

Dienstleistungen für die Langzeitarchivierung digitaler Daten – Herausforderungen und Geschäftsmodelle

Die langfristige Erhaltung und Bewahrung von digitalen Unterlagen stellt für Archive und Bibliotheken eine zentrale Herausforderung der nächsten Jahrzehnte dar. Die Lösungssuche konzentriert sich zurzeit hauptsächlich auf akademische Projekte an Universitäten und Forschungsinstitutionen, während Dienstleistungen von privaten Firmen kaum nachgefragt oder angeboten werden. In diesem Artikel wird die Frage erörtert, ob und wie sich ein solcher Markt in den nächsten Jahren entwickeln könnte und wer darauf als Anbieter und Kunde auftreten wird. Es werden mögliche Dienstleistungen präsentiert und vier Geschäftsmodelle vorgestellt, mit denen sich die Langzeitarchivierung von digitalen Daten auf einem freien Markt betreiben lässt.

The long-term preservation and storage of digital documents represents a central challenge for archives and libraries over the next decades. The search for solutions is currently centred primarily in academic projects at universities and research institutions, while the services of private companies are rarely requested or offered. This article addresses the question of whether such a market could be developed in the coming years and who will play the parts of provider and customer. Potential services are presented and four possible business models are described which would enable a long-term preservation of digital data to be conducted in the commercial business sector.

EINLEITUNG

Die Langzeitarchivierung digitaler Daten und Unterlagen ist seit mehreren Jahren ein häufig diskutiertes Thema. Man versteht darunter die Gewährleistung der technischen, intellektuellen und logistischen Verfügbarkeit und Nutzbarkeit von digitalen Informationen über lange Zeiträume und unterschiedliche Technologiegenerationen hinweg. Zurzeit sind es hauptsächlich staatliche und akademische Institutionen, die sich um Lösungen in diesem Bereich bemühen. Bibliotheken und Archive haben häufig einen gesetzlichen Auftrag zur unbeschränkten Aufbewahrung von Informationen, unabhängig vom dabei verwendeten Informationsträger. Bei Firmen aus der Privatwirtschaft ist hingegen das Problembewusstsein und der Bedarf nach Dienstleistungen in diesem Bereich noch wenig entwickelt, ungeachtet dessen ob sie nun selbst davon betroffen sind oder als Anbieter von Lösungen auftreten.

In der Privatwirtschaft wird eine zeitlich unbeschränkte Archivierung nur selten gefordert, in vielen Branchen oder in bestimmten Teilgebieten sind Aufbewahrungsfristen von zwanzig und mehr Jahren aber durchaus gebräuchlich. Man denke hier vor allem an Firmen, die aufgrund der von ihnen hergestellten Produkte langfristigen Haftpflichtrisiken und -ansprüchen unterliegen oder Krankenversicherungen und Rentenkassen, die ihre Kundendaten so lange aufbewahren müssen wie sie diese Kunden betreuen, also bis zu deren Tod und meistens sogar darüber hinaus.

Dabei unterscheiden sich die Problemfelder bei staatlichen Institutionen und privaten Firmen eigentlich gar nicht so sehr, in beiden Fällen geht es einerseits um die Sicherung der Nachvollziehbarkeit von Geschäftsvorgängen einerseits und um die Verfügbarkeit, Lesbarkeit und Verstehbarkeit von langfristig wertvollen digitalen Informationen andererseits.

Diese Diskrepanz zwischen vornehmlich akademischen Projekten bei staatlichen Stellen und dem Desinteresse von Unternehmen aus der Privatwirtschaft ist hauptsächlich dafür verantwortlich, dass zum jetzigen Zeitpunkt noch kaum Dienstleistungen für die Langzeitarchivierung von digitalen Daten angeboten oder nachgefragt werden. Man kann sich aber trotzdem bereits darüber Gedanken machen, welche Geschäftsmodelle sich im Bereich der digitalen Langzeitarchivierung denn überhaupt realisieren lassen und wer als möglicher Kunde oder Anbieter solcher Dienstleistungen auftreten könnte. Auf diese Fragen möchte dieser Artikel näher eingehen und mögliche Antworten liefern. Dabei werden zuerst fünf grundlegende Thesen zur zukünftigen Entwicklung von Dienstleistungsangeboten für die digitale Langzeitarchivierung vorgestellt und anschließend die Marktteilnehmer präsentiert, also die potenziellen Kunden und Anbieter solcher Dienstleistungen. Danach folgt eine nähere Beschreibung der Dienstleistungen und zum Schluss werden vier mögliche Geschäftsmodelle vorgestellt und Schlussfolgerungen gezogen.

THESEN

These 1: Öffentliches Problem

Digitale Langzeitarchivierung stellt für staatliche Archive und Bibliotheken die zentrale Herausforderung der nächsten zwanzig Jahre dar, um ihren gesetzlichen Auftrag weiterhin erfüllen zu können. Sie spielt jedoch im privaten Sektor heute keine wesentliche Rolle. Ausnahmen zeigen sich in spezialisierten Gebieten wie etwa bei Produktdaten mit langfristigen Haftpflichtrisiken (beispielsweise in der Automobil- und Luftfahrtbranche), wo in Zukunft neben der klassischen Mikroverfilmung auch vermehrt digitale Lösungen zum Einsatz kommen werden.

These 2: Für große Firmen uninteressant

Im Gegensatz zu den ständig steigenden Milliardenumsätzen bei Lösungen für »Enterprise Content Ma-



Heinz Gnehm

Foto privat



Peter Keller-Marxer

Foto privat



Stephan Heuscher

Foto privat

nagement« (ECM) und Speichersysteme (mit gesetzlichen »Compliance«-Anforderungen und »Zehnjahresfristen«) wird das Problem der digitalen Langzeitarchivierung im privaten und öffentlichen Sektor auf absehbare Zeit keinen Markt generieren, der für große Hersteller interessant genug wäre, erhebliche Investitionen in die Entwicklung und Forschung für Produkte zu tätigen oder langfristiges Risikokapital in Dienstleistungsangebote zu investieren.

