Interview mit einem Experten für die Langzeitarchivierung digitaler Daten beim BASE
(25. September 2024)
Länge: 2:00h

PM:

Können Sie einmal beschreiben, was Ihre Tätigkeiten und Zuständigkeiten beim BASE sind?

Experte:

Ich bin Referent im Fachgebiet "Langzeitdokumentation", das Teil der Forschungsabteilung ist. Und ich bin dort Referent für digitale Langzeitarchivierung. Ich benutze meistens die Formulierung "Langzeitarchivierung digitaler Daten". Da kommen wir gleich auch nochmal bei der Begriffsabgrenzung dazu. Und mein Hintergrund ist: Ich habe vor längerer Zeit Philosophie und Informatik studiert und habe vor fast 20 Jahren angefangen, im Bereich Langzeitarchivierung zu arbeiten. Zuerst als Softwareentwickler und dann in einer Reihe von Forschungs- und Entwicklungsprojekten. Und das war meistens im Bereich Forschungs- und Gedächtnisorganisationen, Kultureinrichtungen, Wissenschaftseinrichtungen. Zum einen im Bibliothekskontext. Und eine Zeit lang war ich auch eher mit dem Thema Forschungsdatenmanagement beschäftigt. Aber immer an der Schnittstelle zwischen IT, Gedächtnisorganisation - was für mich der Oberbegriff ist für Bibliotheken, Archiven, Museen und so weiter - und Wissenschaft.

Meine Tätigkeit: Das Fachgebiet heißt Langzeitdokumentation. Damit ist allgemein der Informations- und Wissenserhalt zum Bereich Zwischen- und Endlagerung gemeint. Und Langzeitarchivierung ist zum gewissen Grad ein deutscher Kunstbegriff. Das ist eine Diskussion, die sich insbesondere bei den Bibliotheken entsponnen hat, wie man digitale Objekte nutzbar hält. Wie man die aufbewahren kann über einen langen Zeitraum. Und man hat dann den Begriff Langzeitarchivierung gewählt, ohne - ich sage mal vereinfacht - ohne Archivare zu fragen.

Denn wenn Sie einen Archivar fragen, dann wird der sagen: "Es gibt gar keine Langzeitarchivierung, weil jede Archivierung ist auf Dauer. Da gibt es kein Enddatum." Das ist Archivaren wichtig. Deswegen verwenden sie meist den Begriff Langzeitarchivierung nicht. Zur Unterscheidung: Wenn man eher aus dem IT-Bereich kommt, dann wird unter Archivierung oft nur so etwas wie Speichern verstanden. Für Leute mit IT-Hintergrund macht dieser Zusatz Langzeitarchivierung nochmal deutlich, dass es um einen Zeitraum geht, wo halt eben technologische Veränderungen unter anderem eine Rolle spielen. Und so wird das üblicherweise auch im Deutschen verstanden. Langzeitarchivierung ist das Erhalten der Nutzbarkeit von digitalen Informationen bzw. digitalen Daten über einen Zeitraum, in dem Kontextveränderungen, technologische Veränderungen und gesellschaftliche Veränderungen bedeutsam werden. Also Langzeitarchivierung meint immer digital. Es gibt nicht in dem Sinne eine analoge Langzeitarchivierung. In dem Sinne ist es ein deutscher Kunstbegriff. Und so kann man das auch ein bisschen von der Langzeitdokumentation abgrenzen. Langzeitdokumentation, das ist für uns einfach das Digitale und das Analoge. Und diesen Begriff kenne ich auch ehrlich gesagt nur im nuklearen Entsorgungsbereich.

In anderen Kontexten - Bibliotheken Archiven, Museen - ist mir der Begriff Langzeitdokumentation bisher nie vorgekommen. Das heißt, Langzeitdokumentation ist bereits speziell für unseren Aufgabenbereich der Begriff. Und wenn wir jetzt noch eine letzte Begriffspedanterie oder eine Präzisierung machen wollen, dann kann man natürlich sagen: Dokumentation bezeichnet auch stärker den Prozess, während etwas entsteht - es festzuhalten, also zu dokumentieren. Während mit der

Archivierung gemeint ist: Es sind in der Vergangenheit Unterlagen entstanden und die werden dann übernommen und dauerhaft aufbewahrt. Das ist sozusagen das Begriffsfeld, auf dem man sich bewegt.

PM:

Gibt es dafür auch englische Synonyme?

Experte:

Im Englischen ist der typische Begriff für Langzeitarchivierung "Long-Term Preservation". Und Langzeitdokumentation... Da bin ich mir nicht exakt sicher, was im internationalen Kontext der Begriff ist, der benutzt wird. Könnte ich aber nachschauen, weil es gibt zumindest ein paar internationale Gremien oder Institutionen, die sich mit diesen Fragen beschäftigen. Aber da würde ich auch selber nur nachgucken, was die dafür verwenden. Aber wenn es um digitale Informationen geht, dann ist Langzeitarchivierung im Deutschen und Long-Term Preservation im Englischen der Begriff. Manchmal muss man halt auch aufpassen, dass der Begriff "digitale Langzeitarchivierung", wenn man ihn verwendet, so von Leuten verstanden wird, als gehe es darum, Unterlagen zu digitalisieren und sie dann zu erhalten. Das ist nicht der Punkt bei diesem Begriff. Aber das ist nicht der Aufgabenbereich, um den es mir geht. Und da würde ich sagen, dass dies ein Sonderfall ist. Um es möglichst präzise und frei von Missverständnissen auszudrücken, sage ich oft "Langzeitarchivierung digitaler Daten" oder "Langzeitarchivierung digitaler Objekte". Manchmal haben Museen gewisse Vorbehalte, weil die unter einem Objekt etwas ganz Bestimmtes verstehen. Beim Datenbegriff: Manche Leute verstehen den nicht so allgemein und dann fassen sie halt nicht Dokumente darunter. Aber ich meine es im weitestmöglichen Sinne.

PM:

Mit welcher Hardware und Software haben Sie in Ihrem Umfeld zu tun? Haben sie eigene Server und wie arbeiten Sie mit Speichermedien?

Experte:

Wir sind derzeit noch im Aufbau begriffen. Deswegen gibt es - sowohl was die Daten betrifft, die aufzubewahren sind, als auch für die Infrastruktur oder die Werkzeuge, die wir benutzen - wenig Definitives. Und das liegt auch daran: Es gibt diesen Paragrafen aus dem Standortauswahlverfahren, der § 38. Und da lautet ja der zweite Absatz sinngemäß: "Alles Weitere regelt eine Verordnung." Wir warten noch auf diese Verordnung und wir bereiten uns natürlich darauf vor. Wir bereiten gerade eine Ausschreibung für ein Langzeitarchivierungssystem vor. Deswegen kann ich Ihnen jetzt noch nicht sagen, wir benutzen als Instrument oder als Infrastruktur diese Werkzeuge, diese Produkte, diese Anbieter. Denn das befindet sich noch im Aufbau und die Ausschreibung wird vorbereitet. Das andere: Das, was wir meistens Speicherdaten nennen, weil das auch der Begriff ist, der im Gesetz verwendet wird - also für die Teilaufgabe der Langzeitarchivierung die digitalen Daten, die wir werden aufbewahren müssen - , die werden wir ja erst bekommen. Aber da rechnen wir damit, dass wir so ziemlich mit fast allem, was man sich vorstellen kann, konfrontiert werden. Bei den Eckpunkten, die für die Verordnung des BMUV [Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz] für die Öffentlichkeitsbeteiligung erarbeitet wurden, lag eine Liste bedeutsamer Daten bei. Die ist nicht abschließend, aber wenn man da reinschaut, dann ist das eher fachlich und inhaltlich bestimmt. Aber man kann daran natürlich schon viel daraus ableiten, was das bedeuten wird. Natürlich werden es zum einen vom Medientyp her die normalen Büroanwendungen sein, Office-Dokumente. Aber da steht auch so etwas drin wie Geodaten. Also Sachen, in denen geologische Sachverhalte ausgedrückt werden. Es

werden auch Baupläne und ähnliches genannt. Und daraus kann man schon ableiten, dass wir mit Formaten, wie sie im Bereich BIM [Building Information Modeling] halt eben vorkommen, konfrontiert sein werden. Dass wir mit den Daten konfrontiert werden, die halt eben bei der geologischen Untersuchung des Untergrunds, z. B. bei Bohrlöchern entstehen. Und dass das natürlich nicht nur als einzelne Datei, sondern auch in Datenbankform uns gegenübertreten wird.

PM:

Das heißt dann nicht nur die Archivierung von einzelnen Dateien, sondern auch von in sich abgeschlossene Datenbanken?

Experte:

Das ist eine gute Frage. Ich vermute, dass es früher oder später darauf hinausläuft. Aber Stand jetzt wissen wir es noch nicht. Dazu kommt natürlich auch: Wenn die Aufgabe in ein paar Jahren losgeht, wird sie natürlich auch viele Jahre und Jahrzehnte andauern. Das heißt, selbst wenn wir wüssten, was der Stand jetzt ist, gibt es ja einen technologischen Fortschritt und wir werden mit neuen Formen konfrontiert werden, die es einfach jetzt noch nicht gibt. Das gehört zur Natur dieser Aufgabe dazu.

PM:

Datenbank-Management-Systeme [DBMS] sind ja auch sehr mit der Zeit gegangen, denke ich. Da stellt sich die Frage: Selbst wenn man jetzt eine gut strukturierte Datenbank wie eine Oracle-Datenbank hat, kann ich die in 50 Jahren noch auslesen?

Experte:

Naja, das ist ein internationales Forschungsfeld, der Erhalt von Datenbanken. Und da ist gearbeitet worden an Exportformaten, die man langfristig gut darstellen kann. Aber das sind dann in vielen Fällen Einzelfallentscheidungen oder szenarioabhängige Entscheidungen, was man denn erreichen will, wie man dann mit diesen Daten bzw. Datenbanken umgeht. Es läuft manchmal unter dem Stichwort "Preservation Planning". Ich finde, bei Webseiten lässt sich das gut nachvollziehen. Webseiten sind auch sehr komplexe Gebilde. Und je nachdem, was der eigene Zweck ist, kann es ganz unterschiedlich sein, wie man eine Webseite aufbewahrt.

Unter Umständen interessiert einen nur der visuelle Eindruck. Und es reicht quasi, einen Screenshot dafür zu machen, für bestimmte Zwecke. In anderen Fällen will man aber nicht nur eben Texte markieren und nutzen können, sondern vielleicht auch Interaktivität erhalten. Und dann stellen sich ganz andere Anforderungen.

