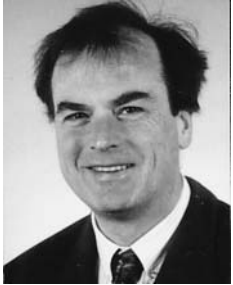


Bewertung von Softwaresystemen zur Langzeitarchivierung digitaler Objekte

Foto privat



Uwe M. Borghoff

Foto privat



Peter Rödiger

Foto privat



Jan Scheffczyk

Foto privat



Franz Schmalhofer

Dieser Beitrag basiert auf den Ergebnissen der nestor-Expertise »Vergleich bestehender Archivierungssysteme« und beschreibt, wie ein Kriterienkatalog zur Unterstützung der Bewertung und Auswahl eines Softwaresystems zur Langzeitarchivierung digitaler Objekte aufgebaut und eingesetzt werden kann. Die Einleitung skizziert die Besonderheiten digitaler Objekte und die Verantwortung digitaler Archive bezüglich der langfristigen Erhaltung von Information. Eine Eingrenzung und grobe Einteilung der Softwaresysteme liefern grundsätzliche Vorgaben für die Entwicklung eines Kriterienkatalogs. Dieser ist in den Kontext eines Entscheidungsprozesses einzupassen. Als Ausgangsbasis zur Herleitung von Kriterien dienen Referenzmodelle, Technologien und Anwendungen sowie die Ergebnisse verwandter Untersuchungen. Die Herleitung funktionaler Kriterien basiert auf dem OAIS-Referenzmodell und einem ergänzenden Inhaltsmodell. Nicht-funktionale Kriterien der Art Aufwand und Qualität kommen hinzu. Die wesentlichen Bestandteile des Kriterienkatalogs sind damit begründet. Eine Betrachtung bestehender Softwaresysteme findet sich im Resümee der nestor-Expertise. Die abschließende Diskussion befasst sich mit der praktischen Anwendung des Kriterienkatalogs sowie weiteren Unterstützungsmöglichkeiten bei der Produktauswahl.

This contribution is based upon results of the nestor expert's report »Comparison of Existing Archival Systems« and describes how a set of criteria for evaluating and selecting a software system for long-term preservation of digital objects can be developed and utilized. By way of introduction, the special features of digital objects and the responsibilities of digital archives in terms of long-term preservation of information are outlined. The specifications and general characteristics of software systems provide the basis for the development of a criteria catalog. This catalog must be tailored to the decision-making process. As a starting point for deriving the appropriate criteria, use is made of reference models, technologies, and applications as well as the results of related studies. The derivation of functional criteria is based on the reference model for an Open Archival Information System (OAIS). Non-functional criteria regarding cost and quality are then added. The essential elements in a set of criteria are thus laid out. A survey of existing software systems is to be found in the resumé by the nestor expert group. In closing, there is a discussion of practical applications of such a criteria catalog as well as further support options in selecting a product.

EINLEITUNG

Die Erhaltung und Bereitstellung von Informationen, die durch digitale Objekte repräsentiert werden, stellt Bibliotheken, Archive und Museen vor neue Herausforderungen, denn die rein physische Erhaltung dieser Objekte sichert nicht den Erhalt der Information. Bedingt durch eine anwendungsspezifische Verwendung von Formaten zur Darstellung von Information entsteht ein Bedarf an unterschiedlicher interpretierender Aufbereitung. Die Sicherstellung dieser Interpretierbarkeit – vor allem über einen langen Zeitraum hinweg – erweitert die Verantwortlichkeit von Archiven. Zum Ausdruck kommt dies in der im OAIS-Referenzmodell¹ festgelegten Regelung des Aufgabenbereiches eines Archivs: *An OAIS is an archive, consisting*

of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community. Diese Definition vermittelt zusätzlich das Bild eines Archivs als Organisationsform bestehend aus Personen und Systemen. Außerdem sind die Art der Objekte und die zu nutzenden Systeme im OAIS-Referenzmodell nicht weiter bestimmt. In einem Archiv, das die Verantwortung für digitale Objekte übernimmt, spielen Softwaresysteme eine entscheidende Rolle. Insbesondere ist es sinnvoll, die Interpretation der digitalen Objekte der Software zu übertragen. Funktionalität und Qualität und letztlich auch die Vertrauenswürdigkeit eines Archivs werden neben dem organisatorischen Umfeld maßgeblich durch Softwaresysteme bestimmt. Der Auswahl eines Softwaresystems kommt daher große Bedeutung zu.

Erwartungen der Nutzer oder Vorgaben des Gesetzgebers bzw. übergeordneter Instanzen bestimmen die Anforderungen an ein Archiv. Im Rahmen einer Anforderungsanalyse formulieren Verantwortliche eines Archivs unter Berücksichtigung verfügbarer Ressourcen dessen zu realisierende Funktionalität und Qualität. Somit stehen Merkmale des Archivs als Vorgabe für die Auswahl von Softwaresystemen zur Verfügung. Um eine Produktauswahl treffen zu können, ist in einem weiteren Schritt der Grad der Übereinstimmung der Vorgaben mit den Eigenschaften eines Softwaresystems festzustellen. Dies erfordert eine entsprechende Charakterisierung der Systeme. Diese Auswahlproblematik und die damit verbundene Erarbeitung von Systembewertungen stellt sich für digitale Archive immer wieder. Wegen der Komplexität der Anforderungen und der Vielzahl vorhandener Systeme ist damit hoher Aufwand verbunden. Zur Unterstützung des Auswahlprozesses wurde im Rahmen der Initiative nestor die Expertise »Vergleich bestehender Archivierungssysteme«² beauftragt. In ihr wurde ein Kriterienkatalog entwickelt, der eine vergleichende Bewertung von bestehenden Archivierungssystemen erlaubt. Nachfolgende Ausführungen geben Einblick in Vorgehensweise und Resultate dieser Studie.

EINGRENZUNG UND EINTEILUNG DER SOFTWARESYSTEME

Das Angebot für Softwaresysteme zur Archivierung digitaler Objekte ist ein Marktsegment mit dynamischer

Entwicklung. Derzeit ist eine Vielzahl von kommerziellen und nicht-kommerziellen Produkten, Komponenten und Zusatzprodukten am Markt, die die Archivierung mehr oder weniger unterstützen. Archivierungsfähigkeiten (im weitesten Sinne) finden sich in den Produktkategorien DMS (Dokumentenmanagementsysteme), DAMS (Digital Asset Managementsysteme), CMS (Content Managementsysteme), Archivierungssysteme, Repositories und Storagemanager.

