

H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, K. Huth (Hrsg.)

# nestor Handbuch

Eine kleine Enzyklopädie  
der digitalen Langzeitarchivierung

Version 2.3

Kapitel 12.4

Systematische Planung von  
Digitaler Langzeitarchivierung

nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung  
hg. v. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, K. Huth  
im Rahmen des Projektes: nestor – Kompetenznetzwerk Langzeitarchivierung und  
Langzeitverfügbarkeit digitaler Ressourcen für Deutschland  
nestor – Network of Expertise in Long-Term Storage of Digital Resources  
<http://www.langzeitarchivierung.de/>

Kontakt: [editors@langzeitarchivierung.de](mailto:editors@langzeitarchivierung.de)  
c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen,  
Dr. Heike Neuroth, Forschung und Entwicklung, Papendiek 14, 37073 Göttingen

Bibliografische Information der Deutschen Nationalbibliothek  
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen  
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter  
<http://www.d-nb.de/> abrufbar.

Neben der Online Version 2.3 ist eine Printversion 2.0 beim Verlag Werner Hülsbusch,  
Boizenburg erschienen.

Die digitale Version 2.3 steht unter folgender Creative-Commons-Lizenz:  
„Namensnennung-Keine kommerzielle Nutzung-Weitergabe unter gleichen Bedingungen 3.0  
Deutschland“  
<http://creativecommons.org/licenses/by-nc-sa/3.0/de/>



Markenerklärung: Die in diesem Werk wiedergegebenen Gebrauchsnamen, Handelsnamen,  
Warenzeichen usw. können auch ohne besondere Kennzeichnung geschützte Marken sein und  
als solche den gesetzlichen Bestimmungen unterliegen.

URL für Kapitel 12.4 „Systematische Planung von Digitaler Langzeitarchivierung“ (Version  
2.3): <urn:nbn:de:0008-20100305244>  
<http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-20100305244>



*Gewidmet der Erinnerung an Hans Liegmann (†), der als Mitinitiator und früherer Herausgeber des Handbuchs ganz wesentlich an dessen Entstehung beteiligt war.*

## 12.4 Systematische Planung von Digitaler Langzeitarchivierung

Hannes Kulovits, Christoph Becker, Carmen Heister, Andreas Rauber

*Durch ständige technologische Veränderungen weisen digitale Objekte eine geringe Lebensdauer auf. Digitale Langzeitarchivierung ist somit zu einer dringlichen Aufgabe geworden. Zur langfristigen Bewahrung digitaler Objekte müssen diese mit Tools zur Langzeitarchivierung bearbeitet werden. Die Wahl eines spezifischen Tools für die Format-Migrationen oder Emulationen und die Einstellung spezifischer Parameter ist jedoch eine sehr komplexe Entscheidung. Die Evaluierung, ob und zu welchem Grad potentielle Alternativen spezifische Anforderungen erfüllen und die Erstellung eines soliden Plans zur Erhaltung einer bestimmten Gruppe von Objekten lässt sich als „Planung von Langzeitarchivierung“ zusammenfassen. Derzeit wird die Langzeitarchivierungsplanung manuell, meist ad-hoc, mit wenig oder keiner Softwareunterstützung durchgeführt. Dieses Kapitel stellt einen Workflow vor, der hilft, diesen Planungsprozess zu systematisieren.*

### Einführung

Es gibt eine Reihe von Strategien und Tools, welche die digitale Langzeitarchivierung unterstützen, jedoch fehlt oftmals eine Entscheidungshilfe für die Auswahl der optimalen Lösung. Für die Wahl einer geeigneten Archivierungsstrategie und eines konkreten Tools müssen komplexe Anforderungen bedacht werden. Sorgsame Dokumentation und gut definierte Vorgehensweisen sind nötig um sicherzustellen, dass das Endergebnis zur Planung von Erhaltungsmaßnahmen den Anforderungen der jeweiligen Einrichtung, insbesondere den Nutzern der Objekte („Designated Community“) entspricht. Dies ist auch eine der Kernaufgabe von TRAC<sup>21</sup> und nestor<sup>22</sup>.