PM:

Ja, stimmt. Dafür gibt es ja Tools wie die Wayback Machine.

Experte:

Genau. Es gibt da auch internationale Konsortien, die sich um Standards kümmern, wie man Web Archiving - das ist der Schlüsselbegriff - realisiert. Und da sind Standard-Archivformate beziehungsweise Containerformate definiert worden wie WARC [Web ARChive], das auch von der Wayback Machine zum Beispiel benutzt wird. Die [Formate] werden auch von nationalen Projekten genutzt, die ihr nationales Internet oder bestimmte Teilbereiche aufbewahren wollen und hierfür Werkzeuge wie beispielsweise Crawler oder Speicherformate verwenden. Und so ähnlich stellt sich dann die Frage bei ganz vielen

Dingen. Was ist das eigentliche Erhaltungsziel? Und daraus kann man dann ein bisschen ableiten, welche Maßnahmen man umsetzen muss, um komplexere Objekte aufbewahren zu können.

PM:

Haben Sie - auch wenn sich dieser Bereich bei Ihnen noch im Aufbau befindet - im Bereich Hardware Vorstellungen, ob Sie Server anmieten oder selbst welche hosten? Hier geht es ja auch um Aspekte wie Sicherheit, Zugriffsschutz und Redundanz.

Experte:

Es sind noch nicht alle Entscheidungen diesbezüglich betroffen. Und deswegen ist auch nicht klar, wer wird Dienstleister sein oder was machen wir selbst. Aber es gibt natürlich Standards, wie zum Beispiel vom BSI, dem Bundesamt für Sicherheit in der Informationstechnik, nach denen wir uns richten werden. An vielen Stellen bewegen wir uns in diesen Bereichen. Es gibt so ein paar Prinzipien, die klar sind. Dass wir natürlich nicht nur einen Standort haben werden, wo unsere Daten gespeichert sind, sondern dass die geografisch voneinander hinreichend getrennt sein müssen, sodass man eine Unabhängigkeit von Kopien der Daten erreicht.

Ich unterscheide typischerweise bei der Langzeitarchivierung drei Ebenen. Die erste Ebene ist da, wo man sich darum kümmert, dass die Bits erhalten werden. Also dass es keine Verluste gibt, keine unkontrollierten Veränderungen. Das ist die Bitstream-Preservation. Ein bisschen kann man sich das an den Speichermedien vorstellen: Dass eine Festplatte kaputt geht, soll nicht zu einem Datenverlust führen. Aber es ist auch ein bisschen komplexer, denn es ist ja nicht damit getan, heutzutage eine Festplatte zu haben, sondern man braucht eine ganze Technologiekette, alleine damit das Betriebssystem überhaupt auf diese Bits zugreifen kann. Also das ist sozusagen die unterste Ebene. Da drauf die Ebene wird manchmal ein bisschen unterschiedlich benannt, aber ich würde sagen, da geht es um die technische Nutzbarkeit. Da spielt dann natürlich so etwas rein wie: Habe ich eine Software, mit der ich dieses Datenformat überhaupt noch verarbeiten kann? Oder brauche ich zusätzliche Services? Also die technische Nutzbarkeit. Manchmal wird diese Ebene "Technical Preservation" oder - das ist vielleicht ein bisschen irreführend - "Logical Preservation" genannt. Und die dritte Ebene darüber, die ist jetzt nicht spezifisch digital, aber es gehört halt auch dazu: Kann ich die Daten noch inhaltlich nutzen? Kann ich sie noch verstehen? Habe ich ausreichende Kontextinformation? Das sind drei Ebenen, um die man sich in der Langzeitarchivierung kümmert. Und diese Frage nach Hardware, Server und Speichermedien ist für mich eher auf dieser Bitstream-Preservation-Ebene angesiedelt. Zumindest wenn es um die Storage-Ebene geht.

PM:

Ja, das kann ich gut nachvollziehen. Das heißt, man müsste auch Software vorhalten, mit der man gespeicherte Daten auch öffnen kann.

Experte:

Genau, das wäre zumindest eine Strategie. Eine Diskussion, die heutzutage gar nicht mehr so richtig aktuell ist, aber die man vor einigen Jahren geführt hat - ich würde fast sagen, in der Anfangszeit, als man angefangen hat, sich konzeptionell mit diesem Problem der Langzeitarchivierung auseinanderzusetzen -, ist die Frage auf dieser Ebene der technischen Nutzung: Migriert man Dateien, digitale Objekte von einem Format in ein anderes, bevor man das eine Format nicht mehr mittels einer Software unterstützen kann? Oder emuliert man alte Umgebungen und die Software, damit man sie weiter betreiben kann, damit man das Objekt nicht anfassen muss? Ich würde sagen, das hat sich

inzwischen ein bisschen als Nebenschauplatz erwiesen, der gar nicht so oft relevant ist. Aber es gibt natürlich Fälle, die diese Fragen sehr plastisch machen. Es gibt in Berlin das Computerspielmuseum zum Beispiel.

PM:

Das kenne ich, da war ich schon mal drin.

Experte:

Ja, genau. Und für die ist das eine ganz harte Frage, wie sie sie die Nutzbarkeit von alten Computerspielen erhalten. Da spielt Emulation natürlich auch eine große Rolle. Insofern ist es ein dankbarer Bereich, denn es gibt viele engagierte Fans, Hobbyleute - Retrogaming ist ja ein eigener Begriff oder Bereich geworden - , die sich um Emulation kümmern, damit man auch auf neuen Hardwaretechnologien ein möglichst authentisches Spielerlebnis der alten Spiele hat. Wenn ich auf Ihren Leitfaden gucke, dann sehe ich, dass Sie ein bisschen diese erste Frage, wer die Stakeholder bei der Frage, was gespeichert wird, sind, übersprungen haben. Ich bin fast ein bisschen dankbar dafür, weil es eine komplizierte Frage ist.

PM:

Das will ich kurz einleiten. Ich hatte die Frage auch [dem Historiker der Gruppe Wissensmanagement der BGE] gestellt, der das Standortauswahlverfahren zeithistorisch begleitet. Wir haben diskutiert, dass es eine Art Gatekeeper geben muss, der entscheidet, welche Informationen erhaltenswürdig sind. Er meinte daraufhin, dass in bestimmten Bereichen tatsächlich er selbst dieser Gatekeeper ist. Insofern ist diese Frage spannend, wer eigentlich Gatekeeper und Stakeholder sind.

Experte:

Ich muss erstmal die Frage zurückgeben: Was verstehen Sie unter Stakeholder?

PM:

Gut, Stakeholder ist erstmal sehr weit gefasst. Also alle, die irgendeinen Einfluss haben, irgendeine Meinung...

Experte:

Heutzutage haben sehr viele Menschen eine Meinung.

PM:

Das ist richtig. Aber klar, auch die Öffentlichkeit ist ein Stakeholder. Es stellt sich die Frage, wie relevant dieser Stakeholder für Sie ist.

Experte:

Natürlich ist der relevant. Zum Ziel der Langzeitdokumentation ist die gesetzliche Formulierung in § 38 gar nicht so klar und fest und in der Verordnung wird es hoffentlich noch etwas präziser. Aber eine andere Formulierung, die aus dem Bereich des OECD/NEA-Forschungsprojekts kommt, was Sie [in Ihrem Interviewleitfaden] auch erwähnen, ist, zukünftigen Generationen eine eigene Entscheidung zu ermöglichen. Das kann man in drei Teilfragen übersetzen. Erstens: Zukünftige Generationen sollen wissen, was sich in einem Endlager befindet und wie die Sicherheit davon einzuschätzen ist. Zweitens:

Wieso es da ist - Wie war der Prozess, wie kam es dazu? Nachvollziehbarkeit. Und drittens: Was ist zu beachten, wenn man an diese eingelagerten Abfälle wieder ranmuss oder wenn man eine Bergung vornehmen muss? Das sind so ein bisschen die inhaltlichen Fragen, anhand derer wir überlegen, was wie gespeichert werden muss.

Und "zukünftige Generationen" ist natürlich ganz breit. Natürlich ist die Öffentlichkeit ein Stakeholder und wir haben da keine feste Klassifikation derzeit, die ich Ihnen präsentieren kann. Was ich sagen kann, ist ein bisschen, wie wir in Bezug auf ein Langzeitarchivsystem unterschieden haben oder unterscheiden. Es ist natürlich noch zu diskutieren, wie viel Einfluss dann wer hat bei der Frage, was wie aufbewahrt wird. Wir haben zum einen die Inhaber der Speicherdaten. Wir haben "Inhaber weiterer relevanter Daten", so nennen wir das. Das hängt damit zusammen, dass manche Sachen vielleicht nicht Speicherdaten sind, aber sie sind möglicherweise trotzdem wichtig zum Verständnis der Speicherdaten.

PM:

Was können das für Daten sein? Was kann man sich darunter vorstellen?

Experte:

Nun ja, nehmen wir an, wir haben Bohrlochdaten. Und die werden als bedeutsam befunden. Aber um die zu nutzen, brauche ich ja weitere Informationen. Und das sind diese weiteren relevanten Daten. Also die Software wäre vielleicht ein Beispiel dafür. Oder Kontextinformationen, Handbücher. Wir werden sicherlich auch andere "Erinnerungsakteure" haben, vielleicht so etwas wie andere Bibliotheken oder Archive, die dort in der Region zum Beispiel sind, die selber Daten auch gesammelt haben oder auch nutzen wollen. Oder auch Personen aus dem Bereich Medien, Kultur, Forschung und Bildung. Wir erleben jetzt zum Beispiel auch öfter Anfragen von Dokumentarfilmern oder Journalisten. Und die haben auch ein Informationsinteresse. Jetzt fragen sie uns nach Interviews und in Zukunft fragen sie uns wahrscheinlich: "Wie sieht denn das Endlager aus?" Also das sind einige vorläufige Unterscheidungen von Akteure. Man kann noch nicht erwarten, dass sie dauerhaft unverändert und für alle Analysen oder Entscheidungen benutzt werden. Aber im Rahmen eines Langzeitarchivs, eines technischen Systems, spielen sie derzeit eine Rolle.

Wir haben die interessierte Öffentlichkeit, die zum Beispiel ein Auskunftsgesuch stellen kann. Das steht allen Menschen offen. Es gibt das Umweltinformationsgesetz, Sie fragen ja später auch nach den gesetzlichen Grundlagen. Also insofern ist potenziell jede Person der Öffentlichkeit relevant. Es gibt auch andere Akteure in verantwortlicher Rolle, vielleicht in Zukunft andere staatliche Stellen, die Informationen benötigen und denen wir Auskunft geben müssen. Es gibt bestimmte externe Systeme, also Datenportale zum Beispiel. Es gibt das sogenannte Geoportal...