Die Expertise ist auf »reine« Archivierungssysteme ausgerichtet, also auf Systeme, die die Archivierung als Kernaufgabe ansehen bzw. die durch die Hersteller als Archivierungsprodukt eingeordnet sind. Folgende Systemkategorien werden nicht den Archivierungssystemen zugerechnet:

Reine Entwicklungswerkzeuge:

Hierbei handelt es sich um Produkte, die die Entwicklung von Endanwendungen unterstützen (z. B. Portal Tool Kits, DBMS-Schemageneratoren). Dieser Entwicklungsprozess setzt Fachwissen voraus, insbesondere was die korrekte Interpretation und Umsetzung vorgelagerter Entwurfsmodelle für die zu realisierenden Systeme anbelangt.

Anwendungsneutrale Grundsysteme:

Diese Systeme (i. W. Datenbankmanagementsysteme und Storagemanager) besitzen i. A. keinen Anwendungsbezug, d. h. es sind keine vordefinierten, anwendungsbezogenen Objekte und Funktionen verfügbar. Während Datenbanksysteme zumindest ein logisches Modell besitzen, auf dem sich ein Anwendungsfall mehr oder weniger leicht abbilden lässt, kennen übliche Storagemanagementsysteme bezüglich der Anwendungsseite nur physische oder logisch primitive Konzepte, wie ganze Festplatten(-gruppen), Blöcke oder Dateisysteme bzw. Dateien.

Produkte, bei denen die Archivierung nur die Teilfunktion einer umfassenden spezialisierten Anwendung darstellt:

Dabei handelt es sich um Produkte, bei denen die Archivierung bzw. die Ablage nur eine untergeordnete Rolle spielt und beispielsweise die Erstellung und Nachbearbeitung komplexer digitaler Objekte (z. B. in CAD-Systemen) oder die Unterstützung eines komplexen Business-Workflows (z. B. in einigen Dokumentenmanagementsystemen) im Vordergrund stehen. Solche Produkte würden wegen der engen Verflechtung mit einer spezifischen Anwendung eine Einschränkung der Allgemeinheit bezüglich der Archivierung bedeuten und eine überflüssige Ressourcenbindung (z. B. bei der Installation oder einem Releasewechsel) für nicht benötigte Funktionalität mit sich bringen.

Neben der Eingrenzung der Softwaresysteme, auf die der Bewertungsprozess zur Anwendung kommen

soll, ist eine Einordnung der verbleibenden Systeme bezüglich folgender Aspekte angebracht:

Abdeckung bezüglich Funktionen und Objekten:

Softwaresysteme werden als anwendungsbezogene Einzelkomponenten bzw. Zusatzprodukte angeboten. Hierbei handelt es sich um Komponenten, die nur einzelne Hauptfunktionalitäten eines Archivierungssystems abdecken, wie etwa in OAIS definiert, oder um Produkte, die eine spezialisierte Aufgabe innerhalb oder außerhalb von OAIS wahrnehmen (z. B. Systeme für Metadatenharvesting, Indizierung, Scannen / OCR, Dateiformatkonvertierung, Dateiformatanalyse, XML-Schema-Validierung). Es existieren Produkte, die konzeptionell nur für die Archivierung spezieller digitaler Objekte vorgesehen sind. Das sind Produkte, die nur die Speicherung von größtmäßig begrenzten Einzelobjekten zulassen oder ablegbare Formate auf einen oder wenige Typen einschränken. Diese Beschränkung geht in der Regel auch mit einem eingeschränkten Metadatenmodell einher – zumindest was Metadaten anbelangt, die für die Langzeitarchivierung relevant sein können. Außerdem gibt es Produkte, die nur für eine eingeschränkte Datenmenge konzipiert sind. Dabei handelt es sich um Produkte aus dem so genannten *Home Office* oder *Small Business* Bereich. Im Angebot befinden sich auch Einzelplatzlösungen mit lokaler Ablage oder nur primitivem externen Zugang.

Entwicklungsstand und Entwicklungspotenzial:

Des Weiteren sind Produkte mit konzeptionell vollständiger Archivierungsfunktionalität auf dem Markt, die an spezifische Anwendungsfälle durch weitgehende Konfigurationsmöglichkeiten oder Entwicklung von zusätzlichen Funktionen unter Nutzung offener Schnittstellen angepasst werden können.

Abdeckung der Langzeitaspekte:

Bis auf einige Ausnahmen sind bei den Archivierungssystemen keine expliziten Mechanismen vorgesehen oder Vorkehrungen getroffen, um die Langzeiterhaltung der Inhalte zu gewährleisten.

Daraus ergeben sich Anforderungen an einen Kriterienkatalog:

Modularität:

Produkte unterschiedlichen Leistungsumfanges (funktionale Abdeckung) sind nicht unmittelbar vergleichbar. Ein modularer, an funktionalen Komponenten orientierter Ansatz sollte dies ermöglichen.

Abstraktion:

Den Produkten, die erst durch ihre Anpassung an die Anwendungssituation ihre »endgültige Gestalt« (Konfiguration) erhalten, sind nicht immer feste Merkmale zuordenbar. Eine Vergleichbarkeit ist hier durch eine abstrahierte, auf Konzepte Bezug nehmende Betrachtung möglich. Verwendete Konzepte geben auch Aus-

DMS, DAMS, CMS

Storagemanagementsysteme

kunft über im Produkt steckendes Entwicklungspotenzial.

Kriterientransparenz:

Für die Langzeitarchivierung mangelt es an anerkannten Techniken. Trotz fehlender expliziter Maßnahmen kann ein Softwaresystem für die Langzeitarchivierung geeignet sein. Ausschlaggebend hierfür sind Eigenschaften, die zur Sicherstellung der Langlebigkeit des Archivs oder der Migrierbarkeit der Inhalte beitragen.

ROLLE DES KRITERIENKATALOGS IM ENTSCHEIDUNGSPROZESS

Eine Produktauswahl erfordert eine Produktbeschreibung, die einen Vergleich der für den spezifischen Anwendungsfall relevanten Produkteigenschaften erlaubt. Um die Auswahl in verschiedenen Entscheidungssituationen zu unterstützen, ist für die Produktbeschreibung eine bestimmte Allgemeingültigkeit und Allgemeinverständlichkeit anzustreben.

Die Vorgabe eines Kriterienkatalogs liefert einen Rahmen für die Erstellung derartiger Produktbeschreibungen. Der Katalog eröffnet auch den Zugang zu einem systematischen Entscheidungsprozess. Er erlaubt eine vergleichende Produktbeschreibung, indem die charakterisierenden Merkmale auf ein einheitliches, entsprechend feines Niveau heruntergebrochen werden und diesen Merkmalen mehr oder weniger normierte Werte zugeordnet werden können. Diese Zuordnung wird als *Rating* bezeichnet. Da die Bewertung mittels Kriterienkatalog von konkreten Anwendungsfällen abstrahiert, sind in einer weiteren Stufe für eine endgültige Auswahl (*Ranking*) individuelle Belange entsprechend einer Anforderungsanalyse zu berücksichtigen. Dies geschieht durch eine Gewichtung der Kriterien und eine Zuordnung der Merkmalswerte zu einer Maßskala.