These 3: Akademische Lösungssuche

Lösungen für die digitale Langzeitarchivierung werden auch weiterhin primär an Universitäten, Bibliotheken und Archiven im Rahmen von öffentlich finanzierten Forschungsprojekten und weitgehend geschlossenen akademischen Kreisen diskutiert und entwickelt. Damit verbunden ist eine ungenügende Nachhaltigkeit bei der Entwicklung, dem Unterhalt und der allgemeinen Nutzbarkeit von technischen Lösungen (»Prototypplastigkeit«) sowie eine stark fragmentierte Investitionstätigkeit in einem personalzentrierten, institutionsspezifischen und wirtschaftlich nicht wirksamen Bereich einiger Dutzend Millionen Euro (in Europa).

These 4: Markt für Spezialdaten

Ein Markt ergibt sich vorab für kleine Firmen, welche Beratung und maßgeschneiderte Lösungen für Unternehmen in sehr spezialisierten und speicherintensiven Datenbereichen (naturwissenschaftliche Forschungsdaten und technische Produktdaten) anbieten. Besonderes Merkmal dieser Kunden ist, dass sie die vollständige Kontrolle über die Erzeugung der Daten (Datenformate, Metadaten) und ihrer späteren Nutzung haben.

These 5: Gemeinsame Kompetenzzentren

Für öffentliche Archive und Bibliotheken ist langfristig nur die Einrichtung und Nutzung von gemeinsamen technischen Infrastrukturen und Kompetenzzentren sinnvoll und finanzierbar. Einzellösungen sind auf lange Sicht zu teuer und zu ineffizient, und es müssen deshalb vermehrt kooperative Modelle mit einer zentralen Datenspeicherung angestrebt werden.

STAKEHOLDERS

Kunden

Grundsätzlich haben nur staatliche oder öffentliche Institutionen wie Archive, Bibliotheken und Museen einen gesetzlichen oder verbindlichen Auftrag zur unbeschränkten Aufbewahrung von digitalen Unterlagen oder Objekten, und sie bilden damit ein primäres Kundensegment für Dienstleistungen im Bereich der Langzeitarchivierung von digitalen Daten.

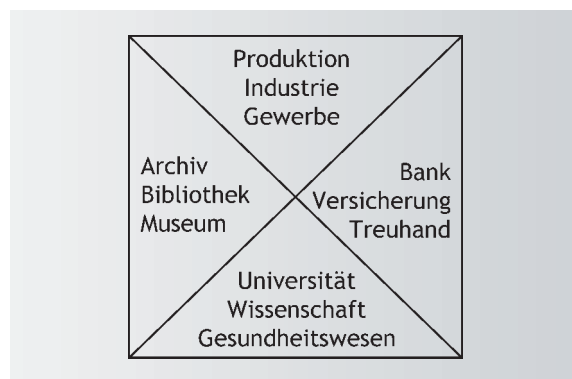


Abb. 1: Kunden

Daneben müssen aber auch andere Institutionen und Unternehmen der Privatwirtschaft digitale Unterlagen für längere Zeiträume aufbewahren, wobei darunter eine Zeitspanne von mindestens fünfzehn Jahren zu verstehen ist. Spitäler müssen beispielsweise Röntgenbilder und weitere digitale Unterlagen ihrer Patienten unter Umständen ein Leben lang aufbewahren, angesichts der großen Datenmenge kein leichtes Unterfangen. Es ist deshalb nicht verwunderlich, dass sich einige Anbieter ganz auf die Archivierung von medizinischen Unterlagen spezialisiert haben und auch eine Auslagerung und den sicheren Zugriff über das Internet anbieten. Eine ähnliche Situation gilt auch für Krankenversicherungen sowie Renten- und Pensionskassen, die die Daten ihrer Versicherten bis ans Lebensende und sogar darüber hinaus aufbewahren und dabei auch noch die Vorgaben des Datenschutzes berücksichtigen müssen.

Im Bereich der Industrie sind es vor allem die Produkthaftungspflicht und die damit verbundenen Risiken, die eine langfristige Archivierung von Konstruktionsplänen, Produktionsunterlagen etc. notwendig machen. So müssen beispielsweise die Zulassungsdokumente für ein neues Medikament mindestens so lange aufbewahrt werden, wie das Präparat auf dem Markt angeboten wird und dann noch einige Jahre darüber hinaus. Im Bereich Automobil- oder Flugzeugbau gelten ähnlich lange Fristen für die Aufbewahrung der Konstruktionspläne und Testunterlagen.

Dabei unterscheidet die Gesetzgebung heutzutage kaum mehr zwischen den einzelnen Informationsträgern, was dazu führt, dass elektronische Unterlagen den herkömmlichen Trägern wie Papier oder Mikrofilm gleichgestellt sind. Viele Unterlagen werden aber weiterhin noch auf Papier oder Mikrofilm archiviert, obwohl die Informationen elektronisch erzeugt worden sind.

Zu erwähnen sind auch wissenschaftliche Institutionen im Bereich Raumfahrt, Luftfahrt, Erdbeobachtung, Experimentalphysik etc. wie etwa das Euro-

päische Kernforschungszentrum (CERN) in Genf, die deutschen Max-Planck- und Fraunhofer-Gesellschaften, Wetterdienste etc. Bei diesen Forschungszentren werden teilweise Datenmengen im Umfang von mehreren Petabytes (PB) pro Jahr erzeugt, die langfristig wertvoll und deshalb für weitere Forschungen aufzubewahren sind.

Ein heute noch unterschätztes, aber in Zukunft wahrscheinlich stark wachsendes Kundensegment bilden Privatpersonen, da heutzutage ein immer größerer Bestandteil unserer persönlichen Kommunikation und Lebensgestaltung über elektronische Kanäle erfolgt und die so erzeugten digitalen Daten auch langfristig zur Verfügung stehen sollen. Es stellt sich die berechnete Frage, ob man heute erstellte CDs mit digitalen Fotos den Kindern und Enkelkindern noch zeigen können. Das digitale Äquivalent der Schuhschachtel ist leider noch nicht erfunden worden, und neben Fotos, Videos und Musikstücken sind es zunehmend Briefe, E-Mails, Familienalben etc., die im privaten Bereich immer mehr in digitaler Form erzeugt, verwaltet und gespeichert werden und auch in zwanzig Jahren noch lesbar und benutzbar sein sollen. Zwar existieren für Privatpersonen bereits Internetdienste zum Auslagern von Daten auf externe Server, dabei steht aber weniger die Langzeitarchivierung als vielmehr das Backup der persönlichen Daten im Vordergrund. Daneben gibt es Forschungsprojekte wie etwa »MyLifeBits« von Microsoft, mit dem sich das gesamte Leben eines Menschen digital verfolgen, speichern und wieder abrufen lassen soll.