Eine sorgfältige Planung der digitalen Langzeitarchivierung unterstützt den Entscheidungsprozess zur Auswahl der optimalen Lösung, indem im Planungsprozess verfügbare Lösungsmöglichkeiten gegen klar definierte und messbare Kriterien evaluiert werden. Sie stellt eine Kerneinheit des Open Archival Information System (OAIS) Referenzmodells dar<sup>23</sup>, insbesondere im Funktionsmodell Preservation Planning – siehe Kapitel 4. Die Planung besteht aus einem konsistenten Workflow, der idealerweise zu einem konkreten Langzeitarchivie-

---

21 OCLC (2007)

22 nestor (2006)

23 CCDS (2007)

ungsplan („*preservation plan*“) führt. Für die Planung der digitalen Langzeitarchivierung muss der Planungsbeauftragte über mögliche Lösungswege, die auf die betreffenden Objekte anwendbar sind, informiert sein. Es wird ein vorzugsweise automatisierter Vergleich von Dokumenten und Objekten vor und nach der Verwendung einer Archivierungsstrategie (z.B. einer Migration oder Emulation) benötigt, um die Qualität der verwendeten Erhaltungsmaßnahme („*preservation action*“) zu evaluieren. Der Prozess der zur Auswahl der Erhaltungsmaßnahme geführt hat, sollte darüber hinaus wiederholbar und auch gut dokumentiert sein, um die Nachvollziehbarkeit sowohl der zu Grunde liegenden Entscheidungen als auch der Gründe für die Wahl der Erhaltungsmaßnahme zu gewährleisten.

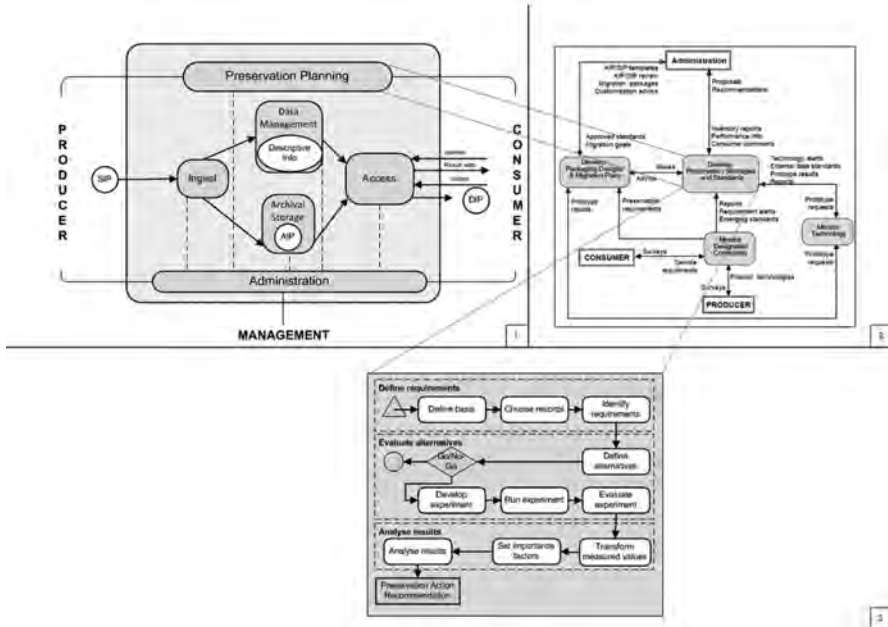


Abbildung 1: OAIS-Modell < Funktionale Entität „Preservation Planning“ < Planungs-Workflow

Der im Folgenden vorgestellte Workflow stellt eine Konkretisierung der funktionalen Komponente „Develop Preservation Strategies and Standards“ aus dem als ISO 14721 verabschiedeten OAIS Modell „Preservation Planning“ dar (Abbil-

dung 1). Der Workflow wurde ursprünglich im Rahmen des Preservation Clusters des EU NoE DELOS<sup>24</sup> (Network of Excellence on Digital Libraries)<sup>25</sup> konzipiert und nachfolgend im Rahmen des EU Projektes Planets<sup>26</sup> (Preservation and Long-Term Access via Networked Services) verfeinert.<sup>27</sup> Der Workflow basiert auf der Nutzwert-Analyse, einem Verfahren ähnlich der Kosten-Nutzen-Rechnung, kombiniert mit experimenteller Evaluierung.<sup>28</sup>

## Der PLANETS Workflow zur Langzeitarchivierung

### Anforderungserhebung („Define requirements“)

Die Phase 1 des Planungsverfahrens ist die Anforderungserhebung. Dazu gehören das Sammeln von Anforderungen von einer möglichst breiten Nutzergrup-