PM:

Ja, jedes Bundesland hat so etwas. Für Hessen kenne ich das. Da kriegen Sie dann irgendwelche City-GML-Modelle, also Kubaturen von Gebäuden für ganz Hessen her und weitere Daten.

Experte:

Es gibt solche Art von Systemen, für die wir als Datenquelle wahrscheinlich irgendwann dienen müssen oder von denen wir auch vielleicht Daten haben wollen. Die BGE setzt ja auch technische Systeme ein, aus denen wir Daten erhalten wollen. Diese externen IT-Systeme sind halt eben ein Punkt. Und dann haben wir natürlich auch intern verschiedene Akteure. Unter Umständen gibt es inhaltliches Interesse an den Daten innerhalb des BASE durch andere Abteilungen. Aber andere Abteilungen sind auch selber Produzenten von Speicherdaten. Also das BASE ist selber jemand, der Speicherdaten produziert. Wir

sind selber in der Produzenten- und Konsumentenrolle. Und dann haben wir nochmal uns als Team und die IT, die vielleicht kein inhaltliches Interesse hat, wonach entschieden wird, ob die Daten aufbewahrt werden. Aber in dem Moment, wo es plötzlich um sehr hochvolumige Daten geht, werden natürlich auch Fragen relevant wie: Haben wir genug Speicherplatz? Wie groß ist unser Breitbandanschluss? Müssen wir den irgendwie ausweiten? Insofern: Das sind eine ganze Reihe von Akteuren, mit denen wir interagieren und wo unterschiedliche Perspektiven bestehen. Aber wenn die Frage ist: „Was wird archiviert?“, dann haben wir in unserer Zuarbeit zu den Eckpunkten der Verordnung halt eben versucht, diesen noch etwas vagen Begriff der potenziell bedeutsam Daten zu konkretisieren, um klarzumachen, dass das Daten sind, um die es geht. Und um die sachlichen Kriterien für die Auswahl der Speicherdaten zu illustrieren.

PM:

Wir können uns ja mal ein Szenario ausdenken, wenn es darum geht, Dinge dauerhaft zu speichern. Wie robust ist denn diese ganze Speicherung, diese Entscheidungsgewalt darüber, in Hinblick auf politische Prozesse? Es ist ja denkbar, dass irgendjemand mal sagt: "Naja, eigentlich brauchen wir das alles hier gar nicht mehr, wir können alles wieder löschen." Ist so etwas theoretisch denkbar?

Experte:

Wir sind ein Bundesamt und haben als solche gesetzliche Aufgaben. Wenn es Gesetz ist, dann müssen wir es umsetzen. Insofern ist die Frage: Wie werden die Gesetze gemacht? Und was da denkbar ist, darüber kann ich hier nicht spekulieren. Ich würde halt Vergleiche machen. Bibliotheken, Archive, Museen haben immer eine gewisse Unabhängigkeit. Gerade Archive dürfen selber darüber entscheiden, was aufbewahrt wird und was nicht. Und das ist gerade gedacht als Instrument der demokratischen Kontrolle. Also diejenigen, die die Unterlagen und die Akten produzieren, sollen nicht selber entscheiden, ob das aufbewahrt wird oder nicht, denn sie haben möglicherweise ein bestimmtes Interesse. Und klassische Archive sind genauso gebaut worden, dass die Entscheidung beim Archiv liegt.

Dazu sind auch etliche Dokumente und Richtlinien entstanden, wo das halt eben festgehalten wird. Bei einer Konferenz habe ich mal einen Kollegen erlebt, der aus dem Archivbereich kam und gesagt hat: "Bei uns wird Demokratie verteidigt" oder "Archive sind die Verteidigung der Demokratie". Ich würde wahrscheinlich ergänzen: "... auf eine sehr langfristige Perspektive". Aber weil wir im Aufbau sind, kann ich Ihnen jetzt nichts Konkretes sagen. Aber ein sehr wichtiges Konzept im Bereich digitaler Langzeitarchivierung ist das der vertrauenswürdigen digitalen Langzeitarchive. Dazu gibt es sogar Standards. Es gibt einen nestor-Kriterienkatalog für vertrauenswürdige Langzeitarchive [nestor ist ein deutsches Kompetenznetzwerk für die Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen]. Es gibt auch einen ISO-Standard. Und darin gibt es neben den vielleicht naheliegenden technischen Aspekten und den Arbeitsabläufen auch eine Vielzahl von organisatorischen Kriterien, wo man sagt: "Liebes Archiv, wenn du als vertrauenswürdiges digitales Langzeitarchiv gelten willst, dann musst du deine Finanzplanung so dokumentieren, dass auch die Öffentlichkeit oder wer auch immer, um den es geht, Vertrauen in deine finanzielle Nachhaltigkeit haben kann." Also diese organisatorischen Nachhaltigkeitsfaktoren, so nenne ich sie mal, sind ein großer Aufgabenbereich in diesen Standards für vertrauenswürdige digitale Langzeitarchive.

PM:

Geht es bei Ihnen dann nur um die Archivierung, um die Speicherung, oder geht es auch darum, eine Art Informationsportal zu schaffen? Sie haben ja gesagt, dass Sie auch mit Auskunftsanfragen rechnen. Gibt

es Informationen, von denen Sie sagen, dass sie primär mit dem Zweck, sie auch offen für andere zur Verfügung zu stellen, gespeichert werden? Ich sehe hier verschiedene Ansätze.

Experte:

Auch das ist nicht entschieden. Als Bundesamt richten wir uns nach Gesetzen und der Zugang zu den Informationen, die wir haben, wird weitgehend durch dieses Umweltinformationsgesetz geregelt. Das heißt, natürlich können Leute Anträge stellen und im Umweltinformationsgesetz sind auch Paragraphen enthalten, die eine proaktive Informationsbereitstellung der Verwaltung bzw. der Behörden vorsehen. Welche Form das annehmen wird, kann ich Ihnen noch nicht sagen. Aber entsprechend dieser Gesetzesgrundlage fände ich es total plausibel zu sagen, dass Informationen, die über ein Portal recherchierbar sind, bereitgestellt werden. Wir müssen an der Stelle aber auch unterscheiden, denn wir werden auch sehr sicher einzelne Daten übernehmen, die Geschäftsgeheimnisse enthalten und bei denen wir erstmal verpflichtet sind, sie auch vertraulich zu behandeln.

PM:

Natürlich auch personenbezogene Daten.

Experte:

Ja, auch personenbezogene Daten. Und die wird es natürlich geben und sie müssen auch zu einem gewissen Grad drin sein. Wie weit, das müssen wir dann in den jeweiligen Fällen sehen. Archive haben aber auch üblicherweise genau dafür besondere Regeln, was den Datenschutz angeht. In dem Sinne, dass bestimmte Rechte aus den Datenschutzgesetzen auch beispielsweise nicht greifen. Also ich kann nicht zum Bundesarchiv hingehen und sagen: "Ich möchte, dass Sie alle Informationen, in denen ich genannt werde, löschen oder mich aus allen Akten rauslöschen." Der Zweck eines Archivs ist halt eben diese Informationssicherung. Aber auch da gibt es natürlich Grenzen. Das Archiv muss es nicht löschen, aber bestimmte Informationsweitergaben sind dann vielleicht trotzdem limitiert. Datenschutzgesetzte spielen also eine Rolle. Gesetze für Urheberrecht und Geschäftsgeheimnisse spielen zum gewissen Grad eine Rolle. Und das ist das Feld, in dem wir uns bewegen. Sie hatten ja auch explizit die Frage, welche Gesetze und Normen betroffen sind.

PM:

Ja genau.

Experte:

Ich habe versucht, mir das mal ein bisschen systematisch rauszusuchen. Aber es ist gar nicht so leicht zu sagen. Also - Es gibt die Archivgesetzgebung des Bundes und der Länder, es gibt das Umweltinformationsgesetz, sehr ähnlich dann das Informationsfreiheitsgesetz und die Datenschutzgesetze. Es gibt dann das Geologiedatengesetz, der Umgang mit Geologiedaten ist dann noch mal speziell. Zu den Sachen mit den Geschäftsgeheimnissen sage ich jetzt lieber nichts, weil ich nicht genau weiß, wo sich das befindet. Aber Urhebergesetz, die Aufbewahrungsfristen, denen Unternehmen unterliegen... Es gibt das E-Government-Gesetz und das Datennutzungsgesetz, die auch nochmal sagen, dass öffentliche Einrichtungen zur Verfügungstellung von Daten verpflichtet sind. Das war bisher eher die Seite der Informationsweitergabe und wir haben dann natürlich auch die Datendokumentations- und Erhebungsregeln, die dann auch ein bisschen mehr auf die kerntechnischen Anlagen eingehen. Also neben dem Atomgesetz, dem Standortauswahlgesetz - hier wird voraussichtlich

auch viel in der entsprechenden Verordnung ausgeführt werden - gibt es dann auch noch die KTA 1404. Ich weiß nicht, ob die Ihnen schonmal begegnet ist.

PM:

Vom Begriff her, ja. Den genauen Inhalt kenne ich nicht.

Experte:

Das ist die Dokumentation beim Bau und Betrieb von Kernkraftwerken. Für diese Art von Verordnungen bin ich zwar kein Experte. Aber es gibt dann auch so etwas wie die Abfallflusskontrollrichtlinien und ESK-Zwischenlagerleitlinien [ESK = Entsorgungskommission] und natürlich die internen Regelungen. Also das sind so Dinge, da sind dann insbesondere die von uns sogenannten Inhaber von Speicherdaten verpflichtet, Dokumentation zu führen und anzulegen. Und das sind Daten, die wir voraussichtlich erhalten werden.

PM:

Das war auch eine Frage, die ich mir aufgeschrieben hatte, also die Frage nach der Unterscheidung zwischen Aufbewahrungs- und Übergabepflichten. Vielleicht können Sie das einmal klarstellen: Einerseits geht es ja um die Übergabe, aber auch um die parallele bzw. redundante Archivierung für einen gewissen Zeitraum. Es wird also danach auf der einen Seite etwas gelöscht und auf Ihrer Seite behalten?