Digital Asset Management-Produkte

potenzielles Kundensegment Privatpersonen

katastrophensichere Auslagerung von Daten

Enterprise Content Management-Produkte

Anbieter

Anbieter von Dienstleistungen rund um die Langzeitarchivierung von digitalen Daten sind zurzeit noch recht selten anzutreffen. Meistens handelt es sich dabei um Informatikunternehmen, die diese Dienstleistung nur als Nebengeschäft betreiben und zusätzlich als nützliches Verkaufsargument für ihre Hard- oder Softwarelösungen verwenden.

Im Softwarebereich sind hier vor allem Firmen wie Open Text, FileNet, EMC², IBM, Hummingbird zu erwähnen, die Produkte für so genanntes »Enterprise Content Management« (ECM) anbieten und dabei auch die Archivierung von digitalen Daten erwähnen. Bei näherer Betrachtung stellt sich aber heraus, dass die Langzeitarchivierung nicht im eigentlichen Sinne verstanden wird und die Archivierung sich auf einen Zeitraum von maximal zehn Jahren beschränkt. Auch ist keines dieser Produkte konform zum Referenzmodell »Open Archival Information System« (OAIS), obwohl mindestens zwei Hersteller das von ihrer Lösung behaupten.

Für die Langzeitarchivierung von digitalen Daten nur wenig besser geeignet sind Lösungen für das so genannte »Digital Asset Management« (DAM). Zurzeit findet eine Marktbereinigung statt, und einige Hersteller von DAM-Produkten sind von großen ECM-Anbietern aufgekauft worden, was durchaus sinnvoll ist, da sich die beiden Produkte in ihren Einsatzbereichen stark überschneiden. Während aber ECM-Systeme aus dem geschäftlichen Umfeld kommen, sind DAM-Produkte häufig aus Katalogsystemen für Fotos und audiovisuelle Medien hervorgegangen. Aus diesem Grund legen DAM-Systeme auch viel Wert auf die Erschließung und Verzeichnung der digitalen Objekte mit Hilfe von möglichst umfassenden Metadaten. Bei ECM-Systemen steht demgegenüber vor allem die Unterstützung der Geschäftsprozesse im Vordergrund, und diese Lösungen bieten denn auch umfangreiche Möglichkeiten zum Modellieren, Verwalten und Messen von Prozessen an, was als »Business Process Management« (BPM) bezeichnet wird.

Daneben gibt es eine wachsende Zahl von Unternehmen, die teils in ehemaligen Minenstollen oder Armeefestungen katastrophensichere Daten- und Rechenzentren betreiben und dort auch die Speicherung von digitalen Unterlagen und unternehmenskritischen Daten anbieten, entweder mittels physischer Lagerung der Backup- oder Archivmedien oder durch eine Übermittlung der Daten über Telekommunikationseinrichtungen. Ein Anbieter garantiert dabei sogar, dass die eingelagerten Daten (in diesem Fall Röntgenbilder) bis zu dreißig Jahre lang zugänglich und lesbar bleiben sollen. Der Hauptzweck einer solchen Auslagerung von Daten ist allerdings nicht in erster Linie die Langzeitarchivierung, sondern vielmehr der Schutz vor Datenverlust oder einem Ausfall der Informatikinfrastruktur des Kunden.

Dienstleistungsunternehmen, die sich mehrheitlich oder ganz auf die Beratung im Bereich der Langzeitarchivierung von digitalen Daten spezialisiert haben, sind zurzeit noch selten. Allerdings gibt es hier zahlreiche Überschneidungen mit anderen Bereichen, hauptsächlich in der Informatik, weshalb Dienstleistungen zur Langzeitarchivierung beispielsweise auch von Informatik- oder sonstigen Beratungsunternehmen erbracht werden können, obwohl sie nicht explizit angeboten oder beworben werden. Es scheint, dass das verfügbare Marktpotenzial hier noch lange nicht ausgeschöpft und eine entsprechende Nachfrage erst am Entstehen ist. Es bleibt dabei abzuwarten, wie stark sich hier die etablierten Informatikfirmen sowie die Hard- und Softwareanbieter profilieren und sich einen entsprechenden Anteil dieses Marktes sichern können.

Unterschiedliche Interessen von Kunden und Anbietern

Ein Grundproblem bei allen Dienstleistungen für die Langzeitarchivierung ist die lange Zeitdauer. Es lässt sich zum jetzigen Zeitpunkt kaum realistisch abschätzen, wie sich heutige Konzepte und Methoden in den nächsten zehn, zwanzig oder fünfzig Jahren bewähren werden. Ein Kunde möchte nicht heute Geld investieren, um dann in zwanzig Jahren oder sogar noch später feststellen zu müssen, dass sich die Investition nicht gelohnt hat, die Daten nur noch teilweise lesbar sind und erneut mit viel Aufwand vor dem kompletten Verlust gerettet werden müssen. Auf der anderen Seite ist der Lieferant einer Dienstleistung im Gegenzug nicht bereit, eine Garantie zur Lesbarkeit und Benutzbarkeit von digitalen Daten über zehn, zwanzig oder mehr Jahre einzugehen und sich damit einem schwer kalkulierbaren Risiko auszusetzen, da das Überleben der Daten in hohem Maße von der zukünftigen Entwicklung der Informationstechnologie abhängt, auf die der Hersteller oder Lieferant aber nur einen begrenzten Einfluss haben kann. Die Interessen von Kunden und Lieferanten liegen in diesem Fall also weit auseinander, und es ist deshalb nötig, dass für Dienstleistungen im Bereich der Langzeitarchivierung ganz neue Modelle der Zusammenarbeit und Partnerschaft gefunden werden müssen, damit sich die Kosten und Risiken zwischen den Partnern aufteilen lassen.

