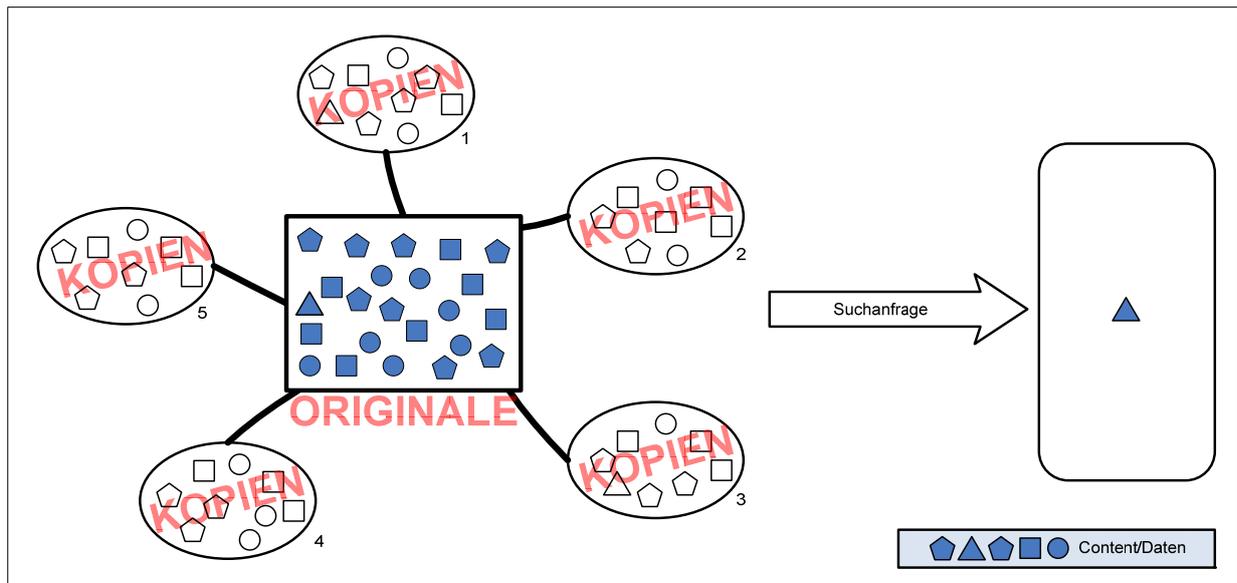


Abbildung 3: Schema zur Data Integration mit zentraler autoritativer Datenbank; weitere Erläuterungen im Text..

Ein Nachteil der Data Integration ist ihre relative Starrheit, da jede Veränderung an der autoritativen Datenbank weitreichende und möglicherweise unvorhersehbare Folgen hat.

In der **realen Welt** wird man meist Kombinationen der beiden Ansätze finden, z.B. föderative Ansätze, die die Inhalte von lokal autoritativen Datenbanken zusammentragen.

3.1.2 Formatvielfalt oder Standardformat?

Neben der Unterscheidung zwischen föderalem und zentralem Data Management stellt sich die Frage, ob Formatvielfalt toleriert wird oder eine Standardisierung angestrebt wird.

Tolerierte Formatvielfalt tritt meist im Zusammenhang mit Data Federation auf: Zum Beispiel wenn Daten in einem Portal unabhängig vom eigentlichen Speicherort einheitlich präsentiert werden, wobei Inhalte aus den verschiedenen Formaten extrahiert und dargestellt werden, ohne die

ursprünglichen Daten zu verändern. Klassische Beispiele sind auch hier wieder die Ausgabemaske einer Desktop-Suchmaschine oder ein Internetportal.

Hierbei ist anzumerken, dass die einheitliche Präsentation aller bekannten Formate – auch in Teilbereichen, beispielsweise Text, Bild oder Ton – utopisch und aufgrund der existierenden Vielfalt teils inkompatibler Datenformate, proprietärer oder offener, nur unvollständig möglich ist.

Das andere Extrem ist das Anstreben eines **vereinheitlichten Formats**, das die bestehende Formatvielfalt ersetzt. Auch dies wird nicht komplett realisierbar sein. Zum einen stellen ständig ändernde Hard- und Software immer neue Anforderungen an ein solches Einheitsformat, die bei der Erstellung eines solchen nur unvollständig absehbar sind. Zum anderen erschwert die Vielzahl unterschiedlicher bereits etablierter Formate eine allgemeine Verbreitung eines neuen Einheitsformates, da dies auch eine Umstellung vorhandener Routinen mit sich bringt.

In der Praxis finden sich dennoch Tendenzen zur Vereinheitlichung, insbesondere gekoppelt mit Data Integration, weil hier eine Festlegung auf wenige Datenformate, die vorausschauend ausgewählt sein wollen, enorme Vorteile hat und durch die zugrundeliegende Infrastruktur leichter zu realisieren ist als bei föderalem Data Management.

Data Integration mit Einschränkung der Formatvielfalt lässt sich am besten dort realisieren, wo auch sonst organisatorisch zentral gehandelt werden kann. Dies ist beispielsweise bei Unternehmen in den Bereichen Verwaltungsvorgänge, Buchführung etc. der Fall. Bei Kooperationen mit externen Partnern ist es dagegen notwendig auch andere Datenformate zu verarbeiten und zu managen.

Reine Format-Konverter sind weder ausreichend noch immer in angemessener Form vorhanden, so dass Data Integration nicht 100% des Data Management in den Bereichen ausmachen kann, in denen Daten mit externen Partnern ausgetauscht werden müssen.

3.2 Erweiterung und Modifizierung des Technologiefeldes

Im „*Hype Cycle for Data Management, 2007*“ werden 21 Technologien aufgeführt (vgl. Abb. 4). Sie sind sortiert nach der Phase öffentlicher Wahrnehmung / Sichtbarkeit, in der sich die Technologie nach der Einschätzung der Firma *Gartner* befindet.

Wir haben diese Technologien als Ausgangspunkt für diese Studie genommen und nach erster inhaltlicher Prüfung eine der 21 Technologien ausgeschlossen, andererseits aber auch einige Technologien bzw. Konzepte hinzugefügt.

immer nachvollziehbare Veränderung, sowohl was die Inhalte, als auch die Position der Begriffe auf dem Hype-Cycle betrifft.

Auch die unterschiedliche Granularität der aufgeführten Begriffe war problematisch. So finden sich konkrete Anzeige- (z. B. Data Quality Dashboard) Dabei erwies sich die zeitliche Instabilität der entsprechenden Hype-Cycles als problematisch. Wie in Abbildung 4 dargestellt, gibt es eine starke und nicht und Suchtools (z.B. XQuery) neben übergreifenden strategischen Ansätzen (z. B. Enterprise Information Management).

„Real-Time Data Integration“ wurde ausgeklammert, da wir im Blick auf die Interessen der Fraunhofer-Gesellschaft zwar die Relevanz von „Data Integration“ nicht anzweifeln, „Real-Time Data Integration“ jedoch eine hochspezialisierte Variante darstellt. Sie vereint die Echtzeitdatenverarbeitung mit der Echtzeit-Integration dieser Daten in mehrere relevante, ggf. unternehmensweite Anwendungen. Ein Nutzen ergibt sich wohl nur in speziellen Unternehmensumgebungen, z. B. bei Call-Centern (Konsolidierung von Kundendaten in Echtzeit), in der Telekommunikationsindustrie oder im Einzelhandel, wenn es um sehr eng verzahnte Lieferkettenabläufe geht, nicht jedoch bei den typischen Aufgaben von Forschungsinstituten.

Der Grund zur Aufnahme zusätzlicher Technologien ist, dass von *Gartner* zum einen Technologien mehr oder weniger gleichrangig aufgeführt werden, die auf sehr unterschiedlichen (Abstraktions-) Ebenen operieren oder aber unter Anwendungsgesichtspunkten "künstlich" getrennt werden (OSS DBMS for mission-critical/non-mission-critical applications).

Darüber hinaus blieben nach unserer Ansicht noch Lücken auf dem Gebiet Data Management, was auf zwei Ursachen zurückzuführen ist:

- Zum einen lässt sich das Themenfeld „Data Management“ nicht eindeutig von verwandten Themen abgrenzen, da es keine allgemeingültige, ausreichend genaue Definition gibt.
- Zum anderen werden bei der Erstellung des Gartner-Hype-Cycle sämtliche Bereiche des Technologiefeldes ausgeblendet, die bereits ausgereift sind und das "Plateau der Produktivität" (fast) erreicht haben.

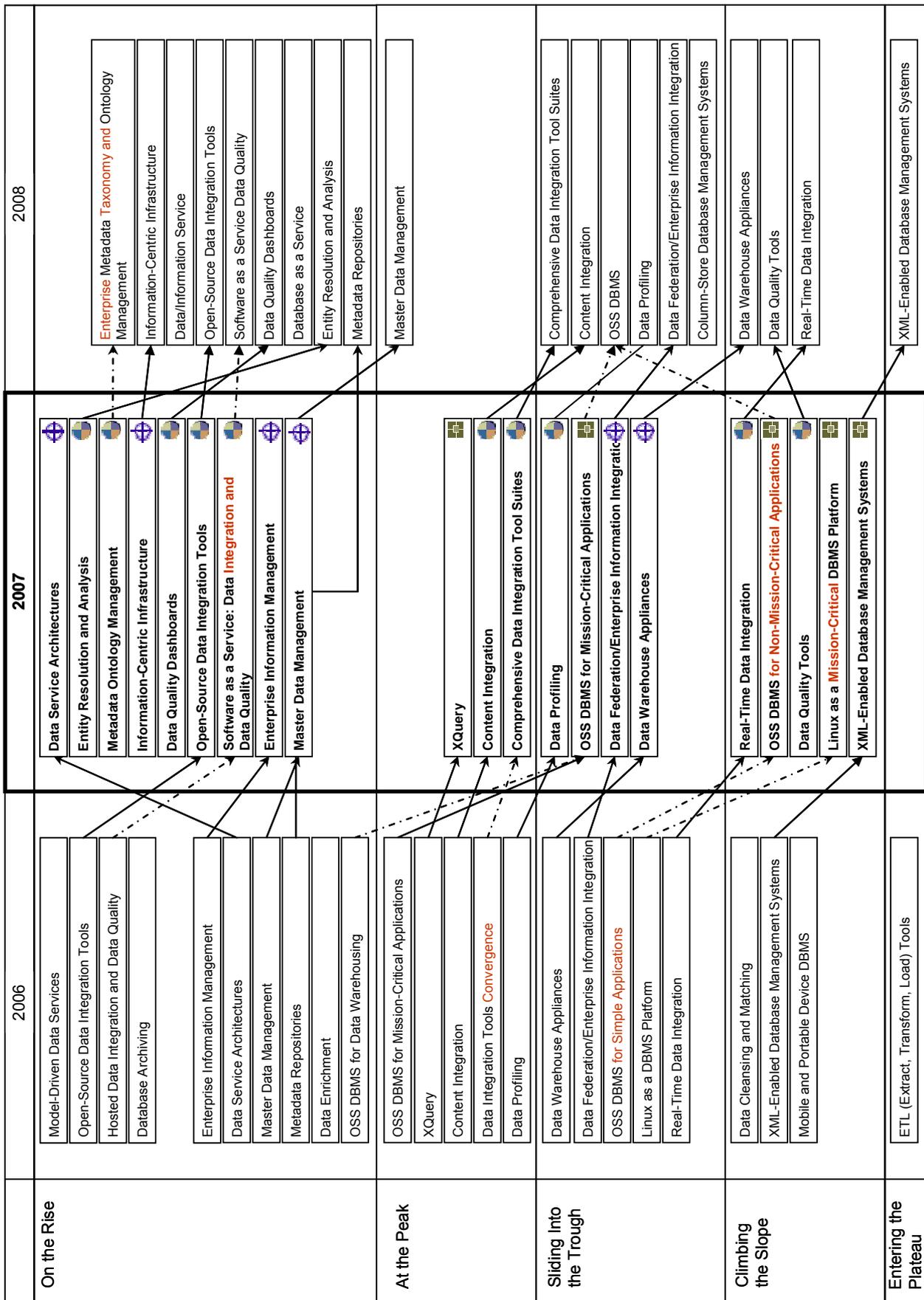


Abbildung 4: Technologien im Gartner-Hype-Cycle „Data Management“ aus den Jahren 2006-2008 inklusive Illustration der durch uns vorgenommene Unterteilung des Themenfeldes in
 + Übergeordnete Begriffe Data Integration und Data Quality Datenbank-Management
 DBMS: Data base management system; OSS: Open source software; weitere Erläuterungen im Text.

Wir haben in der weiteren Betrachtung die Grenzen der Definition des Themenfelds „Data Management“ dem Ziel angepasst, Empfehlungen für eine allgemeine Strategie für die Fraunhofer-Gesellschaft bezüglich des Themenfeldes aussprechen zu können.

3.2.1 Teilfeld 1: Übergeordnete Begriffe

Einige Technologien des Gartner-Hype-Cycles „Data Management“ wurden in der Kategorie „Übergeordnete Begriffe“ zusammengefasst:

- Enterprise Information Management (EIM)
- Information-Centric Infrastructure (ICI)
- Data Service Architecture
- Data Federation / Enterprise Information Integration (EII)
- Master Data Management (MDM)
- Data Warehouse Appliances

Bei der Beschäftigung mit diesen Begriffen wurde festgestellt, dass weitere Begriffe aus der Welt des Datenmanagements notwendig sind, um ein verständliches Bild zu erhalten. Dabei handelt es sich um:

- Business Intelligence (BI)
- Knowledge & Workflow Management
- Enterprise Content Management (ECM)
- Web Content Management (WCM)

Im Folgenden sollen die einzelnen Begriffe erläutert werden. Dabei wird mit dem Umfassendsten begonnen, speziellere Begriffe schließen sich an.