Experte:

Genau. Also die Verordnung wird das voraussichtlich genauer ausführen, das war zumindest unser Anspruch bei unserer Zuarbeit. Nach einer gewissen Zeit erhalten wir Daten, die potenziell bedeutsam sind. Also Daten zur End- und Zwischenlagerung, die bedeutsam sind oder es werden können. Die werden wir von den sogenannten Inhabern von Speicherdaten erhalten. "Inhaber von Speicherdaten" wird in der Verordnung definiert sein. Das wird nicht jeder beliebige Akteur sein können, sondern sie werden anhand von anderen Gesetzen definiert. Aber natürlich klarerweise zum Beispiel die Bundesgesellschaft für Endlagerung (BGE) und die Bundesgesellschaft für Zwischenlagerung (BGZ). Das sind zwei sehr große [Akteure]. Und wenn die Daten zum Beispiel abgeschlossen sind, sind sie uns anzubieten und wir übernehmen sie. Bis dahin dürfen die Inhaber von Speicherdaten diese Speicherdaten aufgrund der Langzeitdokumentation nicht löschen. Nachdem die bei uns angekommen sind und wir den Erhalt bestätigen, machen wir natürlich auch Integritätstests und führen eine Qualitätskontrolle durch, um durchzutesten, dass wir die Daten korrekt erhalten haben und dafür die Verantwortung übernehmen können.

Ab dem Moment haben wir keinen Einwand mehr gegen die Löschung. Aber es kann natürlich sein, dass der Inhaber von Speicherdaten aus anderen Gründen - Steuerunterlagen oder was auch immer - die Daten noch aufbewahren muss. Das können wir nicht beurteilen. Und das können wir den Inhabern von Speicherdaten auch nicht abnehmen. Das geht rechtlich nicht. Also wir können sozusagen nicht der Archivdienst für die Inhaber von Speicherdaten sein, denn das würde bedeuten, wir würden auch die Verantwortung für diese anderen gesetzlichen Verpflichtungen oder vertraglichen Verpflichtungen, die sie eingegangen sind, übernehmen. Aber es ist ein Fall denkbar, wo halt eben die Daten im Endeffekt nur noch für uns aufbewahrt werden, wir sie dann übernehmen und wir sagen dann: "Jetzt dürft ihr sie löschen". Und dann löschen sie als Inhaber von Speicherdaten das auf ihrer Seite und wir sind dann die alleinige Quelle der Daten.

PM:

Wie schaut es aus mit der Aktualität der Daten? Wenn wir irgendwann mal ein Endlager für die hochradioaktiven Abfälle haben - Werden sie dann kontinuierlich mit Daten beliefert oder in Paketen? Wie muss man sich das vorstellen?

Experte:

In der Diskussion für die Verordnung sind jährliche Stichtage, an denen wir Listen mit Listen erhalten und auf Basis dessen dann jeweils Daten zu uns fließen. Es kann auch sein, dass es da auch zu Aktualisierungen kommt. Und manche Daten werden ja auch fortgeschrieben...

PM:

Genau, das ist ja die Frage. Wenn man ein bestehendes Systeme hat, wo man Dinge erweitert, aber die Datei hat immer noch denselben Namen und ist nur ein bisschen größer geworden.

Experte:

Genau. Also an ein paar Stellen hängt es dann wirklich von der Verordnung ab, aber eine Überlegung wäre, welche Daten in diesen jährlichen Übersichtslisten anzubieten sind. Eine Möglichkeit wäre, dass es sich um abgeschlossene Daten handelt, die im letzten Jahr abgeschlossen wurden. Und da wäre das ja per se ausgeschlossen, dass die dann nochmal aktualisiert werden, denkt man. Andererseits: Im Nachhinein wird manchmal ja aktualisiert. Und umgekehrt: Ein Anspruch, so würden wir es derzeit formulieren, ist ja die Nachvollziehbarkeit des Verfahrens. Und das bedeutet, dass auch Zwischenversionen, die für einen bestimmten veröffentlichten Bericht oder für Entscheidungen benutzt wurden, wichtig sind, damit man versteht, wie denn der Stand zu diesem Zeitpunkt war.

PM:

Also eine Art Versionierung wie in der klassischen Softwareentwicklung?

Experte:

Genau, es wird eine Versionierung geben müssen. Also das ist klar. Aber die Details kann ich Ihnen jetzt nicht sagen, zumal eben auch diese Verordnung noch nicht da ist. Und da soll genauer geregelt sein, wann was anzubieten ist.

PM:

OK. Wir haben jetzt schon Aufbewahrungs- und Übergabepflichten angesprochen. Es gab ein Forschungsprojekt [Langzeitbeständigkeit von Paper bzw. "Labest Papier"], da ging es um die Langzeitbeständigkeit von Papier. Und da wurde als Zielmarke 500 Jahre angegeben. Jetzt haben Sie schon gesagt, dass Langzeitarchivierung eine dauerhafte Archivierung meint. Mich interessiert hier: Wie kam man auf 500 Jahre, wenn das Ziel ja eigentlich ist, viel länger zu speichern?

Experte:

Es gibt ein benanntes Datum. Das heißt, die radioaktiven Abfälle sollen 500 Jahre bergbar sein. Und das ist zumindest ein klar identifizierbarer Meilenstein, wo man sagen kann: OK, damit kann man auf alle Fälle testen. Da kommen nochmal andere Faktoren hinzu, aber wenn wir von einem Worst Case ausgehen, dass sich keine Organisation um die Unterlagen kümmert, dann sollte das Papier zumindest diese 500 Jahre überstehen. Wenn Sie die 500 Jahre lesen - ich kann Ihnen nicht den Paragraphen im

Standortauswahlgesetz nennen - dann ist der Grund üblicherweise, dass es diesen Paragraphen gibt, wo drinsteht, dass 500 Jahre lang eine Bergbarkeit gewährleistet sein soll. Das Endlager soll so geplant werden, dass die Entscheidung, die wir jetzt treffen - so gut es eben geht -, auch von zukünftigen Generationen revidiert werden können. Und dazu gehört halt eben die Bergbarkeit der Abfälle nach 500 Jahren.

PM:

Der Vorteil von Papier ist ja: Ich kann das in einen Raum einsperren und 500 Jahre drin lassen. Mit digitalen Daten braucht es ja eine Infrastruktur, die auch gewartet wird. Gut, man könnte vielleicht auch alles auf eine Festplatte packen und mal gucken, wie lange die hält.

Experte:

Bestimmt keine 500 Jahre. Jetzt wird es etwas kompliziert. Weil wir haben ja die Aufgabe, Informationen zu erhalten. Und es ist jetzt nicht primär die Frage, ob analog oder digital. Aber natürlich entsteht ein Großteil aller Informationen heute digital. In der Phase, in der wir uns jetzt zuerst befinden, entstehen diese Informationen. Wir sammeln sie ein und versuchen sie aufzubewahren, zu erschließen, die Nutzbarkeit zu erhalten und sie auch wieder zur Verfügung zu stellen. Wie lange tun wir das? Das hängt von ganz vielen Faktoren ab. Im Idealfall, sag ich mal, für immer. Und dann wäre der normale Betrieb einfach, dass man Systeme, Speichermedien, Hardware etc. natürlich aktualisiert. Und so kann man dann halt eben die Nutzbarkeit langfristig auch erhalten, indem es einfach immer eine Institution gibt, die sich darum kümmert. In diesem OECD/NEA-Projekt "RK&M" - auch wenn es nicht ursprünglich aus diesem Projekt kommt - werden unterschiedliche Zeiten unterschieden, und zwar Short Term, Medium Term und Long Term.

Und Short Term ist - was für die Allgemeinheit ohne diesen Fachkontext erstmal überraschend klingt - bis zum Verschluss des Endlagers. Das ist schon für unsere Verhältnisse ein sehr langer Zeitraum, aber in dem Kontext ist das als Short Term definiert. Medium Term - ich weiß nicht mehr die exakte Formulierung - bedeutet sinngemäß solange, wie sich eine Institution darum kümmert und das beaufsichtigt. Und Long Term ist dann der Zeitraum, wo das Endlager nicht mehr beaufsichtigt wird. So ist, glaube ich, die Formulierung. Und ich finde, es hat eine gewisse Grundplausibilität zu sagen: Wenn das Endlager nicht mehr beaufsichtigt wird, dann wird auch das Archiv möglicherweise dazu nicht mehr beaufsichtigt. Und das würde bedeuten, dass ein digitales Archiv Medium Term funktioniert, weil es halt noch eine Institution gibt, die sich darum kümmert, aber nicht Long Term. Was macht man dann? Das steckt dann zu einem guten Teil in diesem von Ihnen auch zitierten Forschungsprojekt drin, dass man eine Vielzahl von Maßnahmen bereithält, um eine Rekonstruierbarkeit, eine Nachvollziehbarkeit zu ermöglichen. Darauf ist man dann eben angewiesen. Ich glaube, eine total plausible Annahme ist, zu sagen: Wir müssen auch erstmal digital arbeiten und so arbeiten wir auf alle Fälle erstmal in der nächsten Zeit. Und zu verschiedenen Zeitpunkten überführen wir einen Teilbestand unserer digitalen Informationen auf besonders langfristige haltbare Speichermedien oder möglicherweise auch einfach auf besonders robustes Papier. Sie hatten es auch zitiert, wir haben auch Forschungsprojekte zur Langzeitbeständigkeit von Papier. Und mit Papier kann man auch ein paar tausend Jahre überbrücken.

PM:

Ja, ich denke Papyrus hält sich eine Weile.

Experte:

Genau, hängt von vielen Faktoren ab. Meine persönliche Meinung: Wir müssen - Stand dessen, wo wir uns gerade befinden - keine Entscheidung für Speichermedien heute treffen. Aber wir forschen daran, weil wir als BASE, beziehungsweise unsere Nachfolgerinnen und Nachfolger, die sich darum kümmern, müssen irgendwann eine Entscheidung für Speichermedien treffen. Und deshalb betreiben wir Forschung zu diesem Zweck, damit wir irgendwann auf dem dann vorliegenden Stand der Wissenschaft und Technik das bestmögliche Speichermedium und die bestmögliche Speicherart - auch Codierung spielt da eine Rolle - auswählen können.

PM:

Jetzt wird ja in dieser Tabelle [aus dem Final Report des OECD/NEA-Forschungsprojekts "Preservation of Records, Knowledge and Memory (RK&M)] Langzeitarchivierung nur als eine Möglichkeit beschrieben. Ich wundere mich - korrigieren Sie mich gerne -, dass sich diese anderen Approaches [wie Marker und Bildung] weicher anhören. Mir kommt es so vor, als seien hier noch keine gesetzlichen Grundlagen oder Verpflichtungen da. Zusammengefasst scheint sich dieses regulatorische Gerüst meiner Meinung nach sehr um diese Datenspeicherung und Archivierung zu drehen, aber weniger um diese anderen Möglichkeiten.