Ebenso unsicher ist die Festsetzung eines Preises für solche Dienstleistungen. Zahlt ein Kunde bereits heute die Archivierung für die nächsten vierzig Jahre, oder wird ihm jedes Jahr eine neue Rechnung präsentiert, die mit wachsendem Aufwand für den Unterhalt der Daten auch immer höher ausfallen wird? Gehen bei Nichtbezahlung der Rechnung die Daten einfach verloren, oder bekommt der Kunde sie in einer Form zurück, die es ihm erlaubt, die Daten einem anderen Anbieter anzuvertrauen? Auch hier werden eventuell neue Finanzierungsmodelle benötigt, und vielleicht werden in Zukunft die Preise für die digitale Langzeitarchivierung als Versicherungsprämien bezahlt, mit denen sich die Verfügbarkeit und das Risiko eines Verlustes langfristig absichern lassen.

Zwischen den unterschiedlichen Interessen der beteiligten Partner einen Ausgleich zu finden, wird zu einer der großen Herausforderungen der kommenden Jahre werden. Das fängt bereits bei den Begriffen an. Archivierung bedeutet für einen Informatiker nicht das gleiche wie für einen Archivar oder einen Bibliothekar. Ein Mitarbeiter im Bereich Informationsmanagement sieht in der langfristigen Sicherstellung der Verfügbarkeit von digitalen Objekten eine große Herausforderung, während der Informatiker darin vor

allem ein technisches Problem sieht, das sich auch mit technischen Mitteln lösen lässt. Obwohl beide Disziplinen das Wort »Information« im Namen tragen, sind die Auffassungen sehr unterschiedlich, und für langfristige und tragfähige Lösungen muss deshalb zuerst eine gemeinsame Sprache gefunden werden, damit beide Seiten ihre Sicht der Dinge darlegen können und gegenseitig verstanden werden.

Dienstleistungen

Die möglichen Dienstleistungen im Umfeld der Langzeitarchivierung von digitalen Daten lassen sich einerseits nach strategischen (»planen«) und operationellen (»tun«) Tätigkeiten sowie nach dem aktuellen Erkenntnis- und Fortschrittsstand einteilen. Abbildung 2 zeigt diese Einteilung und unterteilt den Erkenntnisstand in die Phasen »Wissen«, »Verstehen«, »Umsetzen«, »Nutzen« und »Bewahren«. Der mittlere Ring umfasst dabei die strategischen, der äußere Ring die operationellen Aufgaben.

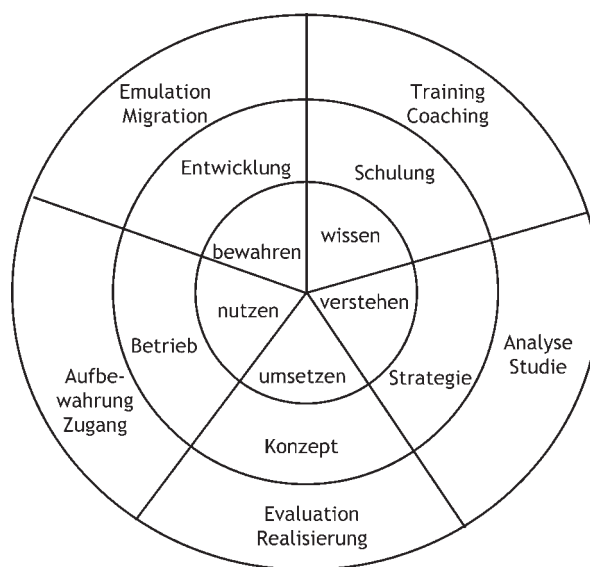


Abbildung 2: Dienstleistungen

Wissen bedeutet, sich der Herausforderungen der Langzeitarchivierung von digitalen Daten bewusst zu sein, die grundlegenden Fragestellungen, Problemfelder und Lösungsstrategien zu kennen und die relevanten Auswirkungen für das eigene Unternehmen oder die eigene Institution erkennen zu können, Lösungsansätze und Produkte nach ihrer Eignung zur Langzeitarchivierung beurteilen und bewerten sowie eine strategische Sicht einnehmen zu können.

Verstehen bedeutet, die Relevanz der Langzeitarchivierung für die eigene Institution korrekt einschätzen und beurteilen zu können sowie im Rahmen von

Konzepten eine tragfähige und langfristige Weiterentwicklung sicherstellen zu können.

Umsetzen bedeutet, die konzeptionell erarbeiteten Lösungen im Umfeld der eigenen Institution realisieren und Datenbestände in solche Lösungen integrieren und darin unterhalten zu können.

Nutzen bedeutet, die eingeführten Systeme und Prozesse für die anstehenden Probleme und Herausforderungen der digitalen Langzeitarchivierung einsetzen und anwenden zu können sowie den laufenden Betrieb sicherzustellen.

Bewahren bedeutet, den Betrieb der Infrastruktur für die Langzeitarchivierung von digitalen Daten im täglichen Einsatz sicher und kosteneffizient sicherstellen und auf lange Sicht aufrechterhalten zu können.

Schulung

Die Schulung und Beratung hat zurzeit noch eine Art Brückenfunktion und muss versuchen, die doch sehr unterschiedlichen Vorstellungen von Informatikern auf der einen und Archivaren/Bibliothekaren auf der anderen Seite auf einen gemeinsamen Nenner zu bringen und die vorhandenen Ängste und Vorbehalte auf beiden Seiten abzubauen. Tatsächlich wird in der Informatik unter dem Begriff »Archivierung« etwas durchweg anderes verstanden, als ein Archivar oder Bibliothekar darunter verstehen würde. Aus diesem Grund ist man insbesondere im angelsächsischen Raum dazu übergegangen, den Begriff »archiving« durch »preservation« zu ersetzen, eine deutsche Entsprechung dazu wäre etwa Bewahrung oder Erhaltung. Am Beginn einer ernsthaften Beratung im Umfeld der digitalen Langzeitarchivierung sollte immer eine Archivierungsstrategie stehen. Man muss sich vorgängig also erst einmal die Frage stellen, welche Unterlagen man wie lange für welchen Nutzerkreis archivieren möchte. Erst nachdem eine solche Strategie formuliert worden ist, kann man sich daran machen, daraus die personellen, finanziellen, organisatorischen und technischen Ausprägungen und Rahmenbedingungen zu formulieren und die Umsetzung in Angriff zu nehmen.