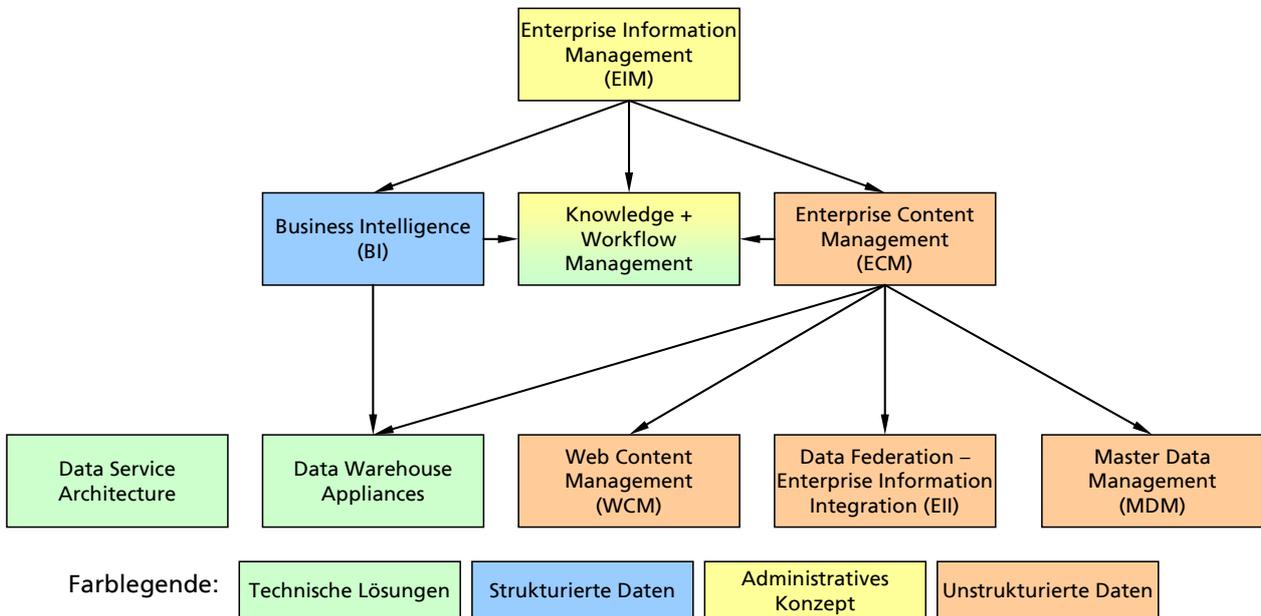


Abbildung 5: Graphische Darstellung der Zusammenhänge in der Kategorie „Übergeordnete Begriffe“

Enterprise Information Management (EIM)

Position on Hype-Cycle: *On the Rise*

Benefit Rating: *High*

EIM ist ein administrativer Ansatz um mit Informationen umzugehen. Es bedeutet, dass man sich an zentraler Stelle und mit den nötigen Mitteln um Belange des Datenmanagements kümmert. Beim EIM geht es, systematisch gesehen, um den Umgang mit Business Intelligence (BI, strukturierte Daten) und Enterprise Content Management (ECM, unstrukturierte Daten). Außerdem beinhaltet EIM noch die Aspekte Sicherheit (inkl. Rechteverwaltung), Daten-Analyse und Business Process Management (BPM). Ein Hauptanbieter ist Microsoft mit Sharepoint 2007.

Information-Centric Infrastructure (ICI)

Position on Hype-Cycle: *On the Rise*

Benefit Rating: *Transformational*

Bei der ICI werden Daten aus den einzelnen Anwendungsprogrammen herausgeholt um die Bildung von voneinander unabhängigen „Datensilos“ zu vermeiden. Im Mittelpunkt steht die Information, nicht mehr das Anwendungs-

programm. Diese „reine“ Information soll bedeutend effektiver genutzt werden können. Ein Problem bei der Behandlung dieses Themas ist, dass der Begriff „Information-Centric Infrastructure“ von Gartner geprägt wurde und auch nur im Zusammenhang mit Gartner auftaucht. Es ist daher schwer zu sagen, was genau damit gemeint ist und wie er von anderen Konzepten abgegrenzt werden kann. In Abgrenzung zu Data Federation (s. u.) scheint es jedoch die physikalische Veränderung der Quelldaten einzuschließen. Eventuell ließe sich sagen, dass ICI eine Form von EIM ist.

Business Intelligence (BI)

Nicht auf dem Hype-Cycle „Data Management“

Bei „Business Intelligence“ handelt es sich um die Gesamtheit aller in strukturierter Form vorhandenen Daten und dem Umgang damit. Hiermit sind z.B. Daten in hoch strukturierten Datenbanksystemen gemeint. In der Fraunhofer-Gesellschaft ist SIGMA das zentrale System in dessen Rahmen BI stattfindet. Im Folgenden wird BI daher nur noch im Zusammenhang mit SIGMA Erwähnung finden.

Enterprise Content Management (ECM)

Nicht auf dem Hype-Cycle „Data Management“

Unter Enterprise Content Management (ECM) versteht man den unternehmensweiten Umgang mit unstrukturierten Daten. Bei Daten in unstrukturierter Form handelt es sich vor allem um menschenlesbare Texte, aber auch um Bilder, Tondokumente, oder Kombinationen daraus (z.B. Webseiten). Von den reinen Daten wird dabei begrifflich „Content“ abgesetzt, der aus den Daten (Inhalten) und den dazu gehörenden Metadaten besteht.

Maschinen haben es naturgemäß schwerer mit unstrukturierten Daten umzugehen als mit strukturierten, gerade im wissenschaftlichen Umfeld (Informations-orientierter Typ) stehen jedoch unstrukturierte Daten im Vordergrund des Interesses.

ECM beinhaltet die *Erfassung, Verwaltung, Speicherung, Bewahrung und Bereitstellung* von solchen Daten und Dokumenten. ECM ist dabei eher ein Konzept als ein Programm oder eine Sammlung von Programmen. Es schließt Web-Content-Management (WCM, s. u.) mit ein, der Schwerpunkt liegt jedoch bei der Inhouse-Informationsbereitstellung. Weitere zentrale Aspekte von ECM sind Data Federation / Enterprise Information Integration (EII, s. u.) und Master Data Management (MDM, s. u.).

Workflow Management und Knowledge Management

Nicht auf dem Hype-Cycle „Data Management“

Die beiden Techniken Workflow Management und Knowledge Management stehen an der Schnittstelle zwischen EIM, BI und ECM. Beim Workflow Management geht es darum Arbeitsprozesse geordnet durchzuführen und nachvollziehbar zu dokumentieren. Beim Knowledge Management steht implizites und explizites Wissen im Mittelpunkt. Auch die Dokumentation abgeschlossener Projekte gehört hierzu. Diese beiden, an sich sehr wichtigen, Themen werden im Weiteren nicht mehr eigens behandelt, da sie nicht in das engere Feld Data Management fallen, waren aber für die Darstellung in Abbildung 5 wichtig.

Data Warehouse Appliances

Position on Hype-Cycle: *Sliding Into the Trough* Benefit Rating: *High*

Diese Technologie beschreibt Komplettpakete für die Einrichtung von Data Warehouses, bestehend aus Hardware, Betriebssystem, DBMS Software und Speicherung. Sie bedient sich, technisch gesehen, des *massive parallel-processing* (MPP) und ist häufig aus einzelnen, voll funktionsfähigen Servern aufgebaut (*shared nothing architecture*). Klassische Anbieter sind Teradata (incl. des aufgekauften Pionier-Unternehmens Britton-Lee) und Netezza.

Den Data Warehouse Appliances werden wir im Weiteren keine Aufmerksamkeit mehr widmen, da es sich um einen sehr technischen Aspekt des Datenmanagements handelt. Als Fazit lässt sich sagen, dass dieses Item wenig aufregend, aber recht nützlich sein könnte.

Web Content Management (WCM)

Nicht auf dem Hype-Cycle „Data Management“

Web Content Management (WCM) ist der bekannteste Unteraspect des ECM. Es geht hierbei darum, einen Internetauftritt auf der Basis einer dahinter liegenden Datenbank aktuell zu halten. WCM wurde wegen seiner Bekanntheit in diese Liste aufgenommen, im Folgenden jedoch nicht weiter behandelt.

Data Federation / Enterprise Information Integration (EII)

Position on Hype-Cycle: *Sliding Into the Trough* Benefit Rating: *Moderate*

Bei Data Federation / Enterprise Information Integration (EII) werden Daten aus verschiedenen Quellen (Anwendungen) gemeinsam und einheitlich dargestellt

(*uniform data access / uniform information representation*). Es kann also als Möglichkeit gesehen werden, kreativ mit einer Vielfalt von uneinheitlichen Datenquellen umzugehen. Das Besondere dabei ist, dass die ursprünglichen Datenquellen nicht verändert werden.

Typischerweise erfolgt dieser Prozess in drei Stufen:

- 1 Ausführung verteilter Suchanfragen an diverse Quellen
- 2 Zusammenfassung der Suchergebnisse als „virtuelle Gesamtschau“
- 3 Weitergabe dieser Gesamtschau an andere Anwendungen, die das Ergebnis weiterverarbeiten oder darstellen.

Die Tatsache, dass diese Methode es erleichtert mit einer Vielzahl von verschiedenen Systemen und Datenformaten vernünftig umzugehen, lässt sie als besonders attraktiv für ein „föderales“ System, wie die Fraunhofer-Gesellschaft, erscheinen. Allerdings scheint es noch einige technische Hürden bis zur einfachen Anwendbarkeit zu geben.

Master Data Management (MDM)

Position on Hype-Cycle: *On the Rise*

Benefit Rating: *High*

Master Data Management (MDM, deutsch: „Zentrale Stammdatenverteilung“) ist eine Gruppe von Methoden und Prozessen, um die Richtigkeit, Vollständigkeit, Aktualität und Konsistenz der wichtigsten Geschäftsdaten über verschiedene Anwendungen, Systeme, Datenbanken, Prozesse, Abteilungen, Organisationen und Niederlassungen hinweg sicherzustellen.¹

Data Service Architecture

Position on Hype-Cycle: *On the Rise*

Benefit Rating: *Transformational*

Technischer Ansatz um Datenspeicherung, Sicherheitsaspekte und Arten des Zugriffs auf Daten voneinander zu entkoppeln. Als Funktionen einer guten *Data Service Architecture* werden genannt:

- Synchronisation von Datennetzen mit den „Master-Daten“

¹ "MDM is a set of disciplines and processes for ensuring the accuracy, completeness, timeliness and consistency of the most important types (or domains) of enterprise data across different applications, systems and databases, multiple business processes, functional areas, organizations, geographies and channels." Dan Power, „The Politics of Master Data Management & Data Governance“, DM Review Magazine, March 2008, http://www.dmreview.com/issues/2007_45/10000894-1.html

- Präsentation hoch-qualitativer kanonischer Daten innerhalb einer *messaging infrastructure*
- Darstellung von Kontrollpunkten um BI und Data-Warehouse-Systeme als lose angebundene Systeme zu steuern

Es handelt sich wiederum um einen sehr technischen Aspekt, der sicher seine Bedeutung hat, jedoch nicht in den Kernbereich unserer Betrachtung fällt.

3.2.2 Teilfeld 2: Data Integration und Data Quality

Die Technologien dieses Technologiefeldes beschäftigen sich primär mit der Integration bestehender Daten in einen Data Pool oder ein Data Warehouse. Die Bewertung der Datenqualität ist damit eng verbunden.

Die verschiedenen Technologien lassen sich bildlich folgendermaßen darstellen:

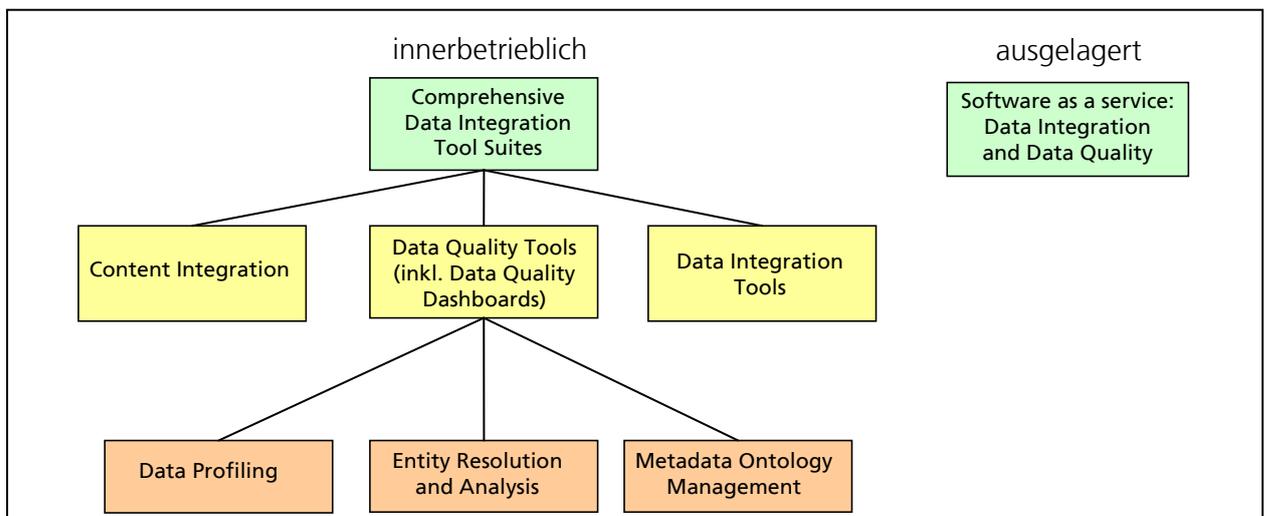


Abbildung 6: Graphische Darstellung der Zusammenhänge in der Kategorie „Data Integration und Data Quality“

Comprehensive Data Integration Tool Suites

Position on Hype-Cycle: *At the peak*

Benefit Rating: *High*

„Comprehensive Data Integration Tool Suites“ sind Kombinationen verschiedener Softwareprodukte, die zu einem „Satz“ zusammengefasst werden. Zentraler Bestandteil ist auch hier die Integration von Daten, also das Extrahieren und Aufbereiten von Daten aus verschiedenen heterogenen Quellen. Anschließend werden die Daten dann in Unternehmenseigene Datenbanken/-warehouses geladen (Extract-Transform-Load Prozess, kurz ELT). Die „Suites“ umfassen auch Tools zum Data Profiling, der Beurteilung von Data Quality und zur zentralen Stammdatenverwaltung.