Experte:

Meines Wissens ist es so, dass wir in Deutschland keine Gesetzgebung zur Markierung haben, also Markers. Aber das ändert sich vielleicht irgendwann. Das ist auch etwas, was wir halt aus fachlicher Sicht sehen, dass es da Handlungsbedarf gibt. Und es ist, glaube ich, etwas, wozu dann auch Diskussionen in zukünftigen Jahren zu führen sind. Generell: Was - glaube ich - der wichtige Wert dieses RK&M-Berichts ist, ist die generelle Strategie, zu sagen: Es reicht nicht, einen Mechanismus, einen Approach zu haben. Sondern man braucht mehrere, die miteinander verzahnt sind, weil über so einen langfristigen Zeitraum mit unvorhersehbaren Ereignissen kann man den Informationserhalt am besten dann sichern, wenn viele Dinge sich gegenseitig verstärken. Wenn man quasi ein Netz aus Maßnahmen hat, die aufeinander verweisen.

PM:

In der Schweizer Verordnung [ENSI-G09 "Richtlinie für die schweizerischen Kernanlagen - Bau- und Betriebsdokumentation"] ging es speziell um die Betriebsdokumentation für die Kernanlagen. Sie sind jetzt schon auf die KTA 1404, das wäre ja das deutsche Pendant, zu sprechen gekommen...

Experte:

...Vielleicht.

PM:

Ja OK, den Inhalt werde ich nochmal checken. Dort [in der ENSI-G09] wurde ja auch der Aspekt mit Aufbewahrungs- und Übergabepflichten aufgemacht. Es wird PDF-A als bevorzugtes Dokumentenformat genannt. Auch in gegenwärtigen Dokumentenmanagementsystemen sind PDFs ja überall anzutreffen. Denken Sie, dass man mit dem PDF-Standard gut für eine Langzeitarchivierung aufgestellt ist?

Experte:

Ich kenne die Diskussion der Langzeitarchivierungs-Fachcommunity zum Thema PDF. Das ist auch ein Containerformat. Da kann sich sehr, sehr viel drunter verstecken. Und PDF-A ist schon mal ein Fortschritt. So würde ich es formulieren. Das heißt nicht, dass man damit auf der Formatebene die

Langzeitarchivierungsaufgabe gelöst hat. In vielen Bereichen ist es ein Fortschritt, aber auch da gibt es nochmal Unterversionen. Dazu bin ich jetzt aber nicht im Detail sprechfähig. Es ist gut möglich, dass wir das [Format] in vielen Fällen verwenden werden. Genauer kann man das nicht sagen. Es gibt unterschiedliche Versionen von PDF-A und man kann auch im PDF-A nicht so schöne Dinge anstellen. Wenn Sie wirklich das interessiert, würde ich Sie an Expertinnen an der TIB Hannover, der Technischen Informationsbibliothek Hannover, verweisen. Da haben wir einige Personen, die wirklich mit dem Hex-Editor in das PDF reinschauen...

PM:

Interessant, aber lassen Sie uns einmal bei der Richtlinie [ENSI-G09] bleiben. Dort wird wieder verwiesen auf die ISO 14721...

Experte:

...und das OAIS-Modell [Reference model for an open archival information system].

PM:

Genau. Ich kannte dieses Modell zumindest vom Namen her aus unseren Grundlagenvorlesungen im Bereich Informatik im Bauwesen. Bis jetzt kam mir das wie ein Relikt von vor Jahrzehnten vor, aber es scheint ja doch eine gewisse Aktualität zu haben.

Experte:

Ich muss mal kurz dazwischenfragen. Sie kannten aus Ihrem Studium das OAIS-Modell?

PM:

Ja, im Kontext der Informationsverarbeitung und von Mensch-Maschine. Dort wurde das Modell in der Einleitung mal erwähnt.

Experte:

Denn als Referenzmodell ist es halt eben eines der zentralen Modelle im Bereich der Langzeitarchivierung.

PM:

In Abbildung 4-12 [Information Object Taxonomy] der ISO 14721 ist ja dargestellt, dass es eben nicht nur das "Information Object" gibt, sondern weitere Informationen zur Interpretation und Repräsentation als Komponenten reinspielen müssen. Auch vorhin haben wir ja über Datenbanken und Managementsysteme gesprochen. Ist dies etwas, was Sie sich systematisch anschauen? Macht man sich wirklich die Mühe, für jedes Informationsobjekt im Vorhinein die Pfeiler darunter zu betrachten?

Experte:

Eine Bemerkung zum OAIS: Das ist natürlich ein Referenzmodell. Es ist im Endeffekt zum Vergleich da und es liefert einen Gutteil der Terminologie im Bereich der Fachgemeinschaft der Langzeitarchivierung. Aber es ist nicht ein Umsetzungsstandard. Und dafür sind eigene Standards geschaffen worden wie z. B. Metadatenstandards. Genauer: Wie man die am besten erstellt und was man da verwendet, wie man das festhält. Also ja, natürlich, in der Sprache des OAIS braucht es Representation Information. Unter gewissen Umständen reicht das schon, wenn es die Formatangabe ist, dass man halt sagt, es ist eine

bestimmte PDF-A-Version. Es gibt andere Fälle, wo man mehr Informationen braucht. Und hierfür greift man - aber nochmal: Wir sind im Aufbau und haben keinen fertigen Standard - natürlich auf Standards zurück, die in dieser internationalen Community bestehen. Und ein wichtiger Umsetzungsstandard ist PREMIS [Preservation Metadata Information Standard] von der Library of Congress. Das ist umsetzungsnäher. Diese Fragen sind also relevant. Was die konkreten Umsetzungsmittel sind, das ist jetzt noch nicht entschieden. Das hängt auch zum gewissen Grad davon ab, wie es das Langzeitarchivierungssystem, was wir anschaffen, macht. Aber die unterstützen eigentlich alle PREMIS.

PM:

Meine Frage zielte auch darauf ab, ob die, die Ihnen die Informationen übergeben, auch Sorge dafür tragen müssen, dass Sie diese unteren Bausteine übergeben. Zum Beispiel, dass sie einen Hinweis erhalten, mit welcher Software eine bestimmte Datei geöffnet werden kann. Ich meine, es macht einen Unterschied, ob ich lediglich einen Container mit mir möglicherweise unbekannten Dateitypen erhalte oder zusätzlich auch die Information, mit welcher Software man die auslesen kann.

Experte:

Das ist nicht entschieden, wie viel wir von den Speicherdaten auch nachfragen können. Aber natürlich, es gibt einen Bedarf nach vielen dieser Informationen, also Kontextinformationen sowohl inhaltlicher als auch technischer Art.

PM:

Ja, da bin ich gestolpert über Abbildung A-1 "Composite of Functional Entities" aus der ISO 14721. Dann ich erst mal gesehen, wie umfangreich das Ganze ist...

Experte:

Wie ich sagte: Es ist ein Referenzmodell. Das heißt nicht, dass man jedes Kästchen mit einer einzelnen Person halt eben ausfüllt. Und manche Sachen sind ja auch Maschinenprozesse. Es hilft, an vielen Stellen in diesen großen, grauen Kästchenkategorien Ingest, Data Management, Archival Storage, Access, Preservation Planning sowie Administration zu denken.

Wie dann die einzelnen Funktionen implementiert sind, ist sehr unterschiedlich. Zur Kommunikation auch im Haus benutze ich gar nicht das OAIS, sondern versuche generell nur zu erklären, was Gedächtnisorganisationen an sich machen. Nämlich Daten übernehmen - das ist hier ein bisschen wie der Ingest. Man muss Daten dann erschließen, also man muss halt wissen: Was ist das? Das ist hier so ein bisschen angedeutet im Data Management. Man muss die Daten bewahren. Das ist hier wahrscheinlich Archival Storage und Preservation Planning. Und na ja, wenn es nur da ist, aber niemand darauf zugreifen kann, dann hätte es auch keinen Sinn. Also muss man auch noch eine Art von Zugriff oder Vermittlung haben. Das ist der Access. Also in diesen vier Grundkategorien hilft es auf alle Fälle, darüber nachzudenken und es dann schrittweise präziser zu machen.

PM:

OK. Über den nächsten Punkt haben wir schon ein wenig gesprochen: Migration. In der Norm sind auch ein paar Handlungsempfehlungen gegeben. Fällt Ihnen hierzu noch etwas ein?

Experte:

Nun, in dieser Langzeitarchivierungs-Community sind natürlich weitere Konzepte entwickelt worden. Das OAIS hat da auch eine eigene Terminologie, die jetzt vielleicht gar nicht so die verbreitetste Terminologie ist. Ein Konzept ist das der "Significant Properties". Was sind eigentlich die wesentlichen Eigenschaften, die erhalten werden müssen? Und das ist meistens von der Zielgruppe oder vom Zielnutzungskonzept abhängig. Ein Beispiel nicht aus unserem Bereich, aber trotzdem nachvollziehbar: Beim Computerspielmuseum ist natürlich die Interaktivität eine "Significant Property". Es gibt andere Fälle, wo vielleicht nur der visuelle Eindruck entscheidend ist. Und deswegen ist das so etwas, worüber man sich bei der Frage, wie man die Dinge erhält, viel Gedanken machen muss. Was sind die Prozesse, die ich später unterstützen will, die möglich sein müssen? Was sind die Zielgruppen? Die "Designated Communities" und die "Significant Properties" sind Schlüsselkonzepte in der Langzeitarchivierungstheorie.

PM:

Gibt es bei all diesen Aspekten, die wir jetzt besprochen haben, auch Vorbilder im europäischen Ausland? Ich kenne mich grob mit den technischen Konzepten der Zwischen- und Endlagerung beispielsweise in der Schweiz und Frankreich aus, aber weiß beispielsweise nicht, ob es so ein Konstrukt wie das BASE auch dort gibt. Wissen Sie das und gibt es einen Erfahrungsaustausch und Best Practice-Beispiele? Die stehen ja genau vor den gleichen Herausforderungen.

Experte:

In Bezug auf Langzeitarchivierung digitaler Daten würde ich sagen, dass es keine Best Practice Beispiele gibt, denn das setzt vielleicht schon zu viel voraus. "Best" heißt ja auch schon, es gibt mehrere und das ist dann der Beste bei Best Practice. Es gibt halt eben Practice [im Sinne von Praxis] und Forschung. Und diese Institutionen - das Pendant in Frankreich und in der Schweiz, die ANDRA und die Nagra - mit denen arbeiten wir - nicht ich, sondern eine Kollegin und ein Kollege von mir - in OECD/NEA-Gremien zusammen. Es gibt halt nicht nur dieses RK&M-Projekt, sondern es gibt auch andere Arbeitsgruppen und in denen wird mit den Pendants oder halbwegs vergleichbaren Institutionen aus dem Ausland zusammengearbeitet und solche Fragen diskutiert. Das ist aber auch da weiterhin ein aktuelles Forschungs- und Diskussionsthema.