Training/Coaching

Beim Training und Coaching geht es in erster Linie darum, den Wissensstand innerhalb einer Organisation in Bezug auf Langzeitarchivierung zu heben. Viele Unternehmen und Institutionen sind sich teilweise der Herausforderungen gar nicht bewusst oder vertrauen zu stark auf kurzfristige technische Lösungen. Man muss allerdings festhalten, dass private Firmen hier eindeutig weniger stark betroffen sind als staatliche Institutionen, die einen gesetzlichen Auftrag zur zeitlich unbeschränkten Aufbewahrung haben. Bei Unternehmen

geht es in erster Linie darum, die Nachvollziehbarkeit von Geschäftsvorfällen für einen gewissen Zeitraum sicherzustellen. Ein späterer Zugriff auf diese Daten ist eher selten. Trotzdem gibt es auch bei Unternehmen Datenbestände, die aus rechtlichen oder anderen Gründen länger aufbewahrt werden müssen.

Strategie

Am Anfang jeder sinnvollen Lösung zur Langzeitarchivierung sollte eine umfassende und langfristig gültige Archivierungsstrategie stehen, die als Führungsinstrument die zentralen Fragestellungen zum »Warum, Was und Wie?« beantwortet. Eine Archivierungsstrategie kann dabei entweder für eine gesamte Organisation, für einen Teilbereich oder auch nur für einen bestimmten Datenbestand erarbeitet werden. Die Strategie legt fest, welche digitalen Informationen überhaupt archiviert werden sollen, welches die künftigen Nutzungszwecke sind und wie der langfristige Betrieb und dessen Finanzierung gesichert werden können. Als Grundlagendokument soll sie dazu beitragen, eine zielgerichtete Führung und Steuerung der Archivierung zu ermöglichen.

Analyse/Studie

Bevor man die Ausarbeitung einer Archivierungsstrategie in Angriff nimmt, ist es notwendig und sinnvoll, in einer Analyse den Ist-Zustand zu erheben und dabei den eigenen Handlungsspielraum zu definieren sowie das weitere Vorgehen festzulegen. Eine Lösung zur Langzeitarchivierung von digitalen Daten kann nicht nur aus einem technischen System bestehen, sondern muss auch organisatorische, personelle und finanzielle Erwägungen einbeziehen. Eine Analyse liefert die nötigen Grundlagen, mit denen sich eine kohärente und langfristig anwendbare Archivierungsstrategie formulieren lässt.

Konzept

Das Erarbeiten und Erstellen von Konzepten für Gesamt- oder Teillösungen im Bereich der Langzeitarchivierung von digitalen Daten erfordert ein umfangreiches Verständnis der Kundenbedürfnisse und deren technischer Infrastruktur und organisatorischer Abläufe. Obwohl die Langzeitarchivierung erst am Ende der Wertschöpfungskette von Informationen steht, ist der Einbezug der vorgelagerten Prozesse von entscheidender Wichtigkeit. Langzeitarchivierung muss als integraler Bestandteil des so genannten »Life Cycle Management« betrachtet und auch entsprechend konzeptionell in die internen Prozesse und Organisationsstrukturen sowie die vorhandene technische Infrastruktur eingebettet werden.

Verständigung zwischen Informatikern, Archivaren und Bibliothekaren auf gemeinsame Grundbegriffe

Assessment/Review

Eine weitere Dienstleistung besteht darin, bestehende oder geplante Lösungen für die Langzeitarchivierung einem Review oder Assessment zu unterziehen und deren Eignung für den vorgesehenen Zweck abzuklären und Schwachstellen sowie Verbesserungsmöglichkeiten aufzuzeigen. Dabei geht es im privaten Sektor hauptsächlich um »Compliance« und »Corporate Governance«, also die Einhaltung gesetzlicher Vorschriften, Qualitätsmanagement und einer guten Geschäftsführung. Diese Vorschriften und Richtlinien können auch als Teil einer Revision oder eines IT-Audits überprüft und nötigenfalls angepasst werden.

Zurzeit sind in den USA und in Großbritannien Bestrebungen im Gang, Richtlinien für eine Zertifizierung des OAIS-Standards zu erarbeiten, mit denen sich anschließend kommerziell erhältliche Soft- und Hardwarelösungen auf ihre Konformität zu OAIS testen und zertifizieren lassen. Eine solche Zertifizierung betrifft aber nur technische Systeme und es wäre deshalb auch denkbar, eine Analyse und Zertifizierung der technischen und organisatorischen Abläufe im Bereich der Langzeitarchivierung für eine gesamte Institution oder auch nur für bestimmte Datenbestände durchzuführen (ähnlich wie dies beim Qualitätsmanagement gemäß ISO 9001 bereits möglich ist). Eine derart zertifizierte Institution könnte sich fortan als sichere, vertrauenswürdige Stelle für die Langzeitarchivierung von digitalen Daten profilieren.

Betrieb

Die Umsetzung von erarbeiteten Konzepten muss natürlich ebenfalls ein Bestandteil von Dienstleistungen der Langzeitarchivierung sein, schließlich will man die Implementierung und Inbetriebnahme von solchen Lösungen nicht allein den Hard- und Softwareherstellern überlassen. Bei der Realisierung geht es tatsächlich nicht nur um die Installation und Konfiguration von Hard- und Softwarekomponenten, sondern auch um die Anpassung von bestehenden und die Gestaltung von neuen Geschäftsprozessen, in deren Verlauf digitale Unterlagen und Informationen erzeugt, verwaltet, gespeichert und schließlich archiviert werden.

Zahlreiche Softwarehersteller bieten heute so genannte Archivierungslösungen unter dem Namen »Enterprise Content Management« (ECM) an, die Archivierung bezieht sich bei solchen Produkten aber wohl eher auf die Auslagerung von nicht mehr benötigten Daten als auf die Sicherstellung der Langzeitverfügbarkeit von digitalen Informationen. Es ist offensichtlich, dass sich Langzeitarchivierung nicht mit einem einzelnen Produkt eines bestimmten Herstellers lösen lässt, sondern dass zusätzlich auch ent-

sprechende Methoden, Prozesse und eine passende Organisation bereitstehen müssen, damit rein technische Lösungen nicht bereits in kurzer Zeit zu einem Datenfriedhof werden, aus denen sich die Daten später kaum mehr in eine neue Umgebung überführen lassen.