Content Integration

Position on Hype-Cycle: *At the Peak*

Benefit Rating: *High*

„Content Integration“ beschreibt die Integration von unstrukturierten Inhalten, die sich in verschiedenen, voneinander isolierten „Datensilos“ befinden. Dieser Inhalt (Content) kann dabei stark diversifiziert sein: Bilder, Dokumente, Videos, Tabellen etc. Das Problem ist hierbei also, dass Content möglicherweise doppelt in verschiedenen Silos vorhanden ist, oder dass Mitarbeiter, die sich nur mit einem Silo beschäftigen, ihren Content in einem anderen Silo finden könnten, ihnen diese aber nicht bekannt sind. Um Content innerhalb eines Unternehmens integrieren zu können, müssen zwei Punkte erfüllt werden:

Die Verantwortung für die Contentpools muss klar verteilt sein.
Die Contentpools müssen gemanagt werden.

Mit dem Managen von Content ist gemeint, dass die Unternehmensleitung eine vereinheitlichte Beschreibung von Content durchsetzen muss, unabhängig davon, in welchen Silos die Daten liegen. Diese Beschreibung kann über Metadaten geschehen oder mit Hilfe von Tools. Ein anderes Beispiel ist die Integration von Unternehmens-Content in E-Learning Kursen, damit die Mitarbeiter an konkreten Beispielen des Unternehmens lernen können. Es geht bei dieser Technologie also allgemein um das sinnvolle Vernetzen von Contents, die an verschiedenen Orten im Unternehmen gespeichert sind. Man beachte, dass „Content“ hier von „Data“ unterschieden wird. Hier geht es also um bereits bearbeitete Dokumente bzw. Informationen, nicht um rohe Daten.

„Content Integration“ als Begriff findet sich primär in Texten und Beschreibungen von Gartner selber, im Internet gibt es andere Begriffe wie „Enterprise Content Integration“, die ähnliche Technologien beschreiben.

Eine konkrete Einsatzmöglichkeit für Content Integration kann in der Harmonisierung von Content zwischen gegenwärtigen und zukünftigen Fraunhofer-Instituten bestehen. Aktuelles Beispiel hierfür wäre der mögliche Beitritt der FGAN-Institute zur FhG.

Suchmaschinen ersetzen Content Integration nicht, ergänzen es aber enorm, da mit ihnen Content gefunden werden kann, der vorher nicht bekannt war. Dieser neu gefundene Content kann dann per Content Integration wieder sinnvoll mit Tags, Metadaten, Anchors oder anderem versehen werden.

Taxonomien werden bei Gartner nicht als Teil der Content Integration gesehen. Wir schließen uns dieser Meinung nicht an, da das Einteilen von Content innerhalb einer Taxonomie dazu führen kann, dass die Daten stärker und systemweit integriert sind.

(Open Source) Data Integration Tools

Position on Hype-Cycle: *On the Rise*

Benefit Rating: *Low*

„Data Integration Tools“ sind Tools zur Integration von Daten (Extraktion und Aufbereiten von Daten aus verschiedenen heterogenen Quellen). In Abgrenzung zum Begriff Content Integration ist hier die Integration von Rohdaten gemeint.

Beim Gartner Hype-Cycle Data Management 2007 werden explizit nur „Open Source Data Integration Tools“ aufgeführt, also nur solche Programme betrachtet, deren Source Code veröffentlicht wurde.

Der Gartner-Bericht gibt eine Marktpenetration von 1% an für diese Art von Tools. Dies wird damit begründet, dass Open-Source Programme meist nur standardisierte Schnittstellen bieten (z. B. Java Database Connectivity), oder sich auf XML-Transformation beschränken. Solche Standards seien aber in den meisten Fällen unzureichend, in denen Data Integration Tools genutzt werden. In der Industrie gibt es jedoch auch Beispiele dafür, dass kommerzielle Anbieter ihre Software auf das Open-Source Modell umstellen, mit der Begründung dass diesem Modell die Zukunft gehört.

Data Quality Tools (incl. Data Quality Dashboard)

Position on Hype-Cycle: *Climbing the Slope*

Benefit Rating: *High*

Data Quality Tools sollen helfen, die Qualität der Daten/Informationen auf deren Grundlage Entscheidungen getroffen werden, zu prüfen. Genutzt wird dies hauptsächlich im Bereich „Business Intelligence (BI)“ (siehe oben).

Für „Datenqualität“ gibt es mehrere, teilweise widersprüchliche, Definitionen, zum Beispiel dass ein Datensatz von hoher Qualität ist, wenn er das reale Objekt, auf das er sich bezieht, korrekt abbildet.

Diese Datenqualität wird durch Datenverluste verringert, die bei der Integration von Daten und Content auftreten können oder aber schlichtweg im Laufe der Zeit entstehen (z. B. neue Telefonnummern oder die Einführung eines neuen Postleitzahlensystems). Data Quality Tools müssen daher regelmäßig aktualisiert werden, um auch solche Verluste schnell und automatisch korrigieren zu können.

Ein „Data Quality Dashboard“ visualisiert die Ergebnisse der Data Quality Tools mit Hilfe von einfachen Punktekarten oder Ampeln. Der Zweck dieser Visualisierung ist die einfachere Analyse des Datenbestandes unter dem Aspekt der Datenqualität. So kann untersucht werden in welchem Bereich der

Datenbank die Datenqualität mangelhaft ist, oder welche Datensätze regelmäßig korrigiert werden müssen.

Data Profiling

Position on Hype-Cycle: *Sliding into the Trough* Benefit Rating: *Moderate*

Das „Data Profiling“ bewegt sich in der Nähe der „Data Quality Tools“ und überschneidet sich mit diesen auch teilweise, weshalb viele Anbieter von Data Quality Tools deshalb auch Data Profiling Tools anbieten.

Beim reinen Data Profiling geht es um die Auswertung der Daten an sich und deren Nutzung, um auf dieser Grundlage zu entscheiden, ob sie für den Profiler nützlich sind oder nicht. Beispiele für (Teil-)Fragen, die mit Hilfe des Data Profiling beantwortet werden können, sind:

- Um was für Daten handelt es sich?
- Wurden Metadaten korrekt den Datensätzen zugeordnet?
- Wie viele Daten welchen Typs wurden abgerufen? (Nutzungsstatistiken)

SaaS Data Integration and Data Quality

Position on Hype-Cycle: *On the Rise* Benefit Rating: *Moderate*

SaaS („Software as a Service“) Dienste bezeichnen eine Möglichkeit für ein Unternehmen Aufgaben an eine andere Firma auszulagern (Outsourcing). In diesem Falle würde ein Unternehmen die Aufgaben der Data Integration und Data Quality an eine externe Firma auslagern. Für den Bereich Data Integration sind bisher nur wenige SaaS Lösungen bekannt,² da die Daten bei jedem Unternehmen sehr unterschiedlich und die Integration dieser Daten stets ein individueller Prozess ist. Bei der Data Quality gibt es bereits einige Anbieter, hier allerdings nur in speziellen Fällen (z. B. bei der Verifikation von Kundenadressen).

² <http://blogs.ittoolbox.com/bi/websphere/archives/informatica-news-here-comes-saas-and-on-demand-data-integration-9551>
Abgerufen: 20. 5. 2008 15:25 Uhr

Metadata Ontology Management

Position on Hype-Cycle: *On the Rise*

Benefit Rating: *High*

Meta-Daten (v. a. in Form von Anmerkungen/Annotations) sind wichtig, da sie die Beurteilbarkeit der Datenqualität verbessern. Metadata Ontology Management vereint alle verfügbaren (heterogenen) Metadaten einer Person, eines Vorgangs oder einem Objekt zu einem auswertbaren Datensatz.

Entity Resolution and Analysis

Position on Hype-Cycle: *On the Rise*

Benefit Rating: *High*

Es geht bei Entity Resolution and Analysis darum, alle Einzeldaten, die zu einer Person, einem Vorgang oder einem Objekt existieren, automatisch als zusammengehörig zu erkennen und zu einem einheitlichen Datensatz zusammenzufassen. Dieser ist dann unter allen Namen, Pseudonymen und Synonymen auffindbar. Dieses Verfahren stammt aus der Forensik und dient dort zur Identifikation von kriminellen Einzelpersonen und Gruppen.

3.2.3 Teilfeld 3: Datenbank-Management und OSS

Der Einsatz von Open Source Software – nicht nur im Bereich Data Management – bietet eine Reihe von Vorteilen gegenüber kommerzieller Software. So ist sie zumeist kostenlos und unterliegt keinen Nutzungsbeschränkungen, der Quelltext ist offengelegt und für Änderungs- und Erweiterungswünsche ist man nicht an einen Softwarehersteller gebunden.

Dem gegenüber steht der mögliche Nachteil, dass es in der Regel keinen Hersteller gibt, der sich verpflichtet, dem Anwender der Software im Falle auftretender Probleme zu helfen. Meistens gibt es eine Gemeinschaft von Nutzern und Entwicklern, die diese Support-Funktion wahrnehmen; dies ist oft aber nicht in dem Sinne möglich, wie es beim Software-Einsatz in Unternehmen erforderlich ist.

Im Gartner-Hype-Cycle wird nicht zuletzt aus diesem Grund zwischen dem Einsatz von OSS in *non-mission-critical applications* und *mission-critical applications* unterschieden. Dies geschieht aber nicht konsequent, wie anhand des Beispiels OSS DBMS im Gartner-Hype-Cycle abzulesen ist: 2007 wird noch unterschieden, 2008 taucht nur noch der zusammengefasste Begriff OSS DBMS im Hype-Cycle auf (vgl. Abbildung 4).

XML-enabled database management

Position on Hype-Cycle: *Climbing the slope* Benefit Rating: *Moderate*

Unterschieden werden XML-Datenbanken wie folgt:

- XML-enabled:
Herkömmliche Datenbanksysteme, die ein Mapping auf oder ins XML-Format erlauben. Man bezeichnet diese Vorgehensweise als datenorientiert.
- Native XML-Datenbanksysteme:
Diese Systeme speichern die Information, ähnlich wie bei der Speicherung von XML-Dokumenten im Dateisystem, direkt als XML-Dokumente ab. Sie werden daher auch als dokumentenorientiert bezeichnet.

Die im Gartner-Hype-Cycle angegebene Technologie könnte evtl. schon „Off The Hype-Cycle“ sein.

XQuery

Position on Hype-Cycle: *At the peak* Benefit Rating: *Moderate*

XQuery (XML Query Language) ist eine vom W3C spezifizierte Abfragesprache für XML-Datenbanken. Sie scheint besonders geeignet, z. B. aus großen XML-Datensammlungen einzelne Teile herauszusuchen.

XQuery hat das Potential, den Umgang mit einer Vielzahl verschiedener Informationstypen zu vereinheitlichen oder sogar die Informationstypen selbst (z. B. relationale Daten und XML). Dabei steht und fällt dies mit der Implementierung des noch recht neuen XQuery-Standards in den tatsächlich genutzten Datenmanagementsystemen.

OSS DBMS for (non-)mission-critical applications

Position on Hype-Cycle: *Climbing the slope (non-mission-critical applications)*
Sliding into the Trough (mission-critical applications)

Benefit Rating: *Moderate*

DBMS steht für Datenbankmanagementsystem und bezeichnet die Software, mit der die eigentliche Datenbank verwaltet wird; OSS DBMS ist die Open Source Software auf diesem Gebiet.

Das DBMS soll eingeflochten sein in Applikationen, die für (Teil-)Projekte gebraucht werden, also beispielsweise ein Abfragefenster einer Datenbank innerhalb einer anderen Applikation.

Unterschieden wird von Gartner zwischen dem Einsatz von OSS DBMS für "mission-critical applications" und „non-mission-critical applications“. Gemeint ist hier die Einbindung/Nutzung von OSS DBMS in (nicht) kritischen Bereichen der Infrastruktur.

Linux as a mission-critical DBMS-platform

Position on Hype-Cycle: *Climbing the Slope* Benefit Rating: *High*

Die Überschrift ist Programm: Linux als Plattform hat auch in diesem Einsatzgebiet neben finanziellen Vorteilen weitere Vorzüge, die im Einzelfall z. B. die Flexibilität betreffen und kann auch in kritischen Bereichen der Infrastruktur eingesetzt werden.

4 Erfassung der Anwenderperspektive: Befragung von Mitarbeitern verschiedener Abteilungen des INT

Die zweite wichtige Informationsquelle unserer Studie war die Befragung von Fraunhofer Mitarbeitern. Da am Fraunhofer INT alle drei Typen wissenschaftlicher Arbeit (vgl. Abbildung 1) vertreten sind wurden zunächst ausgewählte Mitarbeiter dieses Institutes befragt.

Der von uns erstellte Fragebogen ist in drei Fragenkomplexe unterteilt: Begriffsdefinition/bekanntes Methoden, Erfahrungen und Verbesserungsvorschlägen (Version 1.1, siehe Anhang).

Die Fragen waren bewusst allgemein gehalten und auf Antworten in freiem Textformat ausgelegt. Bei der Befragung wurde vom Interviewer die Antwort in Stichworten oder ganzen Sätzen festgehalten (bei Bedarf nach abschließenden Rückfragen im Fall von Unklarheiten).

4.1 Frage 1: Begriffsdefinition/bekanntes Methoden

Was verstehen Sie unter „Data Management“?

Als Definition wurde am häufigsten „Organisation von Wissen und Informationen“ genannt (4 Nennungen). Dabei wurde besonders die Notwendigkeit betont, Informationen schnell und selektiv zu finden bzw. wieder zu finden (3 Nennungen).

Welche Methoden (oder Tools) werden in Ihrem Bereich bereits angewandt?