PM:

Auf welchem Stand stehen diese Institutionen und wie weit ist das BASE?

Experte:

[Wir stehen gut da.] Ich war letzte Woche bei der iPRES-Konferenz [International Conference on Digital Preservation] und da hat ein Projektpartner der ANDRA Ergebnisse eines Forschungsprojekts vorgestellt, wie sie auf Papier digitale Daten kodiert festhalten. Das betraf also das Speichermedium Papier, aber auch, wie man das für digitale Sachen effizient nutzen kann. Und ich würde sagen: Ja, die unterschiedlichen Länder setzen unterschiedliche Schwerpunkte, aber im Wesentlichen hat man einen ähnlichen Stand. Manche haben schon eine eigenständige Archivinstitution. In Großbritannien ist das zum Beispiel Nucleus, ganz im Norden von Schottland. Die haben ein eigenes Gebäude mit Magazinen. Viele andere Länder sind nicht so weit, dass sie von sich sagen können, dass sie jetzt eine eigenständige Archivinstitution haben. Also: Der Stand ist unterschiedlich, würde ich im Wesentlichen sagen. Niemand ist fertig, hat ein vollständiges Konzept, aber es wird an verwandten Fragestellungen international gearbeitet. Und man tauscht sich halt eben auch aus.

PM:

Gibt es Austausch über Europa hinaus?

Experte:

Ja, aber ich bin nicht der Experte dafür. Ich kann Ihnen nicht sagen, wie die Situation in den USA ist in Bezug auf Informationserhalt für Endlagerung. Viele dieser sehr öffentlichkeitswirksamen und fantasieanregenden Ideen kamen ursprünglich aus einem Projekt aus den USA, wo es darum ging, wie weist man die Zukunft darauf hin, dass man hier nicht nach Rohstoffen buddeln sollen...

PM:

...ja, Atomsemiotik und dass man in der Landschaft Pfeiler aufstellt.

Experte:

Genau. Vieles davon stammt aus einem Projekt der USA. Aber ich bin keine Experte dafür.

Experte:

Aber mein Eindruck ist, weil einfach in Europa viele Länder mit Atomkraftnutzung auf engem Raum zusammenkommen, ist hier vielleicht auch die Interaktion und die Erforschung dieses Themas momentan am weitesten.

PM:

Das bietet sich einfach an.

Experte:

Ja, genau. Ja, es gibt auch in anderen Ländern und Kontinenten Personen und Institutionen, die sich damit beschäftigen. Mir sind jetzt - vielleicht ist es Zufall, aber ich vermute, es hat auch ein bisschen mit der Schwerpunktsetzung und diesen Rahmenbedingungen zu tun - im Wesentlichen die europäischen Länder bekannt.

PM:

OK. Dann sind wir jetzt auch schon beim dritten Komplex [des Interviewleitfadens]. Hier geht es zunächst um eine Passage aus dem Forschungsbericht zum Forschungsprojekt Labest Papier. Dort wurde ja darauf verwiesen, dass es eine Art Dreiklang gibt für die Langzeitarchivierung, bestehend aus Papier, digitaler elektronischer und optischer Archivierung. Und es wurde darauf verwiesen, dass sich das so in vielen Normen wiederfindet. Bei uns wird das Single Source of Truth-Prinzip immer sehr hochgehalten, wenn es darum geht, mit Datenbanken zu arbeiten. Auch bei BIM streben wir dieses Prinzip an. Bei besagtem Forschungsprojekt dagegen setzt man auf Redundanz, was aber auch Vorteile mit sich bringt. Können Sie für die Langzeitarchivierung diskutieren, wie man über diese beiden ganz unterschiedlichen Konzepte nachdenkt?

Experte:

Ich kann nicht für diese Passage und das Forschungsprojekt sprechen. Dafür kenne ich es nicht gut genug. Aber ein wichtiger Unterschied, den ich sehe, zwischen dem Fall, den Sie vorgestellt haben und dem Fall in diesem eher archivarischen Kontext ist natürlich, dass im archivarischen Kontext die Daten abgeschlossen sind. Single Source of Truth erscheint mir besonders wichtig, wenn sich Daten verändern,

um sicherzustellen, dass keine Inkonsistenzen auftreten. Diese Gefahr wird, glaube ich, bei diesem Bereich der Archivierung nicht mehr gesehen, denn es sollte ja niemand an den Daten etwas verändern. Das würde auf ein schwerwiegendes Problem hinweisen oder es müsste deutlich gemacht werden, dass hier eine Veränderung festgestellt wurde. Wir kommen unserer Funktion nach, indem wir auf solche Probleme und Veränderungen prüfen und aufmerksam werden. Und viel stärker ist der Aspekt der Bewahrung, dass man sagt: "OK, wenn wir eben nur auf ein Mittel setzen, dann kann es systematische Probleme geben, die den Bestand der Information gefährden."

Bei Single Source of Truth könnte man ja auch sagen, dass im ganz harten Sinne die Information ja auch dann nur an einem Ort gespeichert sein muss, weil wenn man an zwei Orten die Datenbank gespiegelt hat, stellt sich die Frage, wem man denn bei einer Inkonsistenz glauben soll. Die Denkweise hier ist, dass diese Inkonsistenz eigentlich nicht in der regulären Bearbeitung entstehen kann, weil die Daten üblicherweise als abgeschlossen betrachtet werden. Und deswegen ist die geplante Redundanz über unterschiedliche Arten der Speicherung, unterschiedliche Orte, möglicherweise auch unterschiedliche Organisationen, wichtig. Für die Bitstream Preservation, diese erste Ebene, kann man das Grundprinzip eigentlich sehr einfach formulieren: Es geht darum, möglichst viele, möglichst unabhängige Kopien zu schaffen, die dann regelmäßig kontrolliert werden und bei Diskrepanzen oder bei Beschädigungen möglichst schnell wiederhergestellt werden. Das ist eigentlich das ganze Prinzip. Aber dieses "möglichst unabhängige" bedeutet sowohl technologisch als auch organisatorisch und geografisch. Es sind eine ganze Reihe von Arten der Unabhängigkeiten da drin involviert. Und man kennt es aus dem Bereich der Technik eben, wenn man Festplatten nur aus einer Produktionscharge benutzt: Dann kann es durchaus vorkommen, dass die einen gemeinsamen Fehler oder eine gemeinsame Eigenschaft haben, was dazu führt, dass sie im gleichen Moment ausfallen oder in einem Zeitraum, der hinreichend ähnlich ist.

PM:

Denken Sie, dass dieser Dreiklang der verschiedenen Speichermedien wirklich praktikabel in Zukunft durchführbar ist? Gut, wenn man analoge Daten hat, kann man die digitalisieren, dann macht man ein PDF draus. Aber wenn ich umgekehrt mit digitalen Daten starte - ich denke jetzt bei BIM z.B. an eine IFC-Datei, die zwar aus ASCII-Code besteht, aber von einem Menschen nur bedingt interpretierbar ist - , dann stellt sich die Frage, ob man diese Dateien auch ausdrucken kann. Wenn ich von CAD-Planungsdaten ausgehe, kann ich die natürlich in einen Editor einlesen. Aber ergibt es Sinn, diesen Inhalt auszudrucken, nur um die Information auf Papier zu haben?

Experte:

Ich würde nicht sagen, dass es dieser Dreiklang sein muss. Aber man muss sich wirklich das konkrete Szenario anschauen, um das es geht. Ich erinnere mich, dass ich vor langer Zeit mal ein Rechenzentrum besuchen konnte, wo auch Bankdaten aufbewahrt wurden. Und da war klar: Die wollen auch optische Speichermedien haben. Es war wichtig, weil sie dadurch die Unveränderbarkeit garantieren können. Anders als bei einer Festplatte, wo - wenn es vor Gericht geht - jemand das anzweifeln kann. Das ist deutlich schwieriger beim optischen Speichermedium. Das ist ein anderes Szenario. Also es ist sehr fallabhängig, sehr szenarioabhängig, was die richtige Wahl der Medien ist. Also ich will das nicht in dieser Allgemeinheit, wie es hier formuliert ist, für Langzeitarchivierung behaupten. Aber: Unabhängigkeit von Kopien auf ganz verschiedenen Ebenen ist ein wichtiger Faktor. Gezielte Redundanz.

PM:

Wir haben uns bereits ein wenig über Lese- oder Dekodierungsgeräte unterhalten. Wenn ich Papier habe, brauche ich ja außer diesem Informationsträger nichts Weiteres zum Auslesen, das erledigt das menschliche Auge. Mit digitalen Daten sieht es anders aus. Ich weiß nicht, ob das in der Verordnung konkretisiert werden muss, aber wie sehr müssen Sie auch für die Möglichkeit des Auslesens die Verantwortung tragen?

Experte:

Wir übernehmen Verantwortung für die Informationen und ihre Nutzbarkeit. Nach meinem Verständnis wird es in dieser Verordnung nicht primär um physische Objekte gehen, sondern um Inhalte. Und die Frage nach Lese- und Dekodierungsgeräten ist eher etwas in der Terminologie von vorhin für Medium Term und Long Term. Also solange wir eine Institution haben, die sich darum kümmert - wir oder ein Dienstleister - und dafür Sorge trägt, dass wir weiterhin auf die Information zugreifen können. Beispiel Festplatten: Solange man sich in einem Bereich befindet, dass man sich ständig darum kümmern kann und die Festplatten ersetzen und austauschen kann, spricht nichts gegen Festplatten als Speichermedium. Nur in dem Moment, wo wir sagen: "OK, wir haben jetzt einen Informationsstand, den wir so sichern wollen, dass selbst bei Vernachlässigung in 50 oder 100 Jahren noch jemand darauf zugreifen kann" - Dann wird die Speichermedium-Frage relevant.

PM:

Das wäre dann Long Term, ja.