Hier setzen die Beratungsdienstleistungen an, damit nicht nur technische Lösungen verkauft werden, sondern diese Systeme in einen archivischen Kontext eingebunden sind und gleichzeitig sichergestellt werden kann, dass die zum Betrieb dieses digitalen Archivs notwendigen Ressourcen finanzieller, personeller, aber auch intellektueller Art langfristig zur Verfügung stehen.

Aufbewahrung/Zugang

Noch bis vor wenigen Jahren wurde die Diskussion über die Langzeitverfügbarkeit von digitalen Daten hauptsächlich von der Haltbarkeit der dafür eingesetzten Datenträger beherrscht, und man gewann den Eindruck, dass mit einem genügend lange haltbaren digitalen Speichermedium das Problem bereits gelöst sei. Von dieser sehr kurzfristigen Denkweise ist man heute abgerückt, viele Hersteller und Anbieter von Hardware-Speicherlösungen werben aber weiterhin mit bestimmten Typen von Datenträgern und preisen deren Eignung für die Langzeitarchivierung an. Ein Punkt, der dabei ebenfalls oft herausgestrichen wird, ist die Revisionssicherheit, also die Gewährleistung, dass auf einem bestimmten Datenträger gespeicherte Dokumente anschließend nicht mehr verändert werden können. Im Gesetz wird dieser Begriff nicht verwendet, sondern einzig verlangt, dass die Daten mit geeigneten Maßnahmen vor nachträglicher Veränderung geschützt werden müssen. Revisionssicherheit ist deshalb keine notwendige Voraussetzung für die Langzeitarchivierung (es ist eher umgekehrt) und Langzeitarchivierung lässt sich auch nicht allein durch die Verwendung eines bestimmten Datenträgertyps erreichen. Wenn man die Haltbarkeit von digitalen Medien betrachtet, stellt man zwar durchaus fest, dass magneto-optische Speichermedien eine physikalisch längere Haltbarkeit aufweisen als beispielsweise Magnetbänder. In zwanzig und mehr Jahren wird es aber für beide Datenträger keine Laufwerke mehr geben, mit denen man diese Daten wieder einlesen könnte, womit sich die Diskussion über die Haltbarkeit erübrigt.

Natürlich muss eine Lösung zur Langzeitarchivierung immer auch eine technische Komponente zur Speicherung umfassen, es ist aber nicht sinnvoll, hier die Haltbarkeit der Medien in den Vordergrund zu stellen. Vielmehr muss das Ziel sein, eine Lösung zu finden, die sicherstellen kann, dass die Daten in fünf

oder zehn Jahren entweder aufgefrischt oder möglichst verlustfrei und automatisch auf neuere Datenträger umkopiert werden können. Solche Überlegungen müssen bereits bei der Ausschreibung und Evaluation einer Hardware-Plattform zur Speicherung von digitalen Daten berücksichtigt werden.

Neben der Archivierung vor Ort ist auch eine Auslagerung der Daten denkbar. Solche Auslagerungen finden bevorzugt in ehemaligen Minenstollen oder militärischen Festungsanlagen statt. Die Schweiz kann hier gleich mit mehreren Anbietern aufwarten, was damit zusammenhängt, dass zahlreiche militärische Anlagen in den letzten Jahren aufgelöst und an private Unternehmen verkauft worden sind. Die Auslagerung bietet eine katastrophensichere Unterbringung und Speicherung der Daten, nicht aber deren Langzeitverfügbarkeit, da es sich bei den Daten meistens nur um Backups handelt, die häufig und regelmäßig ausgetauscht werden. Die Langzeitarchivierung benötigt natürlich nicht zwingend eine katastrophensichere Unterbringung, als zusätzliche Dienstleistung ist sie aber durchaus sinnvoll.

Entwicklung

Die Entwicklung umfasst die Pflege und laufende Aktualisierung der einmal eingeschlagenen Archivierungsstrategie und der damit verbundenen Archivierungsprozesse, das frühzeitige Erkennen von auftauchenden Problemen und deren Vermeidung, die Sicherstellung der finanziellen Mittel für die Archivierung sowie die Abstimmung der technischen und organisatorischen Aspekte der Langzeitarchivierung.

Eine einmal eingeschlagene Strategie muss laufend an die sich ändernden Rahmenbedingungen angepasst werden, seien diese nun personeller, organisatorischer, technischer oder finanzieller Art. Das bedingt ein vorausschauendes Planen und das frühzeitige Erkennen von Trends.

Emulation/Migration

Eine weitere Klasse von Dienstleistungen umfasst alle Tätigkeiten im Umgang mit den zu archivierenden oder den bereits archivierten Daten. Es kann sich dabei um die Aufbereitung, Übernahme, Erschließung, Konversion, Auffrischung oder Migration der Daten handeln. Zwar werden solche Dienstleistungen zurzeit teilweise bereits angeboten, aber nicht unbedingt im Hinblick auf die Langzeitarchivierung. Tatsächlich ist es für die Langzeitarchivierung wichtig, dass bereits bei der Erzeugung von digitalen Unterlagen an die spätere Archivierung gedacht wird. Je früher man solche Überlegungen bereits in die Erzeugung und Verwaltung von digitalen Daten einbeziehen kann, um-

so einfacher wird es später sein, die digitalen Objekte auch langfristig sichern und verfügbar halten zu können.

Damit die archivierten Daten auch tatsächlich langfristig verfügbar bleiben, müssen sie entweder regelmäßig auf neue Systeme migriert oder deren ursprüngliche Herkunftssysteme emuliert werden. Emulation und Migration bilden nicht die einzigen Strategien zur Langzeitarchivierung von digitalen Daten, es sind aber diejenigen, die zurzeit am häufigsten in staatlichen oder akademischen Projekten erforscht, angewandt und auch weiterentwickelt werden. Da Bibliotheken vor der Herausforderung stehen, neben dem Inhalt von digitalen Objekten auch deren ursprüngliches Aussehen und Verhalten zu bewahren (man denke hier beispielsweise an multimediale CD-ROMs), setzen sie eher auf die Emulation. Archive müssen dagegen den Evidenz- oder Beweiswert von Unterlagen sicherstellen, also die Integrität und Authentizität von digitalen Objekten, und bevorzugen deshalb die Migration.