Bei den genutzten Systemen standen die Ablage in Papierform (Ordner in Schränken, 6 Nennungen) und die elektronische Ablage in Dateiordnerstrukturen (10 Nennungen) eindeutig im Vordergrund. Einzelne Mitarbeiter gaben an, fremde Dokumente gar nicht mehr zu archivieren, sondern sich praktisch nur noch auf Internet-Suchtools (Google wurde genannt) zu verlassen.

Bei den konkret genannten Programmen zum lokalen Datenmanagement kann zwischen verschiedenen Gruppen unterschieden werden:

a) Desktop-Suchtools

Das Desktop-Suchtool ISYS wurde siebenmal genannt. Diese hohe Rate kann damit zusammenhängen, dass dieses Programm zum Zeitpunkt der Befragung gerade am INT eingeführt wurde und damit den Befragten sehr präsent war. Andere Desktop-Suchtools (Google Desktop, xfriend) wurden vereinzelt genannt.

b) Datenbanken

Diverse Datenbanksysteme wurden genannt, sowohl externe Systeme (z. T. Nutzung über das Internet), als auch interne Systeme, die im Laufe der Institutsgeschichte entstanden sind (insgesamt 9 Nennungen). Hier wurde auch SIGMA mehrfach genannt, das bei den Mitarbeitern der Verwaltung eine große Rolle spielt (2 Nennungen).

c) Produkte aus dem Microsoft-Office-Paket

Konkret wurden von einzelnen Befragten Excel, Word und Outlook genannt (insgesamt 5 Nennungen).

d) Literaturverwaltungsprogramme

Hier wurden gelegentlich Endnote, Citavi (für beide gibt es Fraunhofer-Gesellschaft-Lizenzen) und Zotero (freie Software) genannt.

Bekannte Systeme zur Unterstützung des Data Managements, die jedoch nicht eingesetzt werden:

Bei der Software, die bekannt war, jedoch nicht genutzt wurde, gab es keine auffälligen Häufungen. Genannt wurden z.B. LDAP (Lightweight Directory Access Protocol), Adressverwaltungstools, Spezialprogramme für die Auswertung experimenteller Daten, Software von SAP und DATEV, sowie MindMap/MindManager.

4.2 Frage 2: Erfahrungen

Welche positiven oder negativen Erfahrungen haben Sie beim „Data Management“ bereits gemacht? (Lessons learned)

Hier ist zunächst einmal festzuhalten, dass individuell sehr unterschiedlicher Umgang mit Daten zu einer Vielzahl differierender Meinungen führte.

Positiv wurden im Allgemeinen Konzepte beurteilt, bei denen Daten in elektronischer Form übersichtlich abgelegt, leicht wieder auffindbar und mit anderen teilbar sind. Im Einzelnen wurden beispielsweise Datenbanken, die in der Abteilung/Projektgruppe gemeinsam genutzt werden, das institutseigene Intranet sowie das Fraunhofer Intranet und SIGMA für den Bereich Verwaltung genannt. Insbesondere im Fall Intranet wurde die Nützlichkeit und Notwendigkeit der Suchfunktion hervorgehoben.

Mehrfach wurden die Möglichkeiten verglichen, Daten in elektronischer Form oder in Papierform zu speichern. Negative Erfahrungen mit der elektronischen Ablage umfassten die Problematik der Unübersichtlichkeit bei wachsendem Datenaufkommen und notwendiger Strukturierung sowie die Kompatibilität der Daten bei im Laufe der Zeit wechselnden Programmen und Systemen.

Des Weiteren wurde angeführt, dass beim Data Management Disziplin verlangt ist (4 Nennungen), wobei Skepsis herrscht, ob der Mehraufwand zum einen von allen Beteiligten akzeptiert wird und, zum anderen, ob er in vernünftigen Verhältnis zum gewonnenen Nutzen steht.

4.3 Frage 3: Verbesserungsvorschläge

Wo sehen Sie Defizite? Welche Probleme müssen noch gelöst werden?

Die Skala der Antworten reichte hier von der Meinung, dass keine (weiteren) speziellen Data Management-Lösungen gefragt sind (2 Nennungen) bis zur Einstellung, dass der Bedarf an schnelleren, flexibleren Data Management Lösungen wächst. Daneben wurden für unterschiedliche Bereiche ungelöste Probleme aufgeführt:

- Probleme bei der Strukturierung vorhandener Daten inklusive Umstrukturierung (4 Nennungen)
- Unflexible Datenablage (3 Nennungen)
- Überkomplexe Programme (2 Nennungen)
- Veraltende Softwarelösungen (2 Nennungen)
- Unnötige Formatvielfalt (2 Nennungen)

Welche Methoden (oder Tools) fehlen Ihnen?

Hier gab es vor allem Einzelnennungen:

- Desktop-Suchtool
- Bessere Suchalgorithmen für das Internet
- Methode für übersichtliche und komplette Backups
- E-Mail Archivierung

- Funktionierende Adressdatenbank
- Literaturverwaltung
- Wikis
- Synchronisation von Daten
- Datenaustausch mit externen Kooperationspartnern (Probleme mit ftp-Servern)
- Verstärkter Einsatz von Linux und Open-Source-Software

Wie kann Ihnen die FhG, zentral oder auf Institutsebene helfen?

Hier wurde mehrfach betont, dass oft individuelle Lösungen (instituts- oder abteilungsspezifisch) gefragt sind und für den Problembereich eine zentrale Lösung nicht sinnvoll erscheint. Hierfür spricht auch, dass manche in einem Teilbereich wie der Literaturverwaltung mit vorhandenen Lösungen zufrieden sind, andere wiederum nicht.

Eine Standardisierung auf Abteilungsebene wurde von Einzelnen als empfehlenswert beurteilt.

Weitere Vorschläge waren:

- Ansprechpartner in der FhG oder am Institut für Standardprogramme, Tipps & Tricks, Antworten auf gezielte Fragen (3 Nennungen)
- Institutsebene: Schulungskonzept und Informationsweitergabe für vorhandene Tools (3 Nennungen)
- Zentrale Datenbank für hochwertige Artikel (zentral oder auf Institutsebene) und Experten wünschenswert (2 Nennungen)
- evtl. via Instituts-Bibliotheken: Zugang zu internen (Konferenz-)Berichten verbessern (2 Nennungen)
- evtl. Software(weiter)entwicklung zur Anpassung an Bedürfnisse

4.4 Zusammenfassung

Insgesamt lassen sich die mehrfach genannten Anforderungen an ein Data Management Konzept wie folgt zusammenfassen und abstrahieren:

- Übersichtlichkeit der Daten
- Kompatibilität der Daten und Datenstruktur muss gewährleistet sein, sowohl im Hinblick auf den Datenaustausch mit anderen Programmen als auch für den Fall der Migration zu einem anderen System
→ *möglichst nicht-proprietäre Dateiformate*

- Möglichkeit der flexiblen nachträglichen Umstrukturierung von Datenbeständen
- Archivierungslösung (z. B. zufriedenstellende Archivierung von E-Mails)
- Teilen der Information mit anderen Mitarbeitern der Fraunhofer-Gesellschaft sowie Projektpartnern
- Geringer bzw. angemessener zusätzlicher Arbeitsaufwand
- Angemessene Komplexität des Programms (Software weist oft Funktionen auf, die zwar praktisch sind, aber selten genutzt werden, da sie in einer Optionsvielfalt untergehen) / Software mit klar umgrenztem Einsatzbereich stellt eine Arbeitserleichterung dar.
Außerdem ist es wünschenswert, dass die Grundfunktionalität intuitiv nutzbar ist, so dass je nach Bedarf eine stufenweise Erarbeitung ihrer tiefer greifenden Funktionalität optional möglich ist.
- Hohe Toleranz gegen Abweichungen von allgemeinen Regeln (z. B. die Dateiablage betreffend)
Es werden dennoch allgemeingültige Regeln für die Ablage benötigt, vielleicht nicht für die gesamte Fraunhofer-Gesellschaft, aber auf Instituts-/Abteilungsebene oder auf Projektebene; Akzeptanz und entsprechende Konsequenz in der Anwendung dieser Regeln bleiben wahrscheinlich maßgeblich.

5 Gezielte Horizonterweiterung: Befragung ausgewählter Mitarbeiter an anderen Fraunhofer-Instituten

Zur Ergänzung und Verifizierung der Ergebnisse der Umfragen im INT wurden fünf weitere Interviews mit Mitarbeitern anderer Institute durchgeführt. Um ein auf die FhG passendes Gesamtbild zu erhalten wurden dabei Software-orientierte Institute bevorzugt, da dieser Bereich innerhalb des INT unterproportional vertreten ist. Bei der Befragung wurde ein leicht modifizierter Fragebogen verwendet (Version 1.2, s. Anhang).

Aggregation der Antworten:

Frage 1a. Was verstehen Sie unter „Data Management“?

Der Grundtenor war „strukturierte Art und Weise Daten abzulegen und wiederzufinden“. In einem Interview wurde gesagt, dass „unsortiertes Sammeln und Verlassen auf ein Such-Tool“ kein Data Management sei.

Frage 1b. Welche Methoden (oder Tools) werden in Ihrem Bereich angewandt?

Die elektronische Ablage in Dateiodnerstrukturen (5 Nennungen) stand hier ebenso wie bei der internen Befragung im Vordergrund. Drei Interviewpartner sprachen von gemeinschaftlich genutzten „Containern“. Interessanterweise wurde die Ablage in Papierform, die im INT häufig genannt wurde, von unseren externen Interviewpartnern nie angegeben. Möglicherweise spielt sie in Software-orientierten Arbeitsumgebungen nur eine untergeordnete Rolle.

Bei den konkret genannten Programmen zum lokalen Datenmanagement kann hier ebenfalls zwischen verschiedenen Gruppen unterschieden werden:

1. Versionierungstools

Wie zu erwarten war, spielen im Software-orientierten Bereich Softwareverwaltungs- und -versionierungstools eine große Rolle (4 Nennungen). Konkret wurden Subversion, Git und CVS genannt.

2. Wikis

Wikis werden ebenfalls genutzt, konkret wurde die Zusammenarbeit in den jeweiligen Abteilungen als Einsatzzweck angegeben (2 Nennungen).

3. Andere Programme

Des Weiteren wurden CAS als Programm zur Akquise-Verwaltung, BibTex als Literaturverwaltungsprogramm und nicht näher spezifizierte Kommandozeilentools (z.B. Suchtools) genannt.

Frage 1c. Welche weiteren Methoden (oder Tools) kennen Sie?

Genannt wurden: Livelink, MS-Access (für Fotos) und Google Desktop. Außerdem gibt es wohl bereits ein System, bei dem Institute Kundenanfragen zum Thema „Bilderkennung“ einstellen können, für deren Bearbeitung sie selbst nicht die nötige Expertise haben.

Frage 2: Welche positiven Erfahrungen haben Sie beim „Data Management“ bereits gemacht?

Alle Befragten äußerten sich positiv zum Stand der Technik was Softwareentwicklungstools und Versionierungsprogramme betrifft. Diese scheinen ausgereift zu sein. Auch Wikis wurden positiv bewertet, allerdings wurde angemerkt, dass es notwendig ist, gewisse Grundregeln festzulegen und Administratoren zu benennen, die die Einhaltung überwachen. Suchtools (Nennung von Google und Google Desktop) wurden ebenfalls als praktisch unverzichtbar genannt. An dieser Stelle wurde auch BSCW genannt, das für die Verschiebung und den gemeinsamen Zugriff auf große Datenmengen gut geeignet sei.

Frage 3a: Wo sehen Sie Defizite? Welche Probleme müssen noch gelöst werden?

Während individuelle Ablagesysteme meist als ausreichend funktionierend angesehen wurden, entstehen häufig Probleme bei der Zusammenarbeit. Jeweils zweimal wurden Probleme beim gemeinsamen Betrieb von „Containern“ und bei individuellen Vorlieben in Bezug auf „Lieblingstools“ genannt. Als weitere Probleme wurden Konflikte beim gleichzeitigen Betrieb von mehreren Systemen (konkret: Container und Wiki zur Dateiablage) und die Einschränkung durch unflexible Tools benannt.

Frage 3b: Welche Methoden (oder Tools) fehlen Ihnen?

Unter diesem Punkt wurden genannt:

- Desktopsuchmaschinen (2 Nennungen)
- Plattformunabhängiger, gemeinsam nutzbarer Kalender
- Fotokatalogisierungssoftware für den Bereich Marketing

- Automatische Sortiermaschine mit Vorschlägen bei Ablage
- Tool zur (Neu-)Strukturierung von Daten bei laufendem Projekt

Frage 3c: Wie kann Ihnen die FhG, zentral oder auf Institutsebene helfen?

Diverse Interviewpartner betonten, dass es wohl keine allgemeingültige Lösung für Datenmanagement-Probleme gäbe. Außerdem wurde davor gewarnt, sich von technischen Lösungen zu viel zu versprechen, da es am Ende immer auf das Verhalten des einzelnen Mitarbeiters ankomme.

In Bezug auf die FhG wurde der Wunsch geäußert, dass die Zusammenarbeit der Institute besser werden sollte. Dazu wären ein Werkzeug, das Informationen über andere Institute liefert und eine Art FhG-weites Sourceforge hilfreich.

Organisatorisch wird eine Stelle gewünscht, die zentral Data Management-Tools sichtet und bewertet bzw. diese Information aus den Instituten sammelt und auch als Ansprechpartner für diesen Themenbereich zur Verfügung steht.

6 Ist-Zustand: Data Management in der FhG

Der momentane Umgang mit Daten innerhalb der Fraunhofer Gesellschaft (FhG) besteht aus einer Ansammlung unterschiedlichster Methoden. Ein Grund hierfür liegt in dem hohen Maß an Selbstständigkeit, welche die Forschungseinrichtungen der FhG *de facto* genießen.