Experte:

Das ist dann insbesondere Long-Term. Zum gewissen Grad liegt die Schwierigkeit darin, dass man ja manchmal nicht weiß, wann hört eine Organisation auf, sich um bestimmte Dinge zu kümmern, weil das von Rahmenbedingungen abhängt. Ich weiß es aus dem Bereich der Forschungsdaten: Am Anfang der ersten Präsidentschaft von Donald Trump in den USA - ich kann nicht exakt sagen, welche Datenzentren - gab es meines Wissens Forschungsdatenzentren, die nicht mehr die Mittel erhielten und deshalb konnten bestimmte Datenbestände nicht erhalten und fortgeführt werden und sind für die Forschung verloren gegangen. Das ist im gewissen Sinne ein Prozess, der in der Menschheitsgeschichte auch normal ist, dass halt eben nicht alle Daten fortgeführt werden. Manche finden das dann bedauerlich, andererseits übersteigt es auch die Kapazität, alles zu bewahren. Aber solange wir diesen kontrollierten Prozess und eine Organisationseinheit haben, ist Lese- oder Dekodierungsgeräte eine nachrangige Frage.

PM:

Also meinen Sie: Solange einfach die Organisation da ist, mit Personal und vielleicht auch aktiver Forschung am Leben gehalten wird, ist es ein ausreichend stabiles Gerüst.

Experte:

Möglicherweise wird man auch zusätzlich eine Kopie auf besonders haltbaren Speichermedien oder so etwas in naher Zukunft aufbewahren. Aber in dieser Begrifflichkeit Short Term, Medium Term, Long Term ist es streng genommen erst für den Long Term-Fall notwendig, dass man sicherstellt, dass es eine möglichst hohe Unabhängigkeit hat. Es kann natürlich auch andere Faktoren geben. Möglicherweise ist es in einer gewissen Hinsicht finanziell günstiger, auf Lesegeräte oder Dekodierungsgeräte zu verzichten. Es gibt Diskussionen über sogenannten Cold Storage. Das betrifft auch die großen sogenannten Hyperscaler, die großen Anbieter von Rechenzentren, zum Beispiel Amazon, Google und andere. Die

haben riesige Datenbestände, aber auf einen Großteil wird nur extrem selten zugegriffen. Die allermeisten Zugriffe passieren halt eben auf das, was gerade auf Social Media und anderer Stelle sehr aktuell ist, aber halt nur auf einem sehr kleinen Teil der Daten. Und die versuchen, die Informationen immer auf dem besten Speichermedium für diese Zugriffshäufigkeit zu speichern.

Die nehmen in Kauf, dass der Zugriff auf alte Daten länger dauert, weil das dafür dann deutlich günstiger ist. Im Extremfall wären es Bandspeicher für Daten, die extrem selten angefragt werden und wo sie denken, es ist vertretbar. Und die Daten, die am häufigsten angefragt werden, sind auf SSDs oder sogar im RAM-Speicher. Das ist eine Hierarchie von Speichern. Und für Cold Storage wird kontinuierlich an neuen Speichermedien geforscht. Da hat man so etwas wie DNA, Keramik und so weiter im Blick. Und da stellt man sich dann auch immer die Frage: Was sind die Lese- oder Dekodierungsgeräte, die ich dafür benötige oder nicht benötige? Und das sind Fragen bei denen im gewissen Sinne das Hauptkriterium ist, ob es sich auf dem Markt durchsetzen kann. Da werden dann eben ökonomische Fragestellungen relevant.

PM:

Im Forschungsbericht wurde erwähnt, dass es von 2023 bis 2025 auch ein Forschungsprojekt speziell zur Langzeitbeständigkeit digitaler Speichermedien gibt. Das fand ich sehr interessant und es passt ja auch zu dem, was wir gerade besprochen haben. Im Prinzip ist es die logische Fortsetzung der Untersuchung der Langzeitbeständigkeit von Papier. Wissen Sie da mehr über den Scope dieses Projekts und die dort angewandten Methoden?

Experte:

Ich betreue nicht selber das Projekt, sondern eine Kollegin. Das Projekt schaut sich die Haltbarkeit von Speichermedien unter verschiedenen Faktoren an. Auch mit einer Unterscheidung, welche Speichermedien besonders geeignet für den Fall sind, dass sie aktiv gewartet werden oder dass eben eine Vernachlässigung eintritt. Und weil wir uns ja auch nicht auf Basis dieses Projektes für Speichermedien in den nächsten Jahren entscheiden müssen, ist es aus meiner Sicht eher dafür gedacht, einen aktuellen Informationsstand zu gewinnen und in zukünftigen Projekten vergleichen zu können, wie sich das Feld weiterentwickelt. Über die Jahre hinweg kann so ein Gesamtbild aufgebaut werden, welches die Entwicklung von Speichermedien und deren Auswahl umfasst. Es gibt auch einen Methodikanteil darin, wie man Speichermedien bewertet. Konkrete erste Ergebnisse kann ich Ihnen nicht nennen.

PM:

OK. Dann müssen wir wohl das Ende des Projekts abwarten. Wir haben jetzt immer von Informationen gesprochen. Es gibt ja im Wissensmanagement diese Art Pyramide. Wir haben unten Daten, in der Mitte Informationen und darüber das Wissen. In meinem Konzept ist Dreh- und Angelpunkt ja eigentlich immer das Wissen. Ich gehe davon aus, dass es für kerntechnische Anlagen sinnvoll ist, nicht nur auf dieser Dokumentenebene, wo es vorrangig um Daten und Informationen geht, Mittel zu haben, sondern sich verstärkt mit dieser Wissenskomponente zu beschäftigen. Wie sehr ist das für Sie in der Langzeitarchivierung relevant? Auch im Gesetz wird ja "nur" von Informationen gesprochen.

Experte:

Also die Kollegin, die das Projekt zu den Speichermedien betreut, ist auch zuständig für Wissensmanagement. Für uns als Institution und auch in der Langzeitarchivierung hat das eine Relevanz. Ich würde sagen: Dieses Thema ist tatsächlich so ein bisschen der Langzeitarchivierung vorgeordnet oder

übergeordnet. Es geht ja um die Frage: Was sind die relevanten Informationen? Aber es hängt natürlich zum gewissen Grad auch damit zusammen, dass die Informationen interpretierbar bleiben müssen, es wird also Kontextwissen benötigt. Im OAIS gibt es im Prinzip auch ein theoretisches Instrumentarium, um Kontextwissen zu bezeichnen: "Representation Information". Kontextwissen ist nicht nur technisch, sondern es kann auch semantisch sein, um das mal einfach als Oberbegriff für inhaltliches Kontextwissen zu benutzen. Aber das ist etwas, da sind wir wieder im Aufbau. Wir müssen uns Erschließungskonzepte erarbeiten, wodurch die Information auch navigierbar wird. Das können wir jetzt, ohne dass wir über die Informationen verfügen, noch nicht fertig präsentieren oder nicht voraussetzen. Natürlich haben wir eine gewisse Systematik zum Beispiel durch die veröffentlichte Entwurfsliste in den Eckpunkten für eine Verordnung, was auf alle Fälle bedeutsame Daten sind. Da ist schon eine gewisse Logik drin, wie man das dann macht, wie man das strukturiert, kategorisiert oder Klassifikationen oder Ontologien möglicherweise auch selber entwickelt.

Da schauen wir uns Sachen an, aber wir können es nicht vorab entscheiden. Es gibt aus einem OECD/NEA-Projekte - ich glaube, das nennt sich dort RepMet [Radioactive Waste Repository Metadata Management]- eine Art Ontologie-Entwurf. Da sind Entitätenklassen gebildet worden und ich würde sagen: Ja, es muss in diese Richtung gehen, dass man halt eben ein Informationsmodell vom Endlagersystem zusammen mit dem Müll und den Akteuren und dem Standort baut und anhand dessen die Informationen navigierbar macht und miteinander in Beziehung setzen kann. Das fand ich auch in Ihrer Präsentation interessant, dass auch Sie an einer Ontologie sitzen. In die Richtung muss es gehen, denke ich.

PM:

Es gibt ja verschiedene Abstufungen. Das Einfachste wäre eine Taxonomie, das ist ja auch etwas, womit Dokumentenmanagementsysteme arbeiten, also das typische Verschlagworten. Und Ontologie wäre dann schon...

Experte:

...logische Schlussfolgerungen.

PM:

Genau. Reasoning dann hinterher - das ist gerade auch bei uns ein großes Thema. Sie haben ja bei IFC gesehen: Es gibt einzelne Klassen, also IFCDoor, IFCWall, und so weiter. Das ist ja schon so eine Art Taxonomie. Es gibt Forschungsprojekte, in denen es darum geht, für das Bau- oder Umweltwesen "die" Ontologie zu finden oder man strebt an, Ontologien für einzelne Teilbereiche miteinander zu kombinieren. Aber die Frage ist natürlich: Wenn ich möchte, dass Leute von mir aus in 500 Jahren mit dem archivierten Wissen noch etwas anfangen können, braucht es halt dieses Kontextwissen. Wie viel Wert legt man auf Technologien wie Big Data oder KI im Vorfeld, beispielsweise zur Klassifikation? Man könnte ja auch sagen, dass die Leute in 500 Jahren das selbst lösen sollen und wir heute uns nur für die Informationen zuständig fühlen.

Experte:

Ja, ich glaube, dass unterschiedliche zeitliche Phasen unterschiedliche Aufgaben haben. Die Aufgaben, die wir jetzt haben: Ja, da schauen wir uns auch diese Themen Ontologien, KI, Big Data an. Aber stärker für unsere gegenwärtigen Aufgaben. Aber die Informationen dann in eine Form zu bringen, dass selbst wenn es organisatorische Diskontinuitäten gibt, lokale, regionale Katastrophen - was auch immer - , diese noch nach einem Abbruch der Betreuung oder Stewardship verständlich sind? Das ist eine

besondere Aufgabe. Die würde ich nicht in eins setzen. Also wir werden jetzt natürlich schauen, wie wir möglicherweise mit KI-Methoden die Erschließung vereinfachen oder das Finden von Informationen verbessern können. Aber das ist etwas anderes als das, was man in Zukunft aufbereiten wird, damit zukünftige Generationen möglichst voraussetzungslos die Informationen nutzen können. Das muss man unterscheiden. In dem RK&M-Bericht, der von Ihnen zitiert wurde, sind auch ein paar unterschiedliche Instrumente drin. Unter anderem, dass separate Dokumentensätze oder Dokumente erstellt werden, wie der Key Information File oder Set of Essential Records.

Es werden eigene Produkte für bestimmte Zwecke erstellt. Und vielleicht nutzt man da KI-Methoden. Aber wir brauchen die auch vorher schon für andere Aufgaben. Aber in dieser RepMet-Initiative werden Entitäten definiert und in Bezug zueinander gesetzt, was im Bereich Endlagerung für Informationen erfasst werden sollten. Meines Erachtens ist es etwas, was fortgeführt werden muss. Das ist nicht der Weisheit letzter Schluss. Aber das ist eben, wieweit diese Initiative gekommen ist. Und es ist immer schön, wenn man sich an einem Standard orientieren kann. Es gibt halt nur oft dann viel zu viele Standards. Und wir müssen eben schauen, wie wir diese aufgreifen und damit weiterarbeiten.