Die Archivierungsprodukte stützen sich häufig auf weitere Produkte (z. B. Datenbanksysteme) bzw. können für spezielle Anwendungszwecke durch weitere Produkte ergänzt werden. Teilweise ist dafür eine umfangreiche Auswahl an Produkten möglich. Dies bedingt jedoch eine Vielzahl unterschiedlicher Konfigurationsmöglichkeiten, die teilweise erst im Rahmen einer konkreten Anwendung abschließend bewertet werden können. Insgesamt ergibt sich der in Abbildung 1 dargestellte Zusammenhang.

ENTWICKLUNG VON KRITERIEN: AUSGANGSBASIS

Wie in Abbildung 1 erkennbar, stützt sich die Erstellung des Kriterienkatalogs auf folgende Arten von Quellen:

Referenzmodelle:

Referenzmodelle beschreiben für eine bestimmte Anwendungsklasse meist funktionale Anforderungen und ggf. die dazu benötigten Objekte in einer abstrakten Art und Weise und definieren mehr oder weniger formal die dabei verwendeten Begriffe. Die Betrachtung von Referenzmodellen soll folgende Aspekte abdecken:

- Ableitung und Strukturierung von im Wesentlichen funktionalen Kriterien (funktionale Komponenten)
- Schaffung einer Grundlage für eine gemeinsame Begrifflichkeit (Allgemeinverständlichkeit)
- Darstellung des Funktionsumfangs von Entwurfsmodellen und Implementierungen.

Technologien:

Technologien und zugehörige Standards bestimmen maßgebend die Umsetzbarkeit bzw. die Implementierung von Referenzmodellen bzw. konzeptionellen Modellen. Sie konkretisieren oder erweitern die abstrakt beschriebenen Modellelemente. Außerdem werden die Produkte im großen Umfang mit technologischen Begriffen und nicht mit Begriffen eines Referenzmodells beschrieben. Zweck der Betrachtung von Technologien ist hier also:

- Zuordnung der Funktionalitäten: Technische Realisierungen zu Referenzmodellen bzw. konzeptionellen Modellen
- Darstellung von funktionalen Ergänzungen und Verfeinerungen gegenüber Referenzmodellen.

Da im Kriterienkatalog die Beschreibungsmerkmale auf einer konzeptionellen Ebene angesiedelt sind, finden sich Technologien als Merkmalsausprägungen, also als Werte von Kriterien, wieder. Technologien können sich auf die Umsetzungen innerhalb eines Produktes beziehen, aber auch auf eigenständige Produkte, die eine bestimmte Technologie bzw. Standard umsetzen.

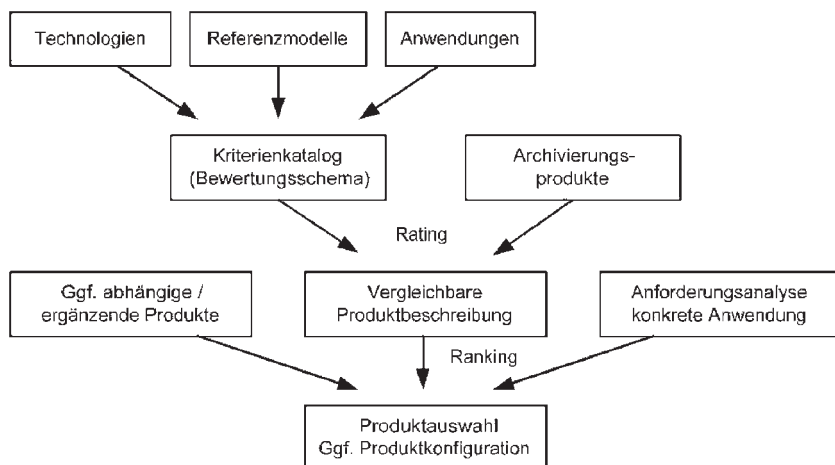


Abbildung 1: Herleitung des Kriterienkatalogs und seine Rolle im Entscheidungsprozess

zen (z. B. HTTP / HTML durch einen Webserver oder SQL durch ein Datenbankprodukt).

Anwendungen:

Weitere entscheidungsrelevante Merkmale, die nicht aus einer Betrachtung von Referenzmodellen und Technologien hergeleitet werden können, ergeben sich aus konkreten Archivierungsszenarien und dem dazugehörigen technischen und organisatorischen Umfeld. Letztlich kann nur die konkrete Anwendung die Relevanz von Kriterien – als Teil der Bewertung – bestimmen. Die Berücksichtigung der Anwendung erfüllt also einen »Doppelzweck«:

- Erweiterung um die Kriterienarten *Qualität* und *Aufwand*
- Bestimmung der Relevanz der Kriterien für den individuellen Anwendungsfall (durch Gewichtung der Kriterien im Rahmen des Ranking).

Wichtige Merkmale von Anwendungsszenarien sind:

- Art und Anzahl der digitalen Objekte
- Art und Umfang der Nutzer- / Nutzergruppen
- Geforderte Dienste für die Nutzer
- Geforderte Qualität (z. B. Verfügbarkeit, Zugriffszeiten).

Ein Anwendungsumfeld wird u. a. beschrieben durch:

- Gesetzliche Vorgaben
- Finanzielle und personelle Ressourcen
- Vorhandene Systeme (Integrations- / Migrationsaufwand).

Referenzmodelle, Anwendungen und Technologien wurden bereits in anderen Untersuchungen betrachtet. In ihrer nachfolgenden Durchsicht wird auf Aspekte der Langzeitarchivierung ein besonderes Augenmerk gelegt.

ENTWICKLUNG VON KRITERIEN: BETRACHTUNGEN VERWANDTER UNTERSUCHUNGEN

Eine Reihe von Untersuchungen beschäftigt sich, teilweise mit sehr unterschiedlichen Blickwinkeln, mit Anforderungen an Archivierungssysteme und den Eigenschaften von Archivierungssystemen. Drei Arbeiten wurden in der Expertise wegen ihrer Berücksichtigung von Implementierungsaspekten sowie ihrer Detailliertheit, ihres expliziten Bezugs zur Langzeitarchivierung und ihrer Aktualität näher betrachtet.