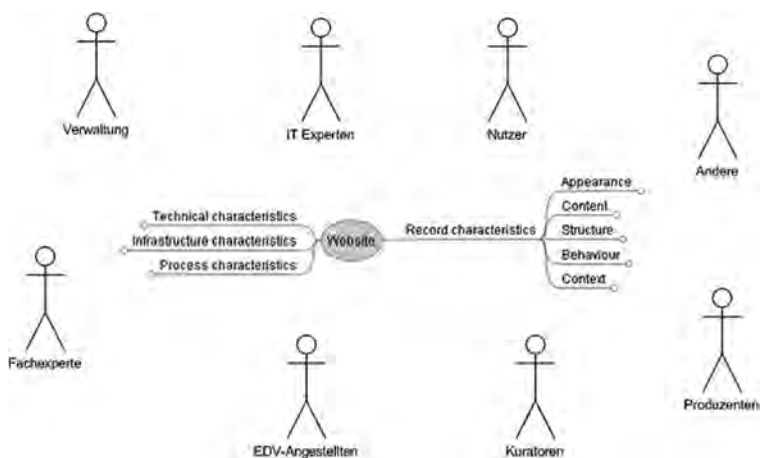


Abbildung 2: Experten, die die Anforderungen auswählen

pe (Abbildung 2), sowie der Faktoren der institutionellen Umgebung, welche die Langzeitarchivierung beeinflussen.

24 <http://www.delos.info/>

25 Strodl (2006)

26 <http://www.planets-project.eu>

27 Farquhar (2007)

28 Rauch (2004)

### *Evaluierung der Alternativen („Evaluate alternatives“)*

Die zweite Phase besteht in der Auswahl der in Frage kommenden Strategien, ihrer experimentellen Anwendung auf ausgewählte Beispielobjekte und der Evaluierung der Alternativen bezüglich der definierten Anforderungen.

### *Analyse der Ergebnisse („Analyse results“)*

In der dritten Phase werden die Alternativen in ihren Stärken und Schwächen verglichen und analysiert. Auf dieser Basis sind dann fundierte und gut dokumentierte Entscheidungen zur Auswahl der optimalen Strategie möglich.

### *Erstellen eines Plans zur Langzeitarchivierung*

#### *(„Build preservation plan“)*

Der Plan zur Langzeitarchivierung wird in der vierten Phase in der funktionalen Entität *„Develop Packaging Designs & Migration Plans“* im OAIS-Modell nach Genehmigung der empfohlenen Strategie in *„Administration“* erstellt. Er legt fest, welche Archivierungsmaßnahmen wie und von wem durchgeführt werden sollen. Änderungen an den Objekten, eine veränderte Umgebung oder neue Technologien machen es unter Umständen notwendig den Plan anzupassen. Eine Überwachung dieser Parameter und daraus resultierende Veränderungen am Plan bewirken einen ständigen Kreislauf im Planungsprozess.

## **Detaillierte Beschreibung des Workflows**

Im folgenden Abschnitt wird auf die drei Kernphasen des Workflows genauer eingegangen, da sich dieses Kapitel auf die Planungsphasen konzentriert.

### *Festlegen der Grundlagen („Define basis“)*

Im ersten Schritt der Phase 1 wird der Kontext des Planungsvorhabens dokumentiert. Dies beinhaltet den Namen des Planes sowie den Namen der Planungsverantwortlichen. Es wird der organisatorische Rahmen dokumentiert, welche Planungsziele die jeweilige Institution hat, was der Planungsgrund ist, welche Zielgruppe angesprochen wird, welche institutionellen Richtlinien zur Langzeitarchivierung existieren (vgl. Kap. 4.2) und welche rechtlichen Bedingungen, personellen sowie finanziellen Ressourcen und organisatorischen Einschränkungen für die Planung wichtig sind.