PM:

Kommen wir nochmal auf mein Konzept zu sprechen. Ich habe Ihnen ja erzählt, dass ich aktuell mit einer Graph-Datenbank arbeite. Mein anfängliches Konzept ging davon aus, dass ich die Wissenselemente möglichst nah an das BIM-Modell bringen will und es wie eine Einheit behandle. Mittlerweile bin ich davon wieder weggegangen und denke mir, dass es zukunftssicherer ist, wenn man das sauber trennt. Man hat das BIM-Modell und natürlich kann jedes Knowledge-Element auf einzelne Elemente darauf verweisen. In BIM-Modellen hat ja jedes Objekt eine Unique ID. Darüber kann man einen Querverweis machen. So steckt nicht alles im BIM mit drin, sondern das eigentliche Wissen steht in der Graph-Datenbank in Form einzelner Datensätze. Ist dieser Ansatz in Hinblick auf eine mögliche Langzeitarchivierung vielversprechend? Ist es sinnvoll, die "Module" auseinander zu halten und nicht einen großen "Block" zu haben?

Experte:

Auch auf die Gefahr hin, ein Klischee oder eine Phrase zu benutzen: Es ist besser, keine Monolithen zu haben. Und dieser Ansatz der Integration klingt für mich schon so, dass es ein großes Modell ist, was dann in der Wartung wahrscheinlich schwieriger wäre als einmal das BIM-Modell und eine Vielzahl an Tools zu haben, die jeweils auf einzelne Standards zurückführbar sind. Und dann habe ich hier nochmal eine separate Graph-Datenbank oder etwas, das sich hoffentlich auch an Standards orientiert oder zumindest Aspekte davon aufgreift, denn Standards sind ja proprietär. Also ich glaube, dass Modularität ein Vorteil ist bei der Erhaltung und bei der Langzeitarchivierung.

PM:

Das nehme ich gerne mit, denn das stützt mein Konzept. Ja prima. Dann sind wir jetzt auch schon am Ende meines Leitfaden angekommen. Gibt es von Ihrer Seite aus noch Themen, über die Sie sprechen möchten?

Experte:

Viele wahrscheinlich. Ihnen geht es ja um Wissensmanagement. Und vielleicht die größte Baustelle oder ein Risiko, was man haben kann in Bezug auf das Digitale, ist glaube ich nicht, wenn es in Form eines klar abgrenzbaren digitalen Objekts einer Datei oder Ähnliches daherkommt, sondern da, wo Information und Wissen in einer Software oder einem Service steckt. Ein bisschen habe ich es auch angesprochen am

Anfang bei der Abgrenzung eines Hardware-Gegenstands von der Software, die zur Steuerung benutzt wird. Also da, wo das Wissen verkörpert ist - nicht einfach in einem Dokument, sondern in Software oder Services. Das ist etwas, glaube ich, wo eine besondere Herausforderung besteht und man wird diese Fälle bekommen.

Ich weiß, dass ist jetzt noch nicht in Form eines mit nach Hause nehmbaren Leitsatzes formuliert, aber ein Großteil der Fragestellungen in der Langzeitarchivierung muss sich eigentlich dahin weiterentwickeln, den Service, Nutzen, Zweck hinter etwas Digitalem zu bestimmen. Das ist derzeit nicht üblich in der Langzeitarchiv-Community. Aber von diesen Fragen - Mache ich jetzt diese oder jene Erhaltungsmaßnahme für dieses einzelne Datenobjekt? - muss ich eigentlich eher zur Nutzbarkeit insgesamt gehen. Was ist hier eigentlich der Service, der genutzt wird? Es handelt sich [bei Ihrem Konzept] um eine Kette von Elementen, die zusammen eigentlich ein Netzwerk bildet, damit jemand etwas nutzen kann, nicht?

PM:

Ja, genau.

Experte:

Man hat eben - manchmal sagt man, es ist ein Technologie-Stack. Aber es ist eigentlich mehr als ein Technologie-Stack, sondern man hat eigentlich ein ganzes Netzwerk von Abhängigkeiten. Man muss also das gesamte Netzwerk benutzen, um auf etwas zuzugreifen. Und man ist dann nicht mehr mit dem Erhalt eines einzelnen Objekts, sondern eines größeren Netzwerks konfrontiert. Und das scheint mir eine besondere Herausforderung zu sein. Und das macht gerade Einzellösungen so fragil, weil es dann nur diesen einen Zugang gibt, um dieses Objekt zu nutzen; sprich: diesen einen einzigen Pfad durch den Dependency-Graph. Also die Abhängigkeiten bilden einen Graphen und - das ist eine fast vergessene Theorie in der Langzeitarchivierungs-Community, dass man sich für die Bewertung der Erhaltbarkeit in so einem Abhängigkeits-Graphen sogenannte View-Paths anschaut - Wie kann ich auf ein Objekt zugreifen? Das ist ein Pfad in diesem Abhängigkeits-Netzwerk. Eigentlich muss man schauen, dass man dieses Netzwerk optimiert, dass man möglichst viele Pfade in ihm hat. Und besonders schwierig wird das jetzt nicht bei typischen digitalen Objekten, sondern wenn das, was ich erhalten will, eigentlich der Service sein muss.

Es gab vor etlichen Jahren mal eine hervorragende Keynote bei der iPRES-Konferenz, das müsste 2008 oder 2009 gewesen sein: "How to preserve virtual worlds." Und da fiel einem auf: Oh, das ist ganz schön schwer. Zum einen hat man die Software nicht lokal laufen, die läuft irgendwo auf einem Server. Und dann noch viel schlimmer: Die virtuellen Welten, die virtuellen Computerspiele leben davon, dass da auch andere Leute drin sind. Eigentlich braucht es fast an diesen Stellen so etwas wie eine Geschichtsschreibung, als dass man da wirklich etwas erhalten kann. Das geht jetzt ein bisschen zu weit, aber an vielen Stellen muss man eher im Sinne des Erhalts eines Ökosystems denken...

PM:

Ja, "Ökosystem" trifft es sehr gut.

Experte:

... und möglichst vieler gangbarer Pfade in einem Netzwerk. So ein bisschen ist das dann halt auch mit "Lösungen" im Sinne von Spezialanwendungen und interaktiven Webseiten, hinter denen ein Service steckt. Wenn ich die Lösung erhalten will, muss ich eigentlich diesen Service erhalten. Ich muss die Software erhalten, die oft aber auch custom made ist. Und dann bin ich eben sehr stark darin, um die

Software zu erhalten, gleich einen ganzen kleinen Technologiezoo bzw. ein Ökosystem zu erhalten. Das scheint mir die größte Herausforderung zu sein. Alle Fälle, in denen Wissen so komprimierbar sein kann und explizit gemacht werden kann, dass es halt eben in einem Dokument oder etwas Ähnlichem ist und solche Abhängigkeiten nicht hat - das sind die guten Fälle. Aber weil wir auch Datenbanken ansprachen: Datenbanken sind ganz oft verbunden mit solchen Services. Da sehe ich die Herausforderung. Da kann es sehr große Aufwände geben.

PM:

Ja, das nehme ich auf jeden Fall mit. Wenn ich jetzt an mein Konzept denke, habe ich natürlich eine Graph-Datenbank, wo die einzelnen Informationen drin sind. Aber klar, die Zusammensetzung findet in einer Game Engine statt und Visualisierung erfolgt mit einem VR/AR-Headset. Ich müsste mir mal anschauen, ob man in der Lage wäre, wenn es jetzt irgendwann nicht mehr diese Brillen gibt oder wenn sich mal die Software oder die Game Engine und so weiter ändert, alleine auf Grundlage der Datenbank die relevanten Informationen zu rekonstruieren. Diese Frage stellt sich also auch: Ist ein Modul nicht nur gut integrierbar, sondern enthält es auch eigenständig für sich Informationen und Wissen?

Experte:

Ja, und ich meine, Sie nutzen Unity. Das ist natürlich gut, weil das ist wahrscheinlich die verbreitetste Game Engine. Und auch das ist natürlich ein Nachhaltigkeitsfaktor, dass wir etwas benutzen, was derzeit verbreitet ist und dadurch vielversprechend ist. Aber sowohl bei Speichermedien als auch bei Software gilt: Diese Marktposition und die Wirtschaftlichkeit von diesen Dingen sind ein wichtiger Faktor, ob es eben langfristig nutzbar ist. Man kann es eben sehr schön an dieser Frage der Speichermedien festmachen, weil da kommen alle paar Jahre Berichte über "das Superspeichermedium". Wissenschaftsjournalisten berichten darüber und es ist auch tatsächlich sehr interessant und fantasieanregend. Aber die kommen fast nie in die Breite und sind fast nie effizient nutzbar. Wieso? Weil sie sich nicht auf dem Markt durchsetzen können. Weil sie theoretisch in einzelnen Fällen besser wären, aber: Jede neue Speichermedium-Technologie, die sich heute behaupten will, muss von den großen Hyperscalern eingesetzt werden. Sonst rechnet sich die Entwicklung und Massenproduktion eben nicht. Dann sind andere Dinge effizienter und es kann sich nicht etablieren. Und dadurch hat es dann auch keinen Bestand, weil dann die Technologiepfade zu sehr einschrumpfen, fragil werden und irgendwann brechen. Ich nannte es vorhin die organisatorischen Faktoren für vertrauenswürdige digitale Langzeitarchive - die muss man halt mitberücksichtigen.

PM:

Ja, das finde ich einen interessanten Punkt, dass man auch einfach guten Gewissens auf das setzen kann, was sich am Markt bereits durchgesetzt hat und nicht immer Hoffnungen in Technologien steckt, die noch nicht da sind.

Experte:

Das ist etwas, wo man dann auch bescheiden und sich bewusst werden muss: Der Markt ändert sich auch und man muss die verschiedenen Arten an Technologien mitverfolgen, damit die nutzbar bleiben. Und für die Zeit, wo wir die digitalen Daten bekommen und damit aktiv arbeiten, haben wir gar keine andere Chance. Also, wir müssen das machen, weil so entstehen die Daten, so bekommen wir sie und wir müssen sie halt weitergeben.

PM:

Ja, OK. Das war doch ein gutes Schlusswort.

Experte:
Ich hoffe, das eine oder andere Interessante war dabei.

PM:
Auf jeden Fall. Ich bedanke mich ganz herzlich für das Gespräch mit Ihnen.