The Scholarly Publishing & Resources Coalition (SPARC) hat ein Papier mit dem Titel *Institutional Repository Checklist & Resource Guide*³ veröffentlicht, das Fragen behandelt, die mit der Einführung und dem Betrieb *institutioneller Repositories* verbunden sind. Dabei handelt es sich um Sammlungen, die den

intellektuellen Output von Universitäten und Colleges bzw. deren Verbünde sammeln und bewahren. Neben Aspekten der Organisation und Kulturpolitik, der Auswahl- und Aufbewahrungspolitik, der Reichweite und Zusammenarbeit werden auch Fragen der technischen Infrastruktur diskutiert. Ein Überblick über Grundkonzepte und Initiativen soll eine individuelle Evaluierung erleichtern. Folgende Themen und Unterpunkte werden dabei aufgegriffen: Aufwand für Entwicklung und Betrieb, Fähigkeit zur Migration und zum Überleben, Langlebigkeit und Formate des digitalen Inhalts, Skalierbarkeit, Outsourcing der Erhaltung, persistente Bezeichnung, Interoperabilität und offener Zugang, OAI-konforme Suchdienste, Nutzerzugriff und Rechteverwaltung. Die Diskussionen finden auf einem sehr allgemeinem Niveau statt, aber zu den einzelnen Themen gibt es umfangreiche Quellenangaben. Ein systematischer Zusammenhang zu OAIS oder anderen Modellen wird nicht hergestellt. Bis auf die Betrachtung von Outsourcing-Modellen werden durch den hier entwickelten Kriterienkatalog die genannten Aspekte abgedeckt. Der Anhang der SPARC-Studie zählt Institutionen und deren Repositories auf, wobei nur Systeme berücksichtigt werden, die unterschiedliche logische Dokumenttypen, also z. B. nicht nur Dissertationen oder disziplinspezifische Publikationsformen, aufnehmen können.

Der Bericht *A Guide to Institutional Repository Software*⁴ des Open Society Institute (OSI) wendet sich an Institutionen, die mit den grundsätzlichen Fragen, die mit der Implementierung eines Repositories verbunden sind, vertraut sind. Neben einer zusammenfassenden Systembeschreibung findet sich ein sehr detaillierter, schematisierter Vergleich weitgehend technisch-funktionaler Eigenschaften. Der Erhaltungsaspekt wird punktuell thematisiert, nämlich bei *Data Preservation Support* untergliedert in *Defined Digital Preservation Strategy*, *Preservation Metadata Support* und *Data Integrity Checks*. Die erwähnten Erhaltungsaspekte sind im Kriterienkatalog aufgegriffen. Die funktionalen, technischen Detailkriterien können relativ einfach den mehr auf der konzeptionellen Ebene formulierten Kriterien unseres Schemas zugeordnet werden. OSI hat für die detaillierte Untersuchung nur Produkte herangezogen, die als Open Source lizenziert sowie freigegeben und öffentlich verfügbar sind und sich darüber hinaus OAI-konform zeigen.

Der *Technology Watch Report* der Digital Preservation Coalition (DPC) mit dem Titel *Institutional Repositories in the Context of Digital Preservation*⁵ diskutiert Anforderungen an ein Repository aus dem Blickwinkel der digitalen Erhaltung. Unter anderem werden Erhaltungsfunktionen beschrieben, die beim Entwurf eines

**Einführung und
Betrieb institutioneller
Repositories**

Repository-Bericht des OSI

**Report der Digital
Preservation Coalition
(DPC)**

Repositories zu berücksichtigen sind: eindeutige und persistente Identifikation, Aufnahme (Ingest), Repräsentationssystem, Technologiebeobachtung, Darstellung (Rendering), Systemarchitektur und Aufzeichnung von Änderungen der Metadaten (Historie). Den Repräsentationsinformationen (Representation Information) wird besondere Bedeutung für die Wiederherstellung der kodierten intellektuellen Inhalte beigemessen. Die funktionalen Anforderungen des DPC-Berichts für die Erhaltung sind in unseren Kriterienkatalog eingeflossen.

Daneben gibt es weitere Untersuchungen durch das Open Archives Forum⁶, die Brigham Young University⁷ und DigiCult⁸. Sie beschäftigen sich mit Anforderungen und Produkten aus dem Bereich digitaler Repositories. Die Langzeitarchivierung wird hierbei nicht weiter vertieft. Hervorzuheben ist die Research Libraries Group (RLG), die sich in ihrem Bericht *Trusted Digital Repositories: Attributes and Responsibilities*⁹ explizit der Langzeitarchivierung widmet. Sie definiert ein Rahmenwerk für die Eigenschaften und Verantwortlichkeiten eines vertrauenswürdigen, verlässlichen und nachhaltigen Repositories, das im Stande ist, den Umgang mit unterschiedlichsten Materialien großer und kleiner Forschungseinrichtungen zu beherrschen.

In den bisherigen Arbeiten ist eine Vielzahl von Kriterien entstanden. Folgende, für eine einheitliche Bewertung nachteilige Eigenschaften sind jedoch festzustellen:

- **Technologienahe Kriterien:**
Die Bewertung erfolgt über die Nennung technischer Lösungen ohne dabei eine konzeptionelle Einordnung der Ansätze vorzunehmen. Diese Vorgehensweise stützt sich damit auf teilweise sehr detaillierte, kurzlebige Kriterien. Dies steht auch der Abschätzung des Entwicklungspotenzials von Produkten im Wege.
- **Unterschiedliche Begrifflichkeit:**
Die einzelnen Untersuchungen bedienen sich für die Beschreibung der Produkte jeweils eigener Begrifflichkeiten, teilweise als Mischung anwendungsspezifischer Begriffe und technologischer Begriffe.
- **Mangelnde Allgemeingültigkeit:**
Bedingt durch die vorgegebenen Untersuchungsziele werden die Produkte oft nach sehr individuellen Gesichtspunkten und Anwendungsszenarien ausgewählt und bewertet.
- **Vernachlässigung nicht-funktionaler Aspekte:**
Kriterien, die den Aufwand, z. B. für den laufenden Betrieb, betrachten oder Kriterien, die die Qualität des Produktes betreffen, hier insbesondere die korrekte Umsetzung der funktionalen Eigenschaften in eine Implementierung, werden meistens nicht oder nur stark vergrößert angeführt.

ENTWICKLUNG VON KRITERIEN: FUNKTIONALE ASPEKTE OAIS-Referenzmodell

Das *Open Archival Information System* ist ein Referenzmodell, das eine als Archiv bezeichnete Organisation aus Personen und Systemen beschreibt, die die Verantwortlichkeit übernommen hat, Informationen zu erhalten und sie für eine bestimmte Zielgruppe (*Designated Community*) verfügbar zu machen. Dabei wird ausdrücklich darauf hingewiesen, dass es sich nicht um die Spezifikation eines Entwurfs oder einer Implementierung handle und tatsächliche Implementierungen die Funktionalität anders gruppieren oder aufbrechen könnten. Somit ist es für einen direkten Vergleich mit einer Implementierung bzw. für eine Entnahme von Bewertungskriterien ungeeignet. Außerdem beschreibt OAIS einerseits Funktionalitäten, die sich auf absehbare Zeit einer softwaremäßigen Implementierung weitgehend entziehen (z. B. *Preservation Planning*), andererseits fehlen Systemmerkmale, die für eine Bewertung einer konkreten Implementierung nötig sind (z. B. Kosten). Die Kernfunktionalität eines Archivierungssystems wird jedoch durch das OAIS-Referenzmodell erfasst. Somit kann es zur Gliederung der funktionalen Kriterien des Kriterienkatalogs und zur Überprüfung der Vollständigkeit dienen. Abweichungen von OAIS ergeben sich aber durch folgende Gesichtspunkte:

- Funktionalitäten außerhalb des funktionalen Modells von OAIS sollen in die Bewertung einfließen können. Dies wäre z. B. die Unterstützung bei der Erstellung von *Submission Information Packages* oder bei der direkten Auswertung von digitalen Datenobjekten.
- Reduzierung der *funktionalen Entitäten* (bzw. deren Teile) des funktionalen Modells von OAIS – neben den bereits erwähnten zurzeit nicht implementierungsrelevanten Funktionen. Auf konzeptioneller Ebene kann die Entität *Data Management* entfallen, da sie als querschnittlicher Dienst (Basisdienst) für die übrigen Entitäten aufgefasst werden kann. Auch Teilfunktionen, wie Fehlerprüfung (*Error Checking in Archival Storage*), können als Basisdienst betrachtet werden.
- Verwendung von Begriffen, die (zumindest bisher) in den bestehenden Produkten bzw. deren zu Grunde liegenden Modellen gebräuchlich sind.
- Modifikation des Informationsmodells: Im Gegensatz zur *Content Information* in OAIS, wo Objekte und *Repräsentationsinformationen* zu *Informationsobjekten* zusammengefasst sind, verwenden wir hier eine Position, in der *Repräsentationsinformationen* als spezielle Metadaten betrachtet werden (traditionelle Sicht: Daten + Metadaten).

Inhaltsmodell

Da sich bisher keine konkreten Methoden und insbesondere keine konkreten Datenmodelle zur Langzeitarchivierung digitaler Objekte etabliert haben und sie sich nicht direkt aus OAIS herleiten lassen, kommt den Eigenschaften eines Systems, die künftige Lösungen berücksichtigen können und vor allem eine Migration (Zugriff / Transformation / Restrukturierung) des Inhalts und seiner Organisation aus bisherigen Systemen bzw. Systemstrukturen erlauben, besondere Bedeutung zu. Daher müssen zur Bewertung des Langzeiterhalts auch Standards und Produkte betrachtet werden, die diesen Aspekt nicht direkt ansprechen. Die Verwendung und der Umgang mit folgenden Strukturarten werden dabei von uns als relevant erachtet:

- Objekte, Objektorganisation
- Metadaten, Metadatenorganisation
- Organisation Objekte – Metadaten
- Sicherung der Integrität dieser Organisation.

Dies charakterisiert funktionale, bezogen auf die Langzeitarchivierung implizite, inhaltsorientierte Systemeigenschaften. Sie erlauben es, z. B. die Migrierbarkeit und Restrukturierbarkeit von Inhalten abzuschätzen. Da wir auch Varianten und Versionen als wichtige Strukturen mit jeweils eigener Semantik erachten, sind sie im Kriterienkatalog als Merkmale gesondert aufgeführt.

ENTWICKLUNG VON KRITERIEN: NICHT-FUNKTIONALE ASPEKTE

Neben den funktionalen Eigenschaften spielen weitere, so genannte nicht-funktionale Systemeigenschaften für eine Bewertung eine Rolle. Diese werden unter den Kriterienblöcken *Aufwand* und *Qualität* behandelt.

Aufwand

Der Aufwand für Aufbau und Betrieb eines Archivierungssystems wird stark von der individuellen Anwendungssituation bestimmt. Maßgeblichen Einfluss haben der Gesamtumfang der digitalen Objekte, ihre Art und die Heterogenität und die Anzahl der gleichzeitig zu bedienenden Nutzer und Produzenten (I/O-Breiten, Lizenzkosten je nach Modell, Nutzerverwaltung). Ebenfalls hohen Einfluss haben Art und Umfang der Dienste (z. B. bei der Erzeugung von Metadaten) und ihre Qualität (z. B. zeitliche Verfügbarkeit, die u. a. durch Redundanz und schnelles Backup / Recovery erkaufte werden muss), die ein Archivierungssystem bereitstellen soll. Je nach Vollständigkeit des Systems kann auch der Entwicklungsaufwand für die Anpassung an individuelle Anforderungen sehr hoch sein. Gegebenenfalls spielen vorhandene Hard- und Softwaresysteme sowie vorhandene Bestände an Inhalten (Daten und Metadaten sowie ihre Organisation) oder gar die Einbettung in weitere Anwendungssysteme eine große Rolle. Besonders schwierig ist der Aufwand für die Langzeiterhaltung abzuschätzen. Dies liegt einerseits an den nicht konsolidierten Mechanismen zur Langzeiterhaltung als auch an der Unvorhersehbarkeit technologischer Weiterentwicklungen. Entsprechende funktionale Eigenschaften (z. B. Umsetzung des Inhaltsmodells) und Qualitätsmerkmale können neben dem Verlustrisiko auch das Kostenrisiko mindern. Ein hinreichend genaues Kostenmodell kann nur durch die Kenntnis der individuellen Anwendung erstellt werden. Doch mit Hilfe des Kriterienkatalogs können wichtige Aufwandsfaktoren wie Produktkosten, personelle Ressourcen und Sachressourcen erfasst werden.