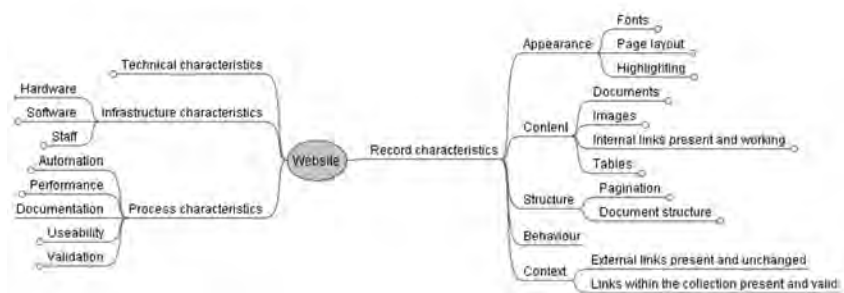


Abbildung 3: Anforderungsform als Mindmap

#### Auswahl der Datensätze („Choose records“)

Im zweiten Schritt werden repräsentative Beispielobjekte ausgewählt, welche die essenziellen Charakteristiken der gesamten Kollektion abdecken. In einem Planungsszenario für die Langzeiterhaltung von digitalen Dissertationen wären das beispielsweise: Eine Dissertation, die sehr viele Abbildungen enthält, eine sehr große bzw. sehr kleine Datei, eine Dissertation, die mathematische Gleichungen/Abbildungen enthält, und eine Datei, die interaktive Elemente beinhaltet. Diese Beispielobjekte werden im späteren Verlauf zur Evaluierung potenzieller Lösungen herangezogen. Üblicherweise werden drei bis fünf Objekte ausgewählt.

#### Identifizieren der Anforderungen („Identify requirements“)

Das Ziel dieses entscheidenden Schrittes ist die Dokumentation der Anforderungen für eine Archivierungsstrategie in klarer und eindeutiger Weise. Diese explizite Definition z.B. der bewahrten Eigenschaften ist eine Kernfordernis sowohl des nestor Kriterienkataloges (Punkt 9.3)<sup>29</sup> als auch des „TRAC“-Kataloges (Punkt B 2.1.)<sup>30</sup>. Allgemeine Ziele und detaillierte Anforderungen werden in einer Baumstruktur, dem sogenannten Kriterienbaum („Requirements tree“, „Objective tree“), gesammelt und strukturiert (Abbildung 3). Die Inhalte des Kriterienbaumes bilden die Basis der quantitativen Evaluierung und ermöglichen dadurch eine objektive Entscheidungsfindung. Während sich die Kriterien im Allgemeinen je nach Kontext und Szenario relativ stark unterscheiden, können einige allgemein gültige Prinzipien festgehalten werden - so hat es sich als

29 nestor (2006)

30 OCLC (2007)



zweckmäßig erwiesen, die Bäume auf der obersten Ebene in vier Hauptkategorien zu unterteilen:

- *Objekteigenschaften* („*Object characteristics*“) beschreiben Anforderungen, inwieweit visuelle und inhaltliche Erlebnisse des Benutzers bei der Betrachtung eines digitalen Objektes erhalten bleiben. Zur Beschreibung der wesentlichen Eigenschaften werden primär fünf Aspekte eines digitalen Objektes herangezogen: Inhalt, Aussehen, Struktur, Verhalten und inhaltlicher Kontext (Metadaten). Bei der nachfolgenden experimentellen Analyse wird gemessen, wie gut diese Eigenschaften bei Anwendung der Erhaltungsmaßnahme erhalten bleiben.
- *Datensatzesigenschaften* („*Record characteristics*“) beschreiben den technischen Kontext der Objekte beziehungsweise die verteilten Strukturen. So können z.B. in Powerpoint-Präsentationen Foliensätze, Videos oder Bilder über eine Verlinkung eingebettet sein oder Webseiten aus zahlreichen Komponenten wie z.B. „Styles-sheets“, „Bildern“, etc. aufgebaut sein. Diese Kriterien werden unter Umständen vom Anwender nicht unmittelbar wahrgenommen, wenn er mit dem digitalen Objekt zu tun hat. Trotzdem sind sie notwendig, um das Objekt in den Originalzustand zu überführen und damit seine ursprüngliche Wirkung und integrierte Funktionalität wieder herzustellen.
- *Prozesseigenschaften* („*Process characteristics*“) beziehen sich auf den Prozess beziehungsweise das Tool selbst. Sie beinhalten unter anderem die Skalierbarkeit eines Verfahrens auf große Datenmengen oder die Komplexität eines Verfahrens, aber auch Aspekte der Automatisierbarkeit, inwieweit manuelle Eingriffe notwendig sind, etc.
- *Kosten* („*Costs*“) spielen normalerweise eine wichtige Rolle beim Treffen der Entscheidungen. Sie können im Prinzip bei den jeweiligen Bereichen im Baum aufgeführt werden. Aus Gründen der besseren Gliederung werden sie aber meist in einem eigenen Zweig gebündelt. Sie können in technische Kosten und Personalkosten unterteilt werden sowie in Initialkosten und laufende Ausgaben.