Bei der Befragung am Fraunhofer INT und von ausgewählten Mitarbeitern anderer Institute kamen interessante und teils überraschende Ergebnisse zu Tage (siehe Kapitel 4 und 5). Bei der Frage nach der Definition des Begriffs „Data Management“ waren sich die meisten Mitarbeiter einig: gemeint ist das sinnvolle Organisieren von Daten und Dokumenten. Zusätzlich wurde die Notwendigkeit betont, diese schnell wieder auffinden zu können. Der Unterschied zwischen Daten- und Dokumentenmanagement, der in der Literatur gefunden werden kann, ist für diese Studie nicht von Belang, da die Grenzen nicht nur fließend sind, sondern die beiden Begriffe einen hohen Überlapp haben.

Als erstes überraschendes Ergebnis der Befragung trat zutage, dass viele der befragten Mitarbeiter sich noch auf klassische Systeme - wie zum Beispiel Papierablagen oder elektronische Ordner - verlassen. Vereinzelt wurde auch darauf hingewiesen, dass fremde Daten gar nicht mehr archiviert werden. Stattdessen verlässt man sich auf Internet-Suchmaschinen (Google wurde häufig genannt), mit denen Daten zu einem späteren Zeitpunkt wieder gefunden werden können.

Spezielle Systeme des Data Management wurden auffallend selten genannt. Stattdessen nutzen die Mitarbeiter häufig das Microsoft Office Paket oder Desktopsuchmaschinen. Bei letzteren ist zu beachten, dass erst vor kurzem eine kommerziell erhältliche Software für die Desktopsuche am INT eingeführt wurde. Dies könnte der Grund für die häufigen Nennungen von Seiten der INT-Mitarbeiter sein.

Speziellere Anwendungen wie Datenbanksysteme, Wikis oder Literaturverwaltungsprogramme wurden ebenfalls, wenn auch seltener, erwähnt. Wikis werden von einigen Mitarbeitern genutzt, meist allerdings nur punktuell. Sie stellen damit „Insellösungen“ dar, die meist für sich alleine stehen. Von einer ähnlichen, oder gar einheitlichen, Herangehensweise an das Data Management kann momentan innerhalb der FhG keine Rede sein.

Die befragten Mitarbeiter wünschen sich auch keine einheitliche Herangehensweise. Im Gegenteil, viele waren mit ihren Methoden des Data Management durchaus zufrieden und würden sich höchstens zusätzliche Schulungen wünschen oder können sich bessere Data Management-Software zwar vorstellen, wissen aber von keinem konkreten Produkt (einzig der konkrete Wunsch nach einem FhG-internen-„Sourceforge“³ wurde geäußert).

Erst wenn es darum geht Daten mit Kollegen, innerhalb eines Projektes oder mit anderen Instituten, auszutauschen und zu bearbeiten wird der Ruf nach spezifischeren Data Management-Lösungen laut. Ein Beispiel für eine solche Lösung im Bereich der Software-orientierten Institute stellt das Open Source-Programm „Subversion“ dar. Dabei handelt es sich um Software, mit der Versionen von Dokumenten, Webseiten oder Sourcecode verwaltet werden können. Solche Programme bieten die Möglichkeit, Daten und Dokumente zu archivieren. Ein solches Feature wird von einigen befragten Mitarbeitern gewünscht.

Die FhG hat zahlreiche Softwareprodukte, die für das Data Management genutzt werden können, bereits lizenziert, wie zum Beispiel Endnote oder Citavi. Die Nutzung dieser Software ist den Instituten freigestellt und nicht vorgeschrieben. In diesem Zusammenhang wurde auch vielfach die Meinung geäußert, dass das Problem nicht die Software sei, sondern meistens der User. Gemeint ist hier, dass das beste Data Management Tool nichts nützt, wenn es nicht richtig angewandt wird.

Die meisten befragten Mitarbeiter sehen weniger in der Software ein Problem als in dem Einhalten verbindlicher Regeln. Bei der korrekten Erfassung von Metadaten, die das spätere Wiederauffinden vereinfachen würden, ist Disziplin nötig. Regeln sind gerade in der Projekt- und Abteilungsarbeit sehr wichtig, damit Daten einheitlich genutzt werden können. Auch in diesem Bereich wurde darauf hingewiesen, dass Schulungen eine Möglichkeit wären, um einheitliche Regeln für alle Mitarbeiter aufzustellen. Es ist dabei weniger die Unlust der Mitarbeiter, dass solche Regeln nicht eingehalten werden. Vielmehr scheint es häufig an der Unkenntnis des richtigen Gebrauchs der Software zu liegen.

Oft war den Mitarbeitern nicht bekannt, dass der Zugriff auf einzelne Softwarelösungen bereits über die FhG möglich ist. So wurde beispielsweise vereinzelt der Wunsch nach einem guten Literaturverwaltungsprogramm geäußert. Die FhG hat seit dem Jahr 2008 eine Fraunhofer-weite Nutzungs- lizenz für Citavi, das nicht nur ein hoch gelobtes Literaturverwaltungsprogramm

³ Bei Sourceforge handelt es sich um eine Website, welche von Softwareentwicklern genutzt werden kann, um ihre Open Source Projekte zu verwalten (siehe: sourceforge.net).

ist, sondern auch Möglichkeiten zum Projektmanagement bietet, wie z. B. das Anlegen von Projektordnern, das Kommentieren von Literatur durch alle Projektteilnehmer, etc. Bereits vorher war mit dem Programm Endnote eine ausgereifte Lösung für die Literaturverwaltung für die Fraunhofer-Gesellschaft lizenziert.

Ein anderes Beispiel für Software, die in der FhG vorhanden, jedoch von den befragten Mitarbeitern wenig genutzt wird, ist Livelink, ein kommerzielles Wissens- und Contentmanagementprogramm. Besonders interessant ist die Nennung der Software BSCW. Dieses Tool ist ebenfalls für das Dokumentenmanagement nutzbar und ursprünglich eine Fraunhofer-Entwicklung (Fraunhofer FIT). Der Bekanntheits- und Nutzungsgrad war bei den Interviewpartnern jedoch sehr gering.

Geradezu einheitlich wirkt dagegen der Bereich der Verwaltung, der mit dem Datenbanksystem SIGMA eine Data Management-Lösung besitzt, die praktisch an jedem Institut der FhG genutzt wird. Dies ist auch verständlich, da die Daten im Bereich der Verwaltung weitestgehend homogen sind, und somit eine einheitliche Lösung optimal ist.

Die geschilderte Situation führt dazu, dass das „Rad immer wieder neu erfunden“ werden muss. Am Fraunhofer INT sollte beispielsweise eine Wiki-Software zur Unterstützung der Abteilungs- und Projektarbeit eingeführt werden. Nachdem eine Mitarbeiterin sich intensiv mit den verschiedenen Wiki-Engines/Programmen beschäftigt hatte, wurde eine Software ausgewählt und am Institut eingeführt. Erst Wochen später lernte dieselbe Mitarbeiterin per Zufall einen Kollegen der ZV kennen, der bereits einige Erfahrung auf dem Gebiet besaß (theoretisch-technisch wie praktisch). Auf klassischem Weg über Ansprechpartner der ZV wäre dieser Kontakt nicht zustande gekommen. Mehrere telefonische Versuche, über offizielle Ansprechpartner der ZV Unterstützung bei dieser Thematik zu erhalten (es gab mehrere Ansprechpartner, die für diese Fragestellung in Frage kamen) verliefen ins Leere. Die Ansprechpartner besaßen weder das gleiche Fachwissen wie ein Experte – was aufgrund der zu bearbeitenden Themenfülle auch gar nicht bei jedem Aspekt möglich wäre - noch konnten sie direkt an jemanden mit entsprechender Expertise verweisen.

Ein ähnliches Problem stellte die Einführung einer Desktopsuche am Institut dar. Nach eingehender Beschäftigung mit den verschiedenen Programmen und der Auswahl einer für das INT idealen Lösung wird nun ein Arbeitsbericht veröffentlicht. Wie kann aber gewährleistet werden, dass andere Institute, die ebenfalls eine solche Desktopsuche einführen möchten, anstatt den gleichen Aufwand erneut zu betreiben sofort auf den Arbeitsbericht aufmerksam gemacht werden können? Hier wäre eine Lösung ideal, welche bei einmaligem

Eingeben einer Suchanfrage alle relevanten Dokumente, Daten und Ansprechpartner liefert (sog. „One-Stop-Shopping“).

Dieser Zustand ist unbefriedigend. Die Verbesserung des Data Management steht daher zu Recht auf der Agenda der IT-Strategiegruppe. Eine Studie aus dem Jahre 2006, die im Auftrag des Fraunhofer-Vorstands in Zusammenarbeit der Fraunhofer Institute IPSI (Institut für Integrierte Publikations- und Informationssysteme), IRB (Informationszentrum Raum und Bau) und IZB (Institutszentrum Schloss Birlinghoven) durchgeführt wurde, beschäftigt sich mit einer ähnlichen Thematik, nämlich den „Anforderungen der Fraunhofer-Mitarbeiter an eine verbesserte IT-Unterstützung der Forschung“ (im folgenden „IPSI-Studie“ genannt).

Wir haben diese Studie für unsere Arbeit genutzt, da einige der Ergebnisse sich mit unseren Untersuchungen überlappen. An dieser Stelle möchten wir uns bei den Autoren für die freundliche Erlaubnis für die Nutzung dieser Studie in unserer Arbeit bedanken. Nebenbei sollte bemerkt werden, dass diese Studie von einem der befragten Mitarbeiter erwähnt wurde; sie gehörte ursprünglich nicht zu der Literatur, die allgemein als Basis für unsere Studie betrachtet wurde. Auch dies ist ein Beispiel dafür, dass das Data- und Dokumentenmanagement der FhG optimiert werden sollte. In dieser Studie wurde eine Befragung von insgesamt 4234 Fraunhofer-Mitarbeitern durchgeführt, der verwertbare Rücklauf lag bei 869 (20,5%). Damit ist diese Befragung repräsentativer als unsere Stichprobe, die gestellten Fragen zielten größtenteils jedoch in andere Richtungen.

Für die Bewertung des Data Management-Bedarfs Experiment-orientierter Institute ist Abbildung 7 interessant. Hierbei wurden die Fraunhofer-Mitarbeiter auf einer vierstufigen Skala (trifft zu, trifft eher zu, trifft eher nicht zu, trifft nicht zu) nach der Zugreifbarkeit, Verfügbarkeit, Bedarf und Relevanz von Datenbeständen befragt. Die Skala wurde im Anschluss auf zwei Kategorien zusammengefasst.

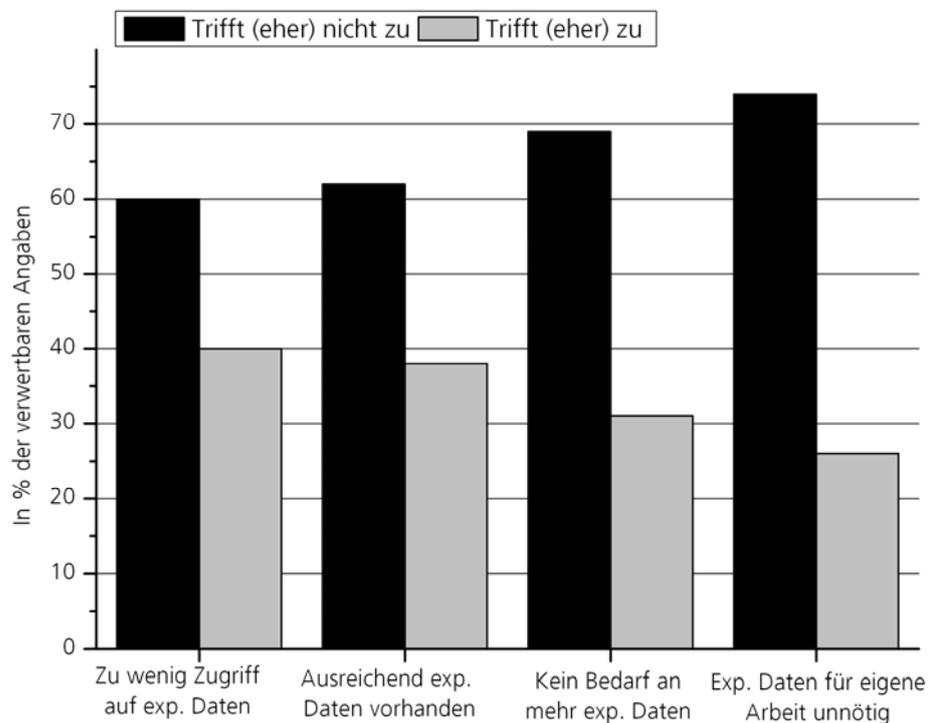


Abbildung 7: Bedarf an externen experimentellen Datenbeständen (graphisch modifiziert nach: „IPSI-Studie“; dort Abbildung 13; Seite 34)

An dieser graphischen Auswertung ist zu sehen, dass ein großer Teil der befragten Mitarbeiter externe experimentelle Daten (also jene von anderen Forschern) für ihre eigene Arbeit benötigten und auch Bedarf an mehr experimentellen Daten haben. Dabei stellt sich die Frage danach, wie diese externen Daten sinnvoll in die am Institut bereits vorhandenen Datenbestände integriert werden können.