Aufwandsfaktoren für Aufbau und Betrieb eines Archivierungssystems

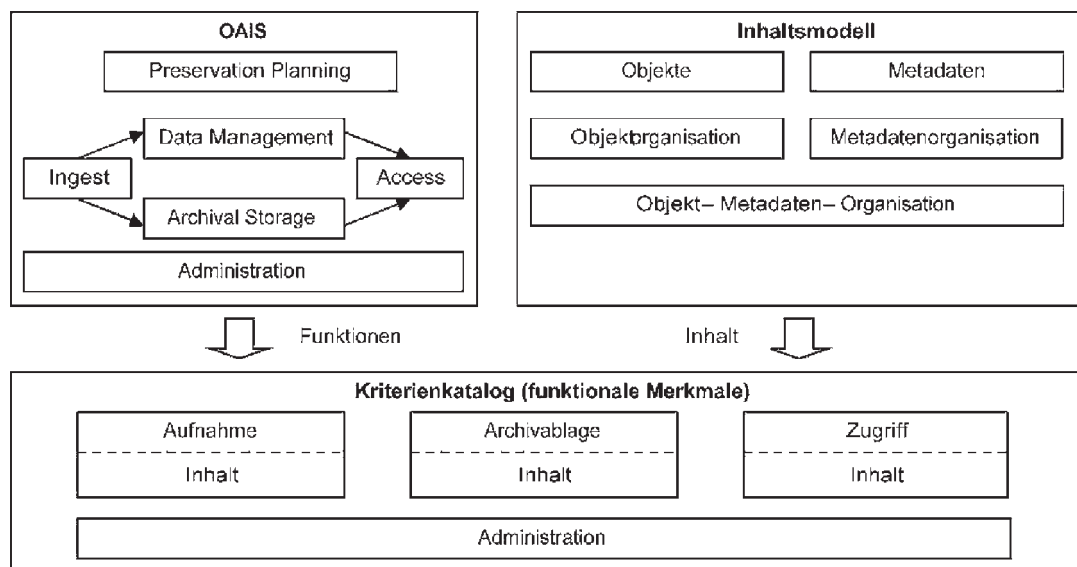


Abbildung 2: Grundlagen für funktionale Merkmale des Kriterienkatalogs

Qualität

Qualität wird hier im Sinne der Sicherstellung der Verfügbarkeit des Systems und – insbesondere unter Langzeitaspekten – der Verfügbarkeit der Inhalte gesehen. Nicht nur funktionale Kriterien bestimmen die Überlebensfähigkeit des Systems und ggf. eine Migrationsfähigkeit der Inhalte aus dem System in ein neues, sondern auch Eigenschaften, die insbesondere den Hersteller, die Dokumentation und die Stabilität als korrekte Umsetzung der funktionalen Eigenschaften betreffen. Eine hohe Verbreitung und Nutzerzahl, ver-

bunden mit der entsprechend wertvollen Datenmenge, können sich positiv auf eine langfristige Sicherung der Inhalte auswirken.

Der Kriterienkatalog

Neben den bisher entwickelten Kriterienarten *funktionale Komponenten* sowie *Aufwand* und *Qualität* enthält das Schema noch Allgemeines, Sonstiges und eine Zusammenfassung, die das Gesamtsystem komponentenübergreifend charakterisiert.

Im Rahmen der Expertise wurde für sechs Archivie-

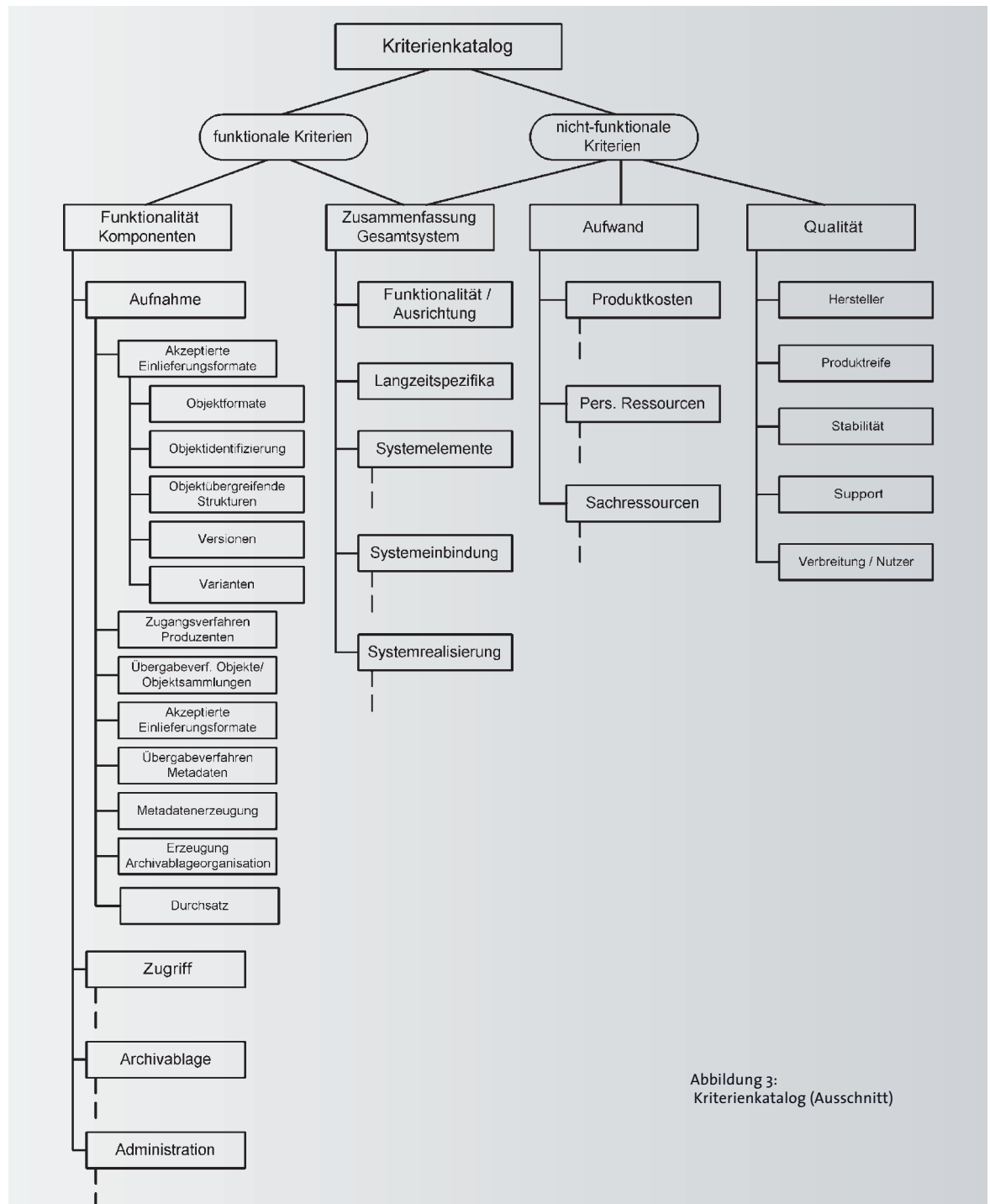


Abbildung 3:
Kriterienkatalog (Ausschnitt)

nungssysteme (DIAS, DigiTool, DSpace, EPrints, Fedora, MyCoRe) mit Verwendung des Kriterienkatalogs eine Produktbeschreibung erstellt. Damit sollte die Anwendung bei unterschiedlichen Konzepten und Systemarchitekturen gezeigt werden. Die Heranziehung dieser sechs Systeme soll nicht als Vorauswahl angesehen werden.

RESÜMEE DER NESTOR-EXPERTISE

Neben der hier beschriebenen Herleitung eines Kriterienkatalogs hatte die Expertise eine Marktsichtung zum Gegenstand. Auf der Basis dieser Marktsichtung und der dazu erstellten Produktbeschreibungen (für weitere 60 Produkte existiert eine Kurzbeschreibung) lässt sich ein Resümee über die konzeptionelle und technische Ausrichtung von Archivierungssystemen, deren Planungen und Entwicklungstendenzen sowie Aspekte der Langzeitarchivierung ziehen.