Aufgrund der langen Zeiträume stellen beide Methoden eine große Herausforderung dar, und es lässt sich zur Zeit noch nicht klar abschätzen, wie gut sich diese und andere Strategien langfristig für die Archivierung eignen werden.

Dienstleistungen im Bereich der Emulation und Migration konzentrieren sich auf die Sicherung der Lesbarkeit und Verfügbarkeit von digitalen Daten. Im Falle der Migration müssen dabei große Datenmengen bewegt werden, während die Emulation ein System benötigt, mit dem sich die ursprünglichen Daten wieder lesen und darstellen lassen.

GESCHÄFTSMODELLE

Abbildung 3 zeigt vier mögliche Geschäftsmodelle zur Langzeitarchivierung von digitalen Unterlagen. Die Geschäftsmodelle werden danach unterschieden, ob die Daten intern oder extern aufbewahrt werden und ob nur die Verfügbarkeit oder auch die Lesbarkeit garantiert werden soll. Verfügbarkeit bedeutet in diesem Fall, dass die Daten in ihrem Originalzustand aufbewahrt und jederzeit wieder in dieser Form zur Verfügung gestellt werden können. Dabei besteht allerdings die Gefahr, dass das Daten- oder Dateiformat in der Zwischenzeit nicht mehr unterstützt und die Daten von keiner Applikation mehr gelesen, angezeigt, ausgedruckt oder anderweitig genutzt werden können. Ein Ansatz, der darüber hinausgeht, ist die Sicherstellung der Lesbarkeit, was bedeutet, dass entweder die Daten von Zeit zu Zeit in neue Formate migriert werden oder mit Hilfe einer passenden Emulation des ursprünglichen Systems weiterhin lesbar bleiben.

**Bibliotheken bevorzugen
die Emulation als Strategie,
Archive die Migration**

	intern	extern
verfügbar	Allmend	Schliessfach
lesbar	Wartung	Versicherung

Abbildung 3: Geschäftsmodelle

Allmend

Mit dem Begriff Allmend wurde im Mittelalter Grundeigentum in Form von Wiesen, Wald und Wasserflächen bezeichnet, das sich im Besitz der Allgemeinheit befand und von allen Mitgliedern der Gemeinschaft genutzt werden konnte. Im Bereich von digitalen Daten versucht zurzeit das Projekt »Lots of Copies Keep Stuff Safe« (LOCKSS) die moderne Umsetzung eines ähnlichen Prinzips vorzunehmen. Im Wesentlichen geht es dabei darum, dass die digitalen Daten einer Institution auf Systeme von anderen, ebenfalls partizipierenden Institutionen kopiert werden. Die untereinander vernetzten Systeme bilden in diesem Fall eine Art von Datenallmend, die allen Mitgliedern des Projekts zur Speicherung ihrer Daten offen steht.

Das LOCKSS-Konzept geht davon aus, dass das Kopieren und Verteilen von digitalen Unterlagen die Chancen erhöht, dass mindestens eine Kopie langfristig verfügbar und lesbar bleibt und sich davon wieder neue Kopien anfertigen lassen. Die beteiligten Institutionen werden sicher unterschiedliche Strategien bei der Langzeiterhaltung anwenden und dabei mehr oder weniger erfolgreich sein und vermutlich auch Teile der eigenen Datenbasis verlieren. Da die Daten aber bei anderen Institutionen weiterhin verfügbar bleiben, lassen sich Datenverluste wieder ausgleichen.

Hier ließe sich auch eine Zusammenarbeit zwischen öffentlichen Institutionen und der Privatwirtschaft denken, indem spezialisierte Firmen als Partner auftreten und beispielsweise die Migration von bestimmten Datenbeständen übernehmen. Durch die verteilte Natur von LOCKSS lassen sich solche »privaten« Partner relativ einfach in das vorhandene Netz einbinden, vorher müssen aber noch Fragen bezüglich des Urheberrechts und der Abrechnung der erbrachten Leistungen geklärt werden.

Wartung

Ein Wartungsvertrag wird heute üblicherweise für den Unterhalt von technischen Systemen wie beispielsweise Hard- und Softwarekomponenten abgeschlossen. Darin eingeschlossen sind beispielsweise

Leistungen wie die Unterstützung bei Problemfällen, die Lieferung neuer Versionen einer Software, der Austausch von defekten Teilen, eine regelmäßige Überprüfung der Funktionsweise etc. Natürlich ließe sich ein solcher Wartungsvertrag auch auf die in solchen Systemen gehaltenen Daten ausweiten.

Das Geschäftsmodell Wartung besteht darin, dass für digitale Daten ein Wartungsvertrag abgeschlossen werden kann, der die Verfügbarkeit und Lesbarkeit der Daten langfristig garantiert und dabei auch die Migration oder Emulation der Daten mit einschließt. Die Kosten eines solchen Wartungsvertrags hängen unter anderem von der Form, dem Inhalt und dem Umfang der digitalen Unterlagen ab und können schwer abgeschätzt werden. Die Kosten können auch pauschal oder nach Aufwand verrechnet werden.

Schließfach

Ein Schließfach ist eine sichere Form der Hinterlegung von Wertgegenständen bei einer Bank oder einer anderen vertrauenswürdigen Institution. Der Zugriff auf das Schließfach ist dabei nur einem ausgewählten Kreis von Personen möglich und die Bank bietet einzig die notwendige technische Infrastruktur, kennt den Inhalt eines Schließfaches aber nicht.