Ähnlich wie unsere Befragung zeigt Abbildung 8 die Hindernisse der Einführung neuer Informationstechnologien, die sich mit denen decken, die wir von unseren Interviewpartnern hörten, als Kommentare zur Nutzung neuer Data Management-Tools. Die Frage, welche in der IPSI-Studie gestellt wurde lautete: „Was sind nach Ihrer bisherigen Erfahrung Hindernisse für die Nutzung

einer neuen Informationstechnologie zur Unterstützung der Forschung und Entwicklung?“

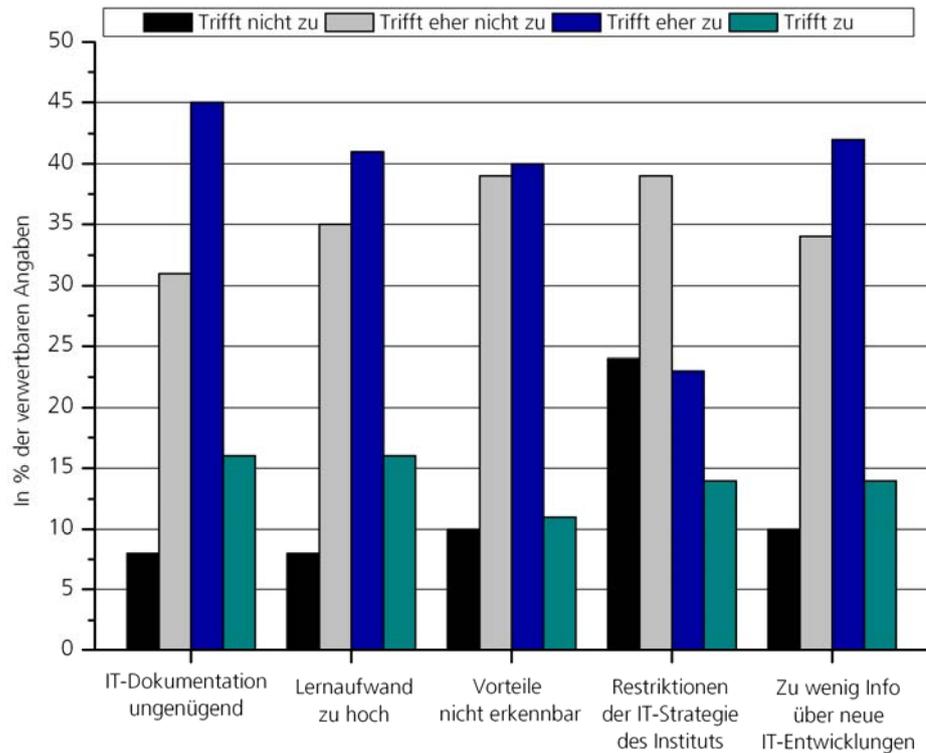


Abbildung 8: Hindernisse bei Einführung einer neuen Informationstechnologie (graphisch modifiziert nach: „IPSI-Studie“; dort Abbildung 17; Seite 39)

In diesem Zusammenhang wurde von den meisten beklagt, dass keine ausreichende IT-Dokumentation vorhanden sei, der Lernaufwand zu hoch sei, die Vorteile der neuen Software nicht klar erkennbar seien und es zu wenig Information über neue IT-Entwicklungen gebe. Restriktionen der IT-Strategie der Institute wurden dagegen von den meisten Befragten nicht bemängelt.

Diese Aussagen ähneln den Ergebnissen unserer Fragebögen. Auch hier werden die Informationen über Software, die genutzt werden könnte als zu gering eingestuft. Als Folgen sind die Vorteile neuer Software nicht bekannt. Die Schuld für diesen Zustand wird auch hier nicht beim Institut gesehen. Im Gegenteil, bei der Frage danach, wie die FhG den Mitarbeitern beim Data Management helfen könnte, wurde die Aussage getätigt, dass sie dies gar nicht könne.

Im Bereich Literaturverwaltung (Abbildung 9) sind die Kollegen des IPSI bei ihrer Befragung ebenfalls zu einem ähnlichen Ergebnis wie wir mit unserer Stichprobe gekommen. In der IPSI-Studie wurde gefragt: „Welche Werkzeuge verwenden Sie zur Verwaltung Ihrer Literatur und zur Erstellung von Literaturanhängen?“ Hier muss aber darauf hingewiesen werden, dass das Literaturverwaltungsprogramm Citavi erst seit 2008 zur Verfügung steht und damit zur Zeit der IPSI-Befragung nicht zum Softwarepool der FhG gehörte.

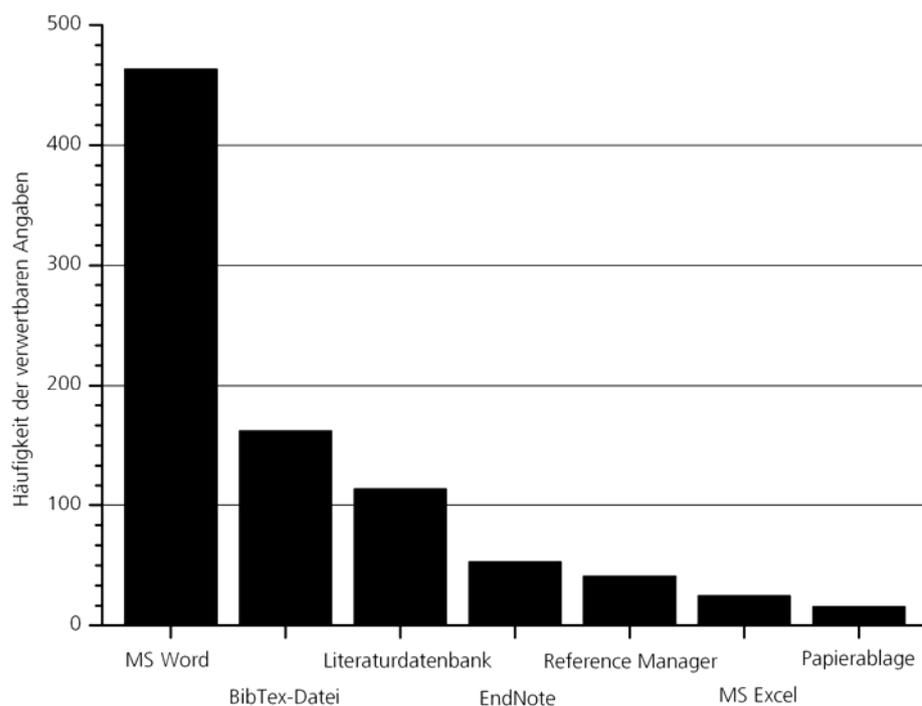


Abbildung 9: Werkzeuge zur Literaturverwaltung (graphisch modifiziert nach: „IPSI-Studie“; dort Abbildung 23; Seite 48)

MS-Word kommt in diesem Bereich eine dominante Stellung zu, gefolgt von der Open-Source-Software Bibtex. Endnote wird von gerade einmal 53 der befragten genannt. Das MS-Office Paket wurde auch bei unserer Befragung auffällig häufig als Tool für das Data- und Dokumentenmanagement genannt.

Etwas überraschend sind die Ergebnisse für den Bereich des „kollaborativen Schreibens“, also beispielsweise bei der gemeinsamen Projekt- oder Abteilungsarbeit (Abbildung 10). Die Frage lautete: „Welche spezielle IT-Unterstützung verwenden Sie, wenn Sie gemeinsam mit mehreren Autoren einen Artikel schreiben?“

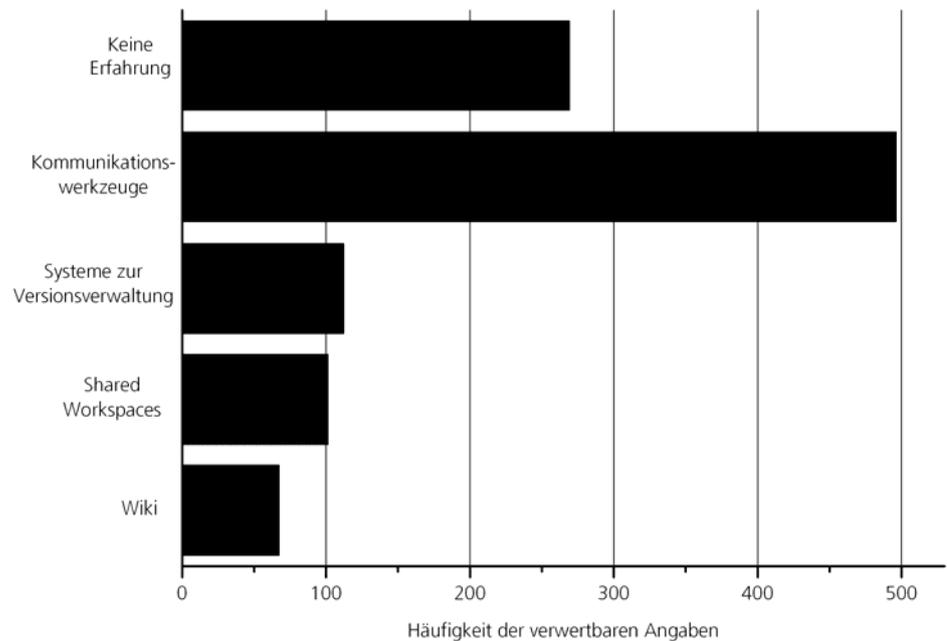


Abbildung 10: IT-Unterstützung beim kollaborativen Schreiben (graphisch modifiziert nach: „IPSI-Studie“; dort Abbildung 24; Seite 49)

Eine enorm hohe Anzahl von Befragten benutzt entweder „klassische“ Kommunikationswerkzeuge, wie zum Beispiel E-Mail, Chat oder das Telefon, oder hat keine Erfahrung mit IT-unterstütztem kollaborativem Schreiben. Hier sollte darauf hingewiesen werden, dass wahrscheinlich der Gebrauch von Wikis oder Systemen zur Versionsverwaltung inzwischen etwas verbreiteter sein dürfte. Generell stimmen diese Ergebnisse jedoch mit denen unserer Stichprobe überein.

Eine weitere Beobachtung ist, dass die Probleme, die Fraunhofer-Mitarbeiter beschäftigten, sehr wenig mit Technologien und Tools zu tun haben, die auf dem Gartner-Hype-Cycle „Data Management 2007“ zu finden sind. Auch wenn die dort aufgeführten übergreifenden Konzepte (natürlich) eine Bedeutung für das Data Management innerhalb der FhG haben, so sind doch die meisten der dort aufgeführten Technologien sehr weit von der Arbeitswirklichkeit und den realen Problemen der Mitarbeiter entfernt. Dies liegt vor allem daran, dass der Hype-Cycle sich vor allem mit solchen Konzepten beschäftigt, die für die Bearbeitung von großen Datenmengen in Unternehmen geeignet sind.

Die meisten wissenschaftlichen und technischen Arbeitsplätze in der FhG sind jedoch in einem viel kleinteiligeren Umfeld anzutreffen, was zum Teil auch die Probleme bei der Kommunikation über das nächste Umfeld hinaus bedingt. Daher ist die Kluft zwischen den abstrakten Konzepten des Hype-Cycle auf der einen, und den alltäglichen Problemen der Mitarbeiter auf der anderen Seite so groß, dass bei den folgenden Betrachtungen kaum noch auf die Gartner-Technologien eingegangen wird.

Insgesamt lässt sich sagen, dass die derzeitige Situation innerhalb der FhG im Bereich „Data Management“ **gravierende Mängel** aufweist, die unbedingt behoben werden sollten. Insbesondere sind zu nennen:

- Mangelhafte Kenntnisse der einzelnen Mitarbeiter über die Möglichkeiten moderner Data Management-Tools
- Mangelhafte Kenntnisse der einzelnen Mitarbeiter über die Möglichkeiten, Unterstützung von zentraler Stelle zu erhalten
- Mangelndes Vertrauen darauf, überhaupt Hilfe von zentraler Stelle zu bekommen
- Mangel an Plattformen zur Zusammenarbeit über die eigene Arbeitsgruppe hinaus
- Fehlende Koordination der vereinzelt Anstrengungen zur Verbesserung des Data Management auf lokaler Ebene, was zu Mehrfachbearbeitung von Fragestellungen führt

7 Bewertung anhand ausgewählter Szenarien

Ausgehend vom ermittelten Ist-Zustand wurden Szenarien erstellt, um die Folgen von verschiedenen Maßnahmen zu analysieren. In der folgenden Beschreibung der Szenarien werden zunächst Handlungsmöglichkeiten vorgestellt, deren wahrscheinliche Auswirkungen zusammengefasst und anschließend Chancen und Risiken bewertet. Diese Bewertungen werden abschließend verglichen, um zu den Empfehlungen im letzten Kapitel zu gelangen.

Die Szenarien steigern sich in vier Schritten vom Nichtstun bis zur totalen Kontrolle.

7.1 Szenario 1: „Status Quo“

Maßnahmen:

In diesem Szenario wird davon ausgegangen, dass keine Maßnahmen ergriffen werden, der *status quo* also einfach erhalten bleibt.

Auswirkungen:

Es ist wahrscheinlich, dass Nichtstun kurzfristig zu keiner Veränderung der Situation führen würde. Andererseits kann angenommen werden, dass es durch das zu erwartende weitere Anschwellen der Datenflut langfristig zu einer Verschlechterung der Lage kommt, da die derzeit verwendeten (minimalistischen) Methoden des Data Management immer weniger geeignet sein werden um damit umzugehen.

Chancen:

Der organisatorische und finanzielle Aufwand sind bei diesem Ansatz minimal. Auch bleibt die Autonomie der einzelnen Institute unangetastet.

Risiken:

Die Schwächen des derzeitigen Zustands (wie Informationsdefizite und mangelhafter Informationsaustausch zwischen Gruppen und Instituten) bleiben bestehen. Da sich die bestehenden Defizite durch höhere Anforderungen in Zukunft eher stärker auswirken werden, sind langfristig Wettbewerbsnachteile zu erwarten.

7.2 Szenario 2: „Information“

Maßnahmen:

Im Rahmen dieses Szenarios wird die Information über vorhandene Data Management-Systeme allgemein verfügbar gemacht und intensiver verbreitet. Dazu wird ein „Help-Desk“ als zentraler Anlaufpunkt für alle Fragen des Data Management eingerichtet. Die Aufgaben der Help-Desk-Mitarbeiter sind:

- Beobachtung des Marktes in Hinblick auf Data Management-Lösungen
- Erprobung vielversprechender Programme
- Aktives Abfragen von verwendeten Lösungen bei den einzelnen Fraunhofer-Instituten
- Kommunikation der Ergebnisse z.B. über das Fraunhofer Intranet, Newsletter, das Fraunhofer Journal und über Schulungen

Dieses Help-Desk würde über die allgemeine Informationstätigkeit (Publikationen unterschiedlicher Art) hinaus nur auf Anfrage, entweder von Instituten oder einzelnen Mitarbeitern, beratend in Aktion treten. Dies bedeutet, dass weiterhin die einzelnen Institute für die Einführung oder Erweiterung ihres Data Management zuständig sind. Andererseits ist es essentiell, dass die Existenz eines solchen Help-Desks und der Umfang seiner Leistungen allen Mitarbeitern bekannt sind. Daher ist es günstig das Help-Desk „Data Management“ in einen größeren Verbund von Kompetenzzentren einzubinden.