Konzeptionelle Ausrichtung:

Die Analyse der Archivierungsprodukte zeigt eine unterschiedliche Schwerpunktsetzung in der Abdeckung von Archivierungsfunktionalitäten. Dies spiegelt sich in einer Reihe von Systemeigenschaften wider. Die archivierbaren Objektarten und Metadaten werden durch die Zielgruppe und den Anwendungszweck des Systems oftmals eingegrenzt. Darüber hinaus versuchen etliche Systeme den gesamten Lebenszyklus eines digitalen Dokuments zu unterstützen. Dabei steht im Allgemeinen der Erstellungs- und Bearbeitungsaspekt im Vordergrund und damit verbunden der Einsatz spezieller Technologien wie Computer Aided Publishing. Aus Effizienzgründen und aufgrund der Konkurrenzsituation im Markt streben die Hersteller zusätzlich die Verwendung der aktuellsten, gelegentlich nicht ausgereiften oder noch in der Standardisierungsphase befindlichen Technologien an. Solche Produktausrichtungen können der Eignung des Systems als Archiv entgegenwirken. So kann der Zugang der Objekte an einen komplexen Prozess gekoppelt sein. Durch modulare Architekturen versuchen die Hersteller eine Entkopplung zu erreichen und somit ein breiteres Anwendungsfeld für ihre Produkte zu erschließen.

Technische Ausrichtung:

Die Systeme zeigen verschiedene Grade der Modularisierung wie etwa bezüglich der produktspezifischen Abhängigkeit der Module untereinander oder bezüglich der durch die einzelnen Module abgedeckten Funktionalität. So nutzen viele Produkte zur Abfrage und Verwaltung von Objekten und Metadaten keine eigene Entwicklung, sondern am Markt verfügbare Datenbanksysteme und Storagemanagementsysteme. Darüber hinaus wird die so genannte Anwendungslogik, z. B. die Darstellung von Objekten für

den Endnutzer, zunehmend auf Basis von Komponententechnologien realisiert. Eine Mischung von Komponenten aus dem Open-Source-Bereich und dem kommerziellen Sektor ist ebenfalls anzutreffen. Spezifische Leistungsmerkmale oder die Nichteinhaltung von Standards können den Austausch und die Freiheit bei der Auswahl von Komponenten jedoch einschränken. Insgesamt ist eine deutliche Tendenz zur Ausbildung von Kernsystemen mit definierten Schnittstellen zu erkennen.

Insbesondere der Einsatz von Web-Technologien verstärkt die Tendenz zur Modularisierung und zur logischen und physischen Verteilung der Systeme. Dies erlaubt grundsätzlich einen höheren Grad an Arbeitsteilung auf der fachlichen Ebene und den ortsunabhängigen Zugang zu Informationen. Ein derartiger kooperativer und föderativer Einsatz erfordert die Definition und Einhaltung von Konventionen. Technische Standards und Empfehlungen hierzu sind weitgehend akzeptiert und in den Produkten umgesetzt. Die Integration in bestehende Techniken ist jedoch teilweise unbefriedigend. Häufig anzutreffende Beispiele für technische Standards sind: Auszeichnungssprachen wie XML zur Beschreibung von Objektorganisation und Metadaten, XSLT zur Aufbereitung von Objekten zur Darstellung für den Endnutzer oder zur Übergabe an andere Systeme, Dublin Core zur Spezifikation von Metadaten. Hingegen werden spezifischere Standards wie METS nur von wenigen Systemen unterstützt. Fehlende Konventionen fördern die unkontrollierte Vergabe von Metadaten und die uneinheitliche Strukturierung von Objektsammlungen oder komplexen Einzelobjekten. Bestrebungen zur Entwicklung und Etablierung derartiger Konventionen sind sowohl auf Herstellerseite als auch auf Anwenderseite im Gange.

Planungen und Entwicklungstendenzen:

Die Hersteller kündigen für ihre Produkte Weiterentwicklungen der Funktionalitäten an. Aus Sicht der Archivierung sind hervorzuheben:

- Unterstützung der Metadatengewinnung und -erstellung mittels Extraktion oder mittels spezieller Editoren
- Schnelles und leichtes Wiederauffinden von Objekten
- Unterstützung der langfristigen Nutzung von Objekten
- Erweiterung der Rechteverwaltung
- Weitergehende Verwendung von Web-Technologien (z. B. Webservices) zur Erleichterung der Föderation und Kooperation
- Vereinfachte Verwaltung physischer Speichermedien

Marktsichtung

Verwendung nicht ausgereifter Technologien

- Erhöhung der Verfügbarkeit
- Erleichterung der Administration.

Für viele Produkte sind E-Commerce, gesetzliche Auflagen zur *revisionssicheren* (meist befristeten) Archivierung oder die Bereitstellung von Publikationen die treibende Kraft für diese Weiterentwicklungen. Die Langzeitarchivierung ist nur für wenige Systeme explizites Ziel.

Aspekte der Langzeitarchivierung:

Nur wenige Produkte realisieren Konzepte zur Unterstützung der Langzeitarchivierung. Selbst bei Produkten mit der Langzeitarchivierung als Zielsetzung sind die Konzepte nur zum Teil verwirklicht. Vorzufinden sind im Wesentlichen:

- Dateiformatregistrierung
- Handlesysteme zur persistenten Identifikation
- Übernahme von Konventionen (Standards) im Bereich der Metadaten
- Konvertierung in bestimmte Ablageformate
- Migrationsunterstützung z. B. durch einfach zu interpretierende Exportformate
- Beschreibung des Kontextes der technischen Nutzung z. B. durch Archivierung der Softwareanteile der Abspielumgebung
- Universal Virtual Computer (UVC).

Bezüglich der Anforderungen der Langzeitarchivierung sind die bisher bestehenden Systeme eher als Zwischenlösung zu betrachten. Aus diesem Grund rücken Systemmerkmale, die die Überlebensfähigkeit eines Systems und der Inhalte bestimmen, in den Vordergrund (vgl. Kriterientransparenz).

Für das Überleben der Inhalte spielt das verwendete Inhaltsmodell eine besondere Rolle. Die Leistungsfähigkeit bisheriger Modelle ist bezüglich der Flexibilität zur Organisation von Objekten und Metadaten und der Festschreibung der Organisation sehr unterschiedlich. Schwachpunkte sind in diesem Zusammenhang:

- Flache oder starre Objekthierarchien
- Fehlende Definierbarkeit der Semantik der Objektbeziehungen
- Eingeschränkte Metadatenmodelle (fehlende Elemente zur Beschreibung von Repräsentationsinformation, eingeschränkte Zuordenbarkeit zu Objekten)
- Unkontrolliertes Vokabular für Metadatenelemente und vor allem deren Werte.