Dieses Modell lässt sich auch auf digitale Daten anwenden, und tatsächlich existieren solche Dienstleistungen bereits. Hauptsächlich geht es dabei um die sichere Aufbewahrung von wichtigen Unternehmensdaten, die den Weiterbetrieb eines Unternehmens nach einem vollständigen oder teilweisen Datenverlust sicherstellen können. Ein Merkmal des Schließfaches ist, dass die darin hinterlegten Objekte nicht verändert werden, im Falle von digitalen Daten also nur die Verfügbarkeit, nicht aber die Lesbarkeit garantiert werden kann. Dafür können die Daten katastrophensicher aufbewahrt und bei Bedarf schnell wieder abgerufen werden. Die Kosten eines solchen Geschäftsmodells lassen sich deshalb besser abschätzen und hängen vor allem vom Umfang der gelagerten Daten ab. Die Kosten umfassen auch das regelmäßige Auffrischen oder Umkopieren der Daten auf neue Speichermedien.

Versicherung

Eine Versicherung bietet gegen Bezahlung einer regelmäßigen Prämie Schutz gegen die finanziellen Auswirkungen eines Schadens, dabei kann es sich um einen Beinbruch, einen Diebstahl, ein Feuer oder ein anderes Schadensereignis handeln. Im Bereich der digitalen Langzeitarchivierung wäre ein solches Schadensereignis beispielsweise der drohende oder bereits eingetretene Verlust der Lesbarkeit von digitalen Unterlagen.

Projekt LOCKSS

Die Originaldaten sind zwar weiterhin verfügbar, können aber von aktuellen Applikationen nicht mehr gelesen und deshalb auch nicht mehr sinnvoll genutzt werden. Mit der Prämie würde man sich demzufolge gegen diesen Verlust der Lesbarkeit versichern und bei Eintreten des Schadensereignisses die ausbezahlte Versicherungssumme für die Migration der Daten oder die Entwicklung einer Emulationslösung aufwenden. Eingeschlossen in der Prämie ist die Sicherstellung der Verfügbarkeit analog des Schließfach-Geschäftsmodells.

Da gewisse Daten- und Dateiformate ein höheres Risiko aufweisen, irgendwann vom Markt zu verschwinden, muss auch die Prämie diesem unterschiedlichen Risiko Rechnung tragen. Die Aufbewahrung eines Microsoft Word-Dokuments würde beispielsweise mehr kosten als die Aufbewahrung des gleichen Dokuments im Format PDF des Adobe Readers. Ein solches Modell würde also gewisse Dateiformate und Vorgehensweisen gegenüber anderen bevorzugen und dadurch automatisch zu einer gewissen Standardisierung führen, was sich wiederum positiv auf die langfristige Verfügbarkeit und Lesbarkeit von bestimmten Datenformaten auswirken und Druck auf die Hersteller von Soft- und Hardwareprodukten ausüben würde, sich diesen Standards anzupassen.

SCHLUSSFOLGERUNGEN

Zusammenfassend lässt sich sagen, dass der Markt für Dienstleistungen im Bereich der Langzeitarchivierung von digitalen Daten zurzeit noch unterentwickelt ist, was vor allem darauf zurückzuführen ist, dass die Privatwirtschaft bis jetzt noch wenig Interesse an diesem Problem gezeigt hat. Das Angebot wird stattdessen stark von technischen Lösungen und Produkten im Informatikbereich dominiert, deren Eignung für die Langzeitarchivierung aber noch bewiesen werden muss. Weitere Angebote existieren vor allem in der Auslagerung von Datenbeständen, aber auch hier wird dem Problem der Langzeitarchivierung noch wenig Bedeutung beigemessen.

Die Entwicklung von Lösungen wird hauptsächlich von Bibliotheken, Archiven und Universitäten mit Geldmitteln der öffentlichen Hand vorangetrieben. Diese Lösungen sind aber häufig spezifisch auf eine bestimmte Institution zugeschnitten und kaum marktfähig.

Die langen Zeiträume, in denen digitale Daten lesbar und nutzbar bleiben sollen, stellen eine weitere Herausforderung dar, da die Hersteller besonders im Hard- und Softwarebereich einem ständigen Markt-

druck ausgesetzt sind und deshalb das alte System von gestern lieber heute als morgen durch ein neues ersetzen möchten. Nachhaltigkeit genießt in der Informatikbranche, die von der ständigen Innovation und dem technologischen Fortschritt lebt, keinen großen Stellenwert. Schließlich ist es genau diese ständige Erneuerung, die dazu geführt hat, dass die Langzeitarchivierung von digitalen Daten überhaupt zu einem Problem geworden ist.

Es bleibt abzuwarten, welche Geschäftsmodelle sich in der Langzeitarchivierung von digitalen Daten letztendlich durchsetzen werden. Die langen Zeiträume verlangen nach neuartigen Modellen der Zusammenarbeit und der Verrechnung von Dienstleistungen, deren Auswirkungen sich erst Jahre später zeigen werden. Das Interesse des privaten Sektors an der Langzeitarchivierung wird in Zukunft sicher zunehmen, und er wird dabei von den Erfahrungen von Archiven und Bibliotheken profitieren können. Hard- und Softwarehersteller müssen sich vermehrt überlegen, wie sie ihre Produkte und Lösungen nachhaltiger gestalten können, ohne dabei die Innovation zu vernachlässigen. Die Langzeitarchivierung von digitalen Daten kann nur dann gelingen, wenn alle involvierten Stakeholder gemeinsam versuchen, tragfähige und kosteneffiziente Lösungen zu entwickeln.

LITERATUR

Chapman, Stephen: Counting the Costs of Digital Preservation: Is Repository Storage Affordable? In: Journal of Digital Information (2003-05-07) Nr. 2, Article 178. <http://jodi.ecs.soton.ac.uk/Articles/vo4/io2/Chapman/> [Stand: 26. April 2005].

Lavoie, Brian F.: The Incentives to Preserve Digital Materials: Roles, Scenarios and Economic Decision-Making. www.oclc.org/research/projects/digipres/incentives-dp.pdf [Stand: 26. April 2005].

Liturg, William G.: Levels of Service for Digital Repositories. In: D-Lib Magazine, Number 5 (2002). www.dlib.org/dlib/may02/lelurgy/05lelurgy.html [Stand: 26. April 2005].

DIE VERFASSER

Heinz Gnehm, Stephan Heuscher und Peter Keller-Marxer sind die Gründer der Firma ikeep AG, Berner Technopark, Morgenstr. 129, CH-3018 Bern. E-Mails: heinz.gnehm@ikeep.com, stephan.heuscher@ikeep.com, peter.keller@ikeep.com