Auswirkungen:

Während der organisatorische Aufwand gering sein dürfte, da nicht in bestehende Kompetenz- und Entscheidungsstrukturen eingegriffen wird, ist mit einem gewissen finanziellen Aufwand zu rechnen, da an zentraler Stelle zusätzliche Aufgaben erfüllt werden müssen. Dem gegenüber steht die Erwartung, dass das Wissen der Institute und der einzelnen Mitarbeiter über Data Management steigt, wodurch langfristig positive Effekte entstehen. Auch ist durch das Aussprechen von Empfehlungen durch die Help-Desk-Mitarbeiter mit einer gewissen Vereinheitlichung der genutzten Lösungen zu rechnen.

Chancen:

Positiv sind der geringe organisatorische und der überschaubare finanzielle Aufwand. Dieser Ansatz würde die Autonomie der Institute unberührt lassen, was die Durchsetzung der Maßnahme sicher erleichtert.

Der größte positive Effekt wird durch die Vermeidung von Mehrfachbearbeitung erzielt, wodurch derzeit ein beträchtlicher Effizienzverlust entsteht. So

kann sich ein Mitarbeiter, eine Abteilung oder ein Institut in diesem Szenario an das Help-Desk wenden um Hilfe bei Problemen mit dem Data Management vor Ort zu erhalten. Diese Hilfe kann in verschiedenen Formen erfolgen, z.B. durch Hinweise auf bereits von der FhG lizenzierte Produkte, der Vorstellung der „besten drei“ Lösungen die für ein spezifisches Problem derzeit auf dem Markt erhältlich sind oder der Vernetzung mit anderen FhG-Einrichtungen, von denen bekannt ist, dass sie bereits ähnliche Probleme gelöst haben. Darüber hinaus ist es denkbar, dass Mitarbeiter des Kompetenzzentrums auch individuell bei der Lösung vor Ort mitwirken könnten, wobei zu überlegen wäre, ob dieser personalintensive Service für die jeweiligen Kunden kostenpflichtig ist.

Ein weiterer positiver Effekt wäre die Sensibilisierung der Mitarbeiter für Probleme im Data Management und deren Lösungen. Allerdings ist zu bezweifeln, ob dies durch „noch einen Newsletter“ gelingt.

Risiken:

Da bei diesem Szenario Anreize zur Vereinheitlichung nur schwach ausgeprägt sind, könnte es dazu führen, dass sich die Institute im Bereich des Data Management weiter voneinander entfernen. Die bekannte Schwäche beim Informationsaustausch über Institutsgrenzen hinweg könnte sich dadurch verstärken.

Außerdem besteht das Risiko, dass auch dieses zentrale Angebot von den einzelnen Mitarbeitern bzw. Instituten nicht wahrgenommen wird, die eingesetzten Ressourcen also wirkungslos verpuffen.

7.3 Szenario 3: „Föderale Struktur“

Maßnahmen:

In diesem Szenario wird ein Kompetenzzentrum eingerichtet, das dem Help-Desk von Szenario 2 ähnlich ist. Dieses Kompetenzzentrum ist jedoch mit zusätzlichen Befugnissen und Aufgaben ausgestattet:

- Erstellung einer Liste von in der FhG gebräuchlichen Tools (davon abweichende Beschaffung muss begründet werden)
- Unterhalt von einem oder mehreren Portalen, in denen lokale Datenbanken zu Kompetenzen, Projekten, etc. zentral abgefragt werden („Data Federation“, die Datenbanken bleiben unter lokaler Kontrolle)
- Betrieb von zentralen Kooperationsplattformen, wie z.B. Wikis oder „Sourceforges“

Darüber hinaus werden von den einzelnen Instituten „Data Management-Beauftragte“ benannt, die die Umsetzung vor Ort koordinieren und das Bereitstellen lokaler Informationen für zentrale Dienste sicherstellen.

Da das beschriebene Kompetenzzentrum einerseits weitreichende Befugnisse hat, Data Management andererseits an vielen Stellen nahtlos in andere Bereiche des Wissens- und IT-Managements übergeht, ist das Kompetenzzentrum „Data Management“ Teil einer übergreifenden Struktur von technischen und strategischen Verantwortlichkeiten.

Auswirkungen:

Sowohl der finanzielle als auch der organisatorische Aufwand sind höher als in Szenario 2. Die positiven Effekte bezüglich der Information der Mitarbeiter sind wahrscheinlich größer. Die zusätzlichen Portallösungen tragen zur Bewältigung der Kommunikationsprobleme bei.

Jedes Institut bekommt eine individuell zusammengestellte (maßgeschneiderte) Lösung angeboten, was sich positiv auf die Effizienz auswirken dürfte.

Vor allem der Verlust an Entscheidungskompetenz der einzelnen Institute könnte zu Widerständen gegen die Einführung des Konzepts führen. Auf der anderen Seite kann die Portallösung nur funktionieren, wenn die Datenbankinhaber einen ausreichend hohen Teil ihres Datenbestandes zur Ansicht freigeben.

Chancen:

Das Risiko der Mehrfachbearbeitung von Data Management-Problemen wird noch stärker als in Szenario 2 gesenkt, da die Mitarbeiter gezwungen sind mit dem lokalen Beauftragten oder dem zentralen Kompetenzzentrum Kontakt aufzunehmen. Auch die Festlegung auf eine Positivliste von Tools senkt den lokalen Rechercheaufwand. Außerdem sind durch den Auftritt der gesamten FhG als Großkunde finanzielle Vorteile bei der Beschaffung und dem Unterhalt der ausgewählten Tools zu erwarten.

Mit der Portallösung wäre darüber hinaus das bestehende Problem der Kommunikation zwischen Abteilungen und Instituten einfacher zu lösen. Die Beschränkung auf einige wenige Programme und Datenbanksysteme für den lokalen Einsatz erleichtern zudem die technische Machbarkeit der Datenföderation. Die Durchsetzung dieses Konzepts wird dadurch erleichtert, dass jede Organisationseinheit „Herr über die eigenen Daten“ bleibt.

Die Ernennung von lokalen Beauftragten weist dem Thema eine höhere Bedeutung zu und stellt sicher, dass alle Informationen, zumindest hinsichtlich der zentralen Ansprechpartner, auf Institutesebene verfügbar sind.

Risiken:

Das Gelingen dieses Konzepts ist von der Kooperation der einzelnen Institute abhängig. Wenn es auf Widerstand oder auch nur Gleichgültigkeit der Institute trifft, ist ein Erfolg sehr fraglich. Hier wird es sehr auf die Qualität der Arbeit des Kompetenzzentrums ankommen: Fehler, z.B. bei der Kommunikation von Erkenntnissen oder bei der Auswahl von Programmen für die Positivliste, würden die Kooperationsbereitschaft der Institute rasch senken und die positiven Effekte wären kaum noch zu realisieren.

Durch die Benennung von lokalen Data Management-Beauftragten und deren Arbeit werden die Institute zusätzlich belastet, wenn auch nur in bescheidenem Ausmaß.

7.4 Szenario 4: „Zentrale Struktur“

Maßnahmen:

Die Fraunhofer-Gesellschaft versucht, in den nächsten Jahren ein zentrales Data Management für alle ihre Forschungseinrichtungen einzuführen. Eine zentrale Abteilung legt dabei die Tools, Abläufe und Standards für das Datenmanagement fest (Enterprise Information Management). Alle Mitarbeiter werden im Umgang mit den vorgegeben Tools geschult und zu deren Einsatz verpflichtet. Möglichst alle, zumindest aber ein großer Teil der Daten (Business Intelligence) der Fraunhofer-Forschungseinrichtungen müssten in autoritative Datenbanken/Data Warehouses integriert werden, damit eine einheitliche Datenbasis vorhanden ist.

Es ist notwendig die Einhaltung dieser Grundsätze (Compliance) zu überwachen, entweder durch die zentrale Abteilung oder durch lokale Data Management-Beauftragte, denen in diesem Fall erhebliche Verantwortung zukommt.

Auswirkungen:

Daten Management ist völlig zentralisiert. Es besteht kaum Flexibilität für die einzelnen Institute. Der Datenaustausch ist durch einheitliche Formate und Vorgehensweisen stark erleichtert.

Chancen:

Die Eindeutigkeit der Daten und der zuverlässige Informationsaustausch sind positiv zu bewerten. Sie machen den Hauptvorteil von Systemen wie SIGMA aus.

Das Data Management-System kann zentral betreut werden, was möglicherweise zu einem Effizienzgewinn führt.

Risiken:

Es ist zu erwarten, dass ein System, das die Wahlmöglichkeiten für die Institute und Mitarbeiter stark einschränkt und Arbeitsprozesse formalisiert, bei Wissenschaftlern auf wenig Gegenliebe stößt. Die Überwindung dieser Widerstände würde eines starken Willens bei der Einführung bedürfen, wobei die Frage ist, wer innerhalb der FhG bereit wäre diese Anstrengung auf sich zu nehmen.

Die große Vielfalt an Information, mit der innerhalb der FhG umgegangen wird (vgl. Kapitel 2.1) erfordert ein hochkomplexes Data Management-System. Da es nicht möglich sein wird dieses „von der Stange“ zu kaufen ist mit hohen Ausgaben für die Beschaffung zu rechnen. Zusätzlich dürfte die Einführung eines solchen Systems mit hohem personellem Aufwand verbunden sein und längere Zeit beanspruchen.

Weitere Probleme sind zu erwarten, wenn Institute der FhG beitreten oder von der FhG in einen anderen Forschungsverbund wechseln. Im ersten Fall müsste das beitretende Institut sein Data Management dem FhG-Regime unterwerfen (was einen Beitritt evtl. weniger attraktiv macht), im zweiten Fall müssten die zentral verwalteten Daten ebenfalls migriert werden.

7.5 Vergleich der Szenarien

Vor dem eigentlichen Vergleich ist festzustellen, dass Szenarien 2 und 3 relevanter als Szenarien 1 und 4 sind. Der Charme von Szenario 1 (Status Quo) liegt lediglich im geringen Aufwand. Allerdings ist durch Nichtstun eine Verschlechterung der Situation zu erwarten. Bloßes Abwarten würde zudem dem Anspruch der FhG eine führende Forschungsorganisation zu sein nicht gerecht. Mit diesem Anspruch kollidiert auch das stark zentralistisch geprägte Szenario 4 (Zentrale Struktur): Während es in einigen Bereichen der FhG (wie etwa der Verwaltung) durchaus berechtigt ist, dem reibungslosen Ablaufen standardisierter Prozesse den Vorzug über individuelle Freiheit und Flexibilität zu geben, bedarf das wissenschaftliche Arbeiten eben dieser Dinge. Abgesehen davon erscheint es unrealistisch, eine so starke Zentralisierung bei der derzeitigen Struktur der FhG umzusetzen.

Bleiben also die beiden Szenarien (Information und Föderale Struktur), die auf zentrale Hilfe für die einzelnen Institute aufbauen. Während in Szenario 2 das zentrale Problem des mangelhaften Datenaustauschs ungelöst bleibt, wird eine Lösung in Szenario 3 mit einem gewissen Kompetenzverlust der Institute erkaufte. Unserer Meinung nach liegt die optimale Lösung zwischen den beiden Szenarien 2 und 3. Durch die Beschreibung dieser Szenarien wurde also ein Raum aufgespannt in dem durch politische und organisatorische Prozesse eine wünschenswerte und machbare Lösung identifiziert werden kann.

Es ist zu prüfen, inwiefern technische Lösungen das Ziel der Datenföderation (Portal) ermöglichen, ohne dass es zu einer Vereinheitlichung der heterogenen Datenbasis kommen muss. Fortschrittliche technische Lösungen dieser Art könnten die unbefriedigende Situation, dass guter Zugang zu Daten meist mit der Einschränkung individueller Gestaltungsfreiheit einhergeht, zumindest mildern.

Als zentrale Voraussetzungen zum Erreichen der positiven Effekte wurde bei beiden Szenarien zum einen die qualitativ hochwertige Arbeit der zentralen Stelle, zum anderen die intensive Kommunikation der Existenz und des Aufgabenspektrums derselben an alle Mitarbeiter identifiziert.

8 Zusammenfassung und Empfehlung

Beschreibung der Ziele:

Das Data Management in der Fraunhofer Gesellschaft muss verbessert werden. Dabei sollten folgende Ziele verfolgt werden:

1. Reduzierung des Kosten- und Zeitaufwandes bei Projekten mit mehreren Beteiligten durch konsequente Anwendung moderner Data Management Methoden.
2. Verbesserung der Kommunikation zwischen FhG-Mitarbeitern, -Instituten und anderen Kooperationspartnern vor allem im Bereich des kollaborativen Schreibens.
3. Erhöhung des Wissensstandes der FhG-Mitarbeiter über vorhandene Data Management Software bzw. Ausbau der Fähigkeiten mit ihr umzugehen.

Begründung:

Die föderale Struktur der Fraunhofer Gesellschaft hat dazu geführt, dass die Institute unterschiedliche Ansätze zum Data Management verfolgen. Dies führte in der Vergangenheit zur Mehrfachbearbeitung ähnlicher Probleme an verschiedenen Stellen, zum Beispiel bei Auswahl und Beschaffung bestimmter Software.

Ähnlich verhält es sich bei der Kommunikation zwischen Instituten, Abteilungen und Mitarbeitern innerhalb der FhG sowie mit externen Kooperationspartnern. Die befragten Mitarbeiter gaben mehrheitlich an sich bei der Projektarbeit, speziell beim kollaborativen Schreiben, traditioneller Kommunikationsformen zu bedienen (E-Mail, Telefon, persönliches Gespräch etc.). Inzwischen finden sich auf dem Markt zahlreiche Software-Lösungen, die die Projektarbeit vereinfachen könnten. Auch in der FhG wurden solche Programme entwickelt. Diese Tools werden jedoch nach unserem Wissensstand kaum, oder nicht ausreichend genutzt.