Bei der Auswahl eines Produktes für die Archivierung ist diesen Gesichtspunkten besondere Aufmerksamkeit zu widmen. Es existieren bereits Systeme, die zumindest konzeptionell die nötige Flexibilität für die Definition und Implementierung von Objekt- und Metadatenorganisation bieten (z. B. Systeme, die sich auf METS abstützen). Fehlende Standards können dann durch Maßnahmen des Anwenders der Systeme, z. B.

durch Festlegung und Einhaltung von Konventionen, ausgeglichen werden.

DISKUSSION UND AUSBLICK

Mit dem Kriterienkatalog steht ein wichtiges Hilfsmittel zur Durchführung des Entscheidungsprozesses bereit. Den Verantwortlichen eines Archivs obliegt es nun, ihre individuellen Anforderungen zu formulieren, die Werte für die Kriterien zu ermitteln, den Grad der Erfüllung zu bestimmen und das Produkt als Ganzes zu bewerten, um letztlich eine Entscheidung treffen zu können. Zu einer adäquaten Bewertung gehört also noch die Auswahl geeigneter Maße und Skalen für die einzelnen Kriterien sowie die Bestimmung eines Verfahrens zur Bildung eines Gesamtwertes für ein System. Häufig kommen hier Nominalskalen zum Einsatz, also die Benennung bestimmter Eigenschaften, z. B. beim Kriterium »akzeptierte Einlieferungsformate« die Namen von Dateiformaten. Sofern man eine sichere Aussage über aufzunehmende Objekte treffen kann, ließe sich in diesem Fall noch relativ einfach der Erfüllungsgrad quantifizieren. Bei Kriterien, bei denen Vagheit oder Unsicherheit auf Seite des Systems oder auf Seite der Anforderungen ins Spiel kommt, wird die Situation komplizierter. Außerdem ist die Verfeinerung eines Kriteriums vorzunehmen, wenn kein passendes Maß für ein Kriterium gefunden werden kann. Für die Ermittlung eines Gesamtwertes für das Softwaresystem muss u. a. untersucht werden, inwieweit sich die Erfüllungsgrade der Kriterien gegenseitig kompensieren können. Die wegen seiner Einfachheit beliebte Verwendung des arithmetischen Mittels (nach der entsprechenden Normierung der Skalen) kann leicht zu Fehlschlüssen führen. Insgesamt stellt aber die Entscheidungstheorie einen ausreichenden Satz von Methoden bereit, um hier den Entscheidungsprozess adäquat abzudecken.

Die eben gemachten Ausführungen deuten die Größe des erforderlichen Aufwands an, so dass sich die Frage stellt, ob für den Entscheidungsprozess nicht weitere Unterstützung geboten werden könnte. Vorstellbar wäre sogar eine Produktzertifizierung zur weitgehenden Abnahme der Entscheidung. Zu berücksichtigen wäre einerseits die Konfigurierbarkeit und Erweiterbarkeit von Softwaresystemen und andererseits die hohe Heterogenität der Anforderungen an Archive sowie der Randbedingungen. So lange jedoch keine allgemein anerkannten Techniken zur Langzeitarchivierung existieren, die u. a. eine Klärung des Informationsbegriffs für spezifische Archivierungsaufgaben voraussetzen, müsste sich die Zertifizierung, die ja in einem gewissen Maß eine Bewertung vorwegnimmt, auf Forderungen beschränken, die sich

Formulierung individueller Anforderungen

bestehende Systeme: Zwischenlösung

Produktzertifizierung

heute bereits hinreichend formulieren lassen und genügend Anwendungsszenarien abdecken. Somit bleibt bezüglich der Produktauswahl nach heutigem Stand eine hohe Verantwortung beim Betreiber eines Archivs. Zur Sicherung der Überlebensfähigkeit seiner archivierten Inhalte muss er neben der Produktauswahl und Konfiguration selbst für ergänzende Konzepte und technische Lösungen sorgen. Wegen des hohen Aufwandes für solche Konzepte und Lösungen sowie für Produktbeschreibungen ist kooperatives Handeln der Betreiber, Nutzer und Hersteller geboten.

¹ CCSDS: Reference Model for an Open Archival Information System (OAIS) – Blue Book. 2002; www.ccsds.org/documents/650xob1.pdf

² Die vollständige Studie ist unter dem Stichwort »nestor-Materialien« auf der Seite www.langzeitarchivierung.de verfügbar

³ Crow, Raym: Institutional Repository Checklist & Resource Guide. 2002. www.arl.org/sparc/IR/IR_Guide_v1.pdf

⁴ Open Society Institute: A Guide to Institutional Repository Software. 2nd Edition. 2004

www.soros.org/openaccess/pdf/OSI_Guide_to_Institutional_Repository_Software_v2.pdf

⁵ Wheatley, Paul: DPC Technological Watch Report, Institutional Repositories in the context of Digital Preservation. 2004. www.dpconline.org/docs/DPCTWf4word.pdf

⁶ Wang, Jing Yuan; Assion, Marcus; Matthaei, Birgit: Open Archives Forum, Inventories – Open Archives Software Tools. 2003. www.oaforum.org/otherfiles/tv-tools.pdf

⁷ Lund, William: Digital Object Library Products. In: RLG DigiNews 2002. www.rlg.org/preserv/diginews/diginews5-5.html

⁸ DigiCULT: Technology Watch Report 1. 2003. www.digicult.info/pages/pubpop.php?file=http://www.digicult.info/downloads/twr2003_01_low.pdf

⁹ RLG: An RLG-OCLC Report, Trusted Digital Repositories – Attributes and Responsibilities. 2002. www.rlg.org/longterm/repositories.pdf

DIE VERFASSER

Uwe Borghoff ist Professor für Informationsmanagement am Institut für Softwaretechnologie, Universität der Bundeswehr München, Fakultät für Informatik, Werner-Heisenberg-Weg 39, 85577 Neubiberg, E-Mail: borghoff@informatik.unibw-muenchen.de

Peter Rödiger ist wissenschaftlicher Mitarbeiter am Institut für Softwaretechnologie, Universität der Bundeswehr München, Fakultät für Informatik, Werner-Heisenberg-Weg 39, 85577 Neubiberg, E-Mail: roedig@informatik.unibw-muenchen.de

Jan Scheffczyk ist wissenschaftlicher Mitarbeiter am Institut für Softwaretechnologie, Universität der Bundeswehr München, Fakultät für Informatik, Werner-Heisenberg-Weg 39, 85577 Neubiberg, E-Mail: jan@informatik.unibw-muenchen.de

Franz Schmalhofer ist wissenschaftlicher Mitarbeiter am Institut für Softwaretechnologie, Universität der Bundeswehr München, Fakultät für Informatik, Werner-Heisenberg-Weg 39, 85577 Neubiberg, E-Mail: franz@informatik.unibw-muenchen.de

**kooperatives Handeln von
Betreibern, Nutzern und
Herstellern**