Empfehlungen:

Insgesamt lässt sich feststellen, dass die Anforderungen an Data Management individuell sehr unterschiedlich sind, daher existiert kein kleinster gemeinsamer Nenner für eine ideale Lösung. Die Umfrageergebnisse deuten darauf hin, dass es derzeit keine technische Lösung für alle Probleme beim Data Management gibt. Möglicherweise wird es sie nie geben, da die Anforderungen der einzelnen Mitarbeiter zu verschieden sind.

Empfehlung 1: Von einer übergreifenden (*one size fits all*) Software-Lösung für den Bereich Data Management in der gesamten Fraunhofer-Gesellschaft ist zurzeit abzuraten.

Allerdings gibt es durchaus Software, die bei der Lösung von Teilproblemen hilft. Software dieser Art wird in höchst unterschiedlichem Maße von den Mitarbeitern eingesetzt. Besonders im Bereich des Software-orientierten wissenschaftlichen Arbeitens werden solche Produkte als ausgereift und außerordentlich nützlich bewertet. Der Bereich Verwaltung stellt durch seine stark strukturierte Arbeitsweise und durch das Programmpaket SIGMA einen Sonderfall dar. Die Zufriedenheit mit SIGMA war bei den befragten Mitarbeitern hoch. In allen anderen Bereichen wurden keine allgemein verbreiteten Lösungen identifiziert. Dies ist umso erstaunlicher, da die FhG einige Programme aus diesem Bereich lizenziert hat und allen Mitarbeitern ohne weitere Kosten zur Verfügung stellt.

Empfehlung 2: Der Bekanntheitsgrad des Portfolios lizenzierter Programme muss erhöht werden.

Empfehlung 3: Bei der Prüfung von Systemen sollten vorzugsweise Open-Source-Systeme und Eigenentwicklungen von Fraunhofer berücksichtigt werden, da diese an die jeweiligen Bedürfnisse angepasst werden können.

Als bedeutendes Problem wurde der Austausch von Informationen über Abteilungs- bzw. Institutsgrenzen hinweg identifiziert.

Empfehlung 4: Ob die existierende Software zum Informationsaustausch die optimale Lösung darstellt, sollte überprüft werden. Die zwei bis vier besten Programme sollten intensiv beworben werden, da viele Mitarbeiter nicht einmal Kenntnis über die verfügbaren Lösungen haben. Dies ist nur ein Symptom eines generellen Mangels an Koordination der Data Management-Anstrengungen

innerhalb der FhG. Letztendlich führt dieser Mangel zur Mehrfachbearbeitung ähnlicher Probleme und damit verbunden zu Effizienzverlusten.

Empfehlung 5: Bereits verwendete Lösungen müssen erfasst und bei Bedarf verfügbar gemacht werden. Ein indirekter Austausch der einzelnen Institute mit einer zentralen Vermittlungsstelle erscheint dabei am realistischsten.

Diese Forderungen verlangen nach verstärkten Anstrengungen im zentralen Bereich.

Empfehlung 6: Es sollte eine Stelle eingerichtet werden, die sich zentral mit den Belangen des Data Management befasst. Eine Einbindung in ein größeres Netzwerk an Koordinierungsstellen ist wünschenswert. Mögliche Aufgaben einer solchen Stelle finden sich im Kapitel zu den Szenarien (speziell unter 7.2 und 7.3). Die Tätigkeit einer solchen Stelle würde im besten Fall zu der Etablierung einer *best practice* und damit einer wünschenswerten Vereinheitlichung der Ansätze führen, ohne dass administrativer Druck ausgeübt werden muss.

Empfehlung 7: Der Einsatz von Programmen (Portalen), die dazu beitragen Daten aus unterschiedlichen Quellen einheitlich präsentieren zu können, sollte verstärkt werden.

Empfehlung 8: Die einzelnen Institute sollten motiviert werden ihr Wissen in höherem Maße miteinander zu teilen. Eine Informationsplattform, die dies ermöglicht, sollte verfügbar gemacht und beworben werden.

Unsere Interviewpartner betonten immer wieder, dass Data Management sehr viel mit den Kenntnissen und dem Verhalten der einzelnen Mitarbeiter zu tun hat.

Empfehlung 9: Es sollten spezifische Fortbildungsangebote zum Data Management gemacht werden. Diese könnte sehr unterschiedliche Formen annehmen. Oft würden schon kurze Hinweise weiterhelfen.

Empfehlung 10: Derzeit scheinen im weiteren Bereich des Data Management Wikis und Desktop-Suchtools große Bedeutung zu haben. Hier sollte verstärkt geschult werden.

Zeitraumen:

Projektvorschlag A: Evaluierung der Möglichkeiten das bestehende zentrale Software-Portfolio in Bezug auf Data Management zu erfassen, zu ergänzen und zu promoten. Zu diesem Zweck sollten zusätzlich lokale Software-Portfolios einzelner Institute und Open-Source-Software analysiert und bewertet werden. Software zum Austausch von Informationen über Instituts- bzw. Abteilungsgrenzen hinweg und die Erarbeitung eines Konzepts zur Einführung solcher Programme bzw. Methoden sollten besonders berücksichtigt werden. (Q1/2009)

Projektvorschlag B: Evaluierung der Möglichkeiten ein kommentiertes Verzeichnis aller Softwarelösungen, die in der FhG genutzt werden, zu erstellen und zu pflegen. Die Beschreibungen der Programme müssten um Anwendererfahrungen ergänzt werden. Es sollte ersichtlich sein, an welchen Instituten die Programme genutzt werden, damit Ansprechpartner identifiziert werden können. Ein geeignetes Information Retrieval System zur Erschließung dieser Informationen wird vorgeschlagen. (Q1/2009)

Projektvorschlag C: Evaluierung der möglichen Aufgaben einer zentralen Stelle „Data Management“ und Vorschläge zur Ausstattung und organisatorischen Ansiedlung derselben. (Q1-Q2/2009)

Projektvorschlag D: Untersuchung der Praktikabilität von Portalen zur übergreifenden Darstellung von Informationen verschiedener Quellen und Erstellung eines Konzepts zur Einbindung einzelner Organisationseinheiten der FhG als Informationsprovider. Bereits existierende Portale, wie die Fraunhofer Publica und E-Prints, werden eingebunden. (Q1-Q2/2009)

Projektvorschlag E: Feststellung von Wissensdefiziten der FhG-Mitarbeiter in Bezug auf Data Management und Erstellung eines Informations- und Schulungskonzepts um den Wissensstand zu erhöhen. (Q1-Q2/2009)

Nebenbedingungen:

Bei der Implementierung unserer Empfehlungen für das Data Management in die bestehende Struktur der Fraunhofer Gesellschaft muss besonders auf die individuellen Bedürfnisse der Institute geachtet werden. Gerade zu Beginn muss geprüft werden, welchem Institutstyp welche Software empfohlen wird, um Vertrauen in diese Institution aufzubauen. Mangelndes Vertrauen der Institute würde dazu führen, dass das Help-Desk für Data Management keine ausreichend Akzeptanz fände. Die richtige Kommunikation der Aufgabe und der Vorteile eines solchen Help-Desks muss nach Möglichkeit alle Mitarbeiter der FhG erreichen.

9 Anhänge

9.1 Fragebögen

(siehe nächste zwei Seiten)

Version 1.1 vom 23.4.2008

Projekt „Data Management“ für IT-Strategiegruppe

- Fragebogen für ein strukturiertes Interview -

Befragte Person:

Datum:

Interviewer:

Einleitung:

„Data Management“ beinhaltet die Gesamtheit aller Verfahren um mit der steigenden Datenflut sinnvoll umzugehen. Wir führen im Auftrag der IT-Strategiegruppe der Fraunhofer-Gesellschaft eine Untersuchung zu diesem Thema durch. Unsere Untersuchung ist die erste von drei Phasen eines Prozesses an dessen Ende eine übergreifende Strategie zum „Data Management“ für die FhG steht. Wir möchten Ihnen nun einige Fragen stellen, um Ihre Meinung zu diesem Thema zu erfahren.

Fragen:

1. Was verstehen Sie unter „Data Management“? Welche Methoden (oder Tools) kennen Sie? Welche Methoden (oder Tools) werden in Ihrem Bereich bereits angewandt?

2. Welche positiven oder negativen Erfahrungen haben Sie beim „Data Management“ bereits gemacht? (Lessons learned)

3. Wo sehen Sie Defizite? Welche Probleme müssen noch gelöst werden? Welche Methoden (oder Tools) fehlen Ihnen? Wie kann Ihnen die FhG, zentral oder auf Institutsebene helfen?

Abschluß:

Wir danken Ihnen für Ihre Auskünfte. Sollte Ihnen noch etwas einfallen, das wir berücksichtigen sollten, setzen Sie sich bitte bis zum _____ (14 Tage) mit uns in Kontakt!

Version 1.2 vom 6.5.2008

Projekt „Data Management“ für IT-Strategiegruppe

- Fragebogen für ein strukturiertes Interview -

Befragte Person:

Datum:

Interviewer:

Einleitung:

„Data Management“ beinhaltet die Gesamtheit aller Verfahren um mit der steigenden Datenflut sinnvoll umzugehen. Wir führen im Auftrag der IT-Strategiegruppe der Fraunhofer-Gesellschaft eine Untersuchung zu diesem Thema durch. Unsere Untersuchung ist die erste von drei Phasen eines Prozesses an dessen Ende eine übergreifende Strategie zum „Data Management“ für die Fraunhofer-Gesellschaft steht. Wir möchten Ihnen nun einige Fragen stellen, um Ihre Meinung zu diesem Thema zu erfahren.

Fragen:

1. a) Was verstehen Sie unter „Data Management“?
b) Welche Methoden (oder Tools) werden in Ihrem Bereich angewandt?
c) Welche weiteren Methoden (oder Tools) kennen Sie?

2. Welche positiven Erfahrungen haben Sie beim „Data Management“ bereits gemacht? (Lessons learned)

3. a) Wo sehen Sie Defizite? Welche Probleme müssen noch gelöst werden?
b) Welche Methoden (oder Tools) fehlen Ihnen?
c) Wie kann Ihnen die Fraunhofer-Gesellschaft, zentral oder auf Institutsebene helfen?

Abschluß:

Wir danken Ihnen für Ihre Auskünfte. Sollte Ihnen noch etwas einfallen, das wir berücksichtigen sollten, setzen Sie sich bitte in den nächsten 14 Tagen mit uns in Kontakt!

9.2 Abkürzungen

BI	Business Intelligence
DBMS	Data Base Management system
ECM	Enterprise Content Management
EII	Enterprise Information Integration
EIM	Enterprise Information Management
FGAN	Forschungsgesellschaft für Angewandte Naturwissenschaften e. V.
FhG	Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e. V.
Fraunhofer INT	Institut für Naturwissenschaftlich-Technische Trendanalysen
Fraunhofer IPSI	Institut für Integrierte Publikations- und Informationssysteme
Fraunhofer ISI	Institut für System- und Innovationsforschung
FTP	File Transfer Protocol
ICI	Information-centric Infrastructure
ICT	Information and Communication Technology
IT	Informationstechnik,
MDM	Master Data Management
MS	Microsoft
OSS	Open Source Software
SaaS	Software as a Service
W3C	World Wide Web Consortium
WCM	Web Content Management
XML	Extensible Markup Language

9.3 Abbildungen

Abbildung 1: Schematische Darstellung der drei Typen wissenschaftlicher Arbeit	4
Abbildung 2: Schema zur Data Federation	8
Abbildung 3: Schema zur Data Integration.	9
Abbildung 4: Technologien im Gartner-Hype-Cycle „Data Management“ aus den Jahren 2006-2008 inklusive Illustration der durch uns vorgenommene Unterteilung des Themenfeldes.	12
Abbildung 5: Graphische Darstellung der Zusammenhänge in der Kategorie „Übergeordnete Begriffe“	14
Abbildung 6: Graphische Darstellung der Zusammenhänge in der Kategorie „Data Integration und Data Quality“	19
Abbildung 7: Bedarf an externen experimentellen Datenbeständen	38
Abbildung 8: Hindernisse bei Einführung einer neuen Informationstechnologie	39
Abbildung 9: Werkzeuge zur Literaturverwaltung	40
Abbildung 10: IT-Unterstützung beim kollaborativen Schreiben	41

9.4 Kurzbiografien der Autoren

Dr. Joachim C. Burbiel

Dr. Burbiel ist Apotheker und hat nach diversen Praktika im In- und Ausland (Liverpool/UK, Queson City/Philippinen) an der Ludwig-Maximilians-Universität München im Fach „Pharmazeutische Chemie“ promoviert. Anschließend war er sechs Jahre als wissenschaftlicher Mitarbeiter am Pharmazeutischen Institut der Friedrich-Wilhelms-Universität Bonn tätig. Seit 2007 ist er wissenschaftlicher Mitarbeiter am INT.

Miloš Jovanović, M.A.

Miloš Jovanović MA studierte an der Universität Düsseldorf Geschichte im Hauptfach und in den Nebenfächern Politik, Medienwissenschaft und Informationswissenschaft. Seit November 2006 promoviert er beim Fraunhofer Institut für Naturwissenschaftlich-Technische Trendanalysen (INT) im Fach Informationswissenschaft (Bibliometrie). Daneben ist er Lehrbeauftragter an der Universität Düsseldorf.

Dr. Silke Römer

Dr. Silke Römer studierte Biochemie und Physik an der Ruhr-Universität Bochum. Nach dem Abschluss als Dipl.-Biochemikerin mit Hauptfach Biophysik promovierte sie an der Universität Witten/Herdecke am Institut für physikalische Biochemie mit den Schwerpunkten Biophysik, Proteinbiochemie sowie Molekularbiologie. Sie ist seit 2007 wissenschaftliche Mitarbeiterin am Fraunhofer INT.