Ein Kriterienbaum unterstützt die Identifikation von Kriterien und wird jeweils an einzelnen Stellen erweitert, an anderen reduziert, falls das eher den Vorstellungen einer Organisation entspricht. Zur vollständigen Identifikation der Kriterien ist meist eine ausführliche Recherche relevanter Literatur für das jeweilige Anwendungsgebiet sowie eine detaillierte Analyse der zu bewahrenden Objekte und Anforderungen erforderlich.

Die Kriterienbäume werden üblicherweise in Workshops erstellt, bei denen Anwender, Techniker und Archivierungsexperten zusammenarbeiten, um die relevanten Anforderungen zu ermitteln und zu strukturieren. Ein zentrales Element der Anforderungsanalyse in diesem Zusammenhang ist stets die quantitative Natur der Nutzwertanalyse. Jede Anforderung sollte soweit als möglich objektiv messbar gemacht werden. Daher wird jedem Kriterium in der untersten Ebene eine Skala zugewiesen, auf der die Erfüllung dieses Kriteriums gemessen wird. Soweit wie möglich sollten diese Kriterien objektiv und automatisch messbar sein, z.B. in Euro pro Jahr oder als prozentuelle Abweichung von der ursprünglichen Auflösung eines Bildes in Bildpunkten. In manchen Fällen müssen jedoch (semi-) subjektive Ordinalskalen zum Zuge kommen. Ein Beispiel dafür ist die Akzeptanz oder der Grad der Offenheit und Standardisierung eines Dateiformates.

Der erstellte Baum ist unabhängig von den betrachteten Alternativen; er dokumentiert die individuellen Anforderungen einer Institution oder Person in Bezug auf die langfristige Archivierung einer bestimmten Kollektion digitaler Objekte. Typischerweise enthalten entsprechende Bäume zwischen 20 und 150 Kriterien auf 3 bis 5 Ebenen. Die Anzahl der Kriterien hängt vor allem von der Art der Objekte ab – je komplexer die Inhalte sind, die in den Objekten abgebildet sind, desto aufwändiger ist die Erstellung des Baumes. Diese Komplexität spiegelt sich dann auch in der Anzahl der Kriterien wider.

#### *Alternativen definieren („Define alternatives“)*

Dieser Schritt betrachtet in Frage kommende Alternativen, wie beispielsweise Migration (vgl. Kapitel 8.3) oder Emulation (vgl. Kapitel 8.4). In diesem Schritt werden die verfügbaren Tools für die in Frage kommenden Strategien ausgewählt. Die Alternativen werden in diesem Schritt ausführlich beschrieben: Name der Alternative, Beschreibung der Alternative, Gründe, warum sie gewählt wurde, Konfigurierungsumgebung und Ressourcen, die für die Ausführung und Evaluierung nötig sind. Wichtig sind insbesondere die Versionsnummer eines Tools, die Parameter-Einstellungen, das installierte Betriebssystem, die Schriftarten, Programmbibliotheken etc.

#### *Fortfahren / Abbruch („Go/No-Go“)*

Unter Berücksichtigung der definierten Anforderungen, der Alternativen und einer Einschätzung der benötigten Ressourcen wird in diesem Schritt entschieden, ob der Prozess der Evaluierung fortgesetzt, abgebrochen oder verschoben werden soll. Außerdem wird entschieden, welche der aufgelisteten Alternativen

evaluiert werden sollen. Pro Alternative wird dokumentiert, weshalb sie in die engere Wahl gekommen ist oder verworfen wird. Beispielsweise kann es sein, dass für eine Alternative Hardware benötigt wird, die in der Anschaffung für die jeweilige Institution von vornherein viel zu teuer ist: Aus Kostengründen kann diese Alternative nicht evaluiert werden. Dieser Grund für die Entscheidung wird dann dokumentiert. Eine weitere Möglichkeit kann sein, dass eine neue Version eines Tools in naher Zukunft verfügbar sein wird. Diese Alternative kann dann in die Liste aufgenommen, die Evaluierung jedoch auf einen späteren Zeitpunkt verschoben werden („*Deferred-go*“).

#### *Experiment entwickeln („Develop experiment“)*

Um reproduzierbare Ergebnisse zu gewährleisten, wird in diesem Schritt ein Entwicklungsplan für jede Alternative spezifiziert, die das Experiment-Umfeld und die Art und Weise der Evaluierung mit einschließt. Dies umfasst die Rechnerumgebung, auf der die Experimente durchgeführt werden, die Konfiguration und das Aufsetzen der Messinstrumente (Zeitmessung etc.). Im Idealfall ist eine standardisierte Test-Umgebung vorhanden.

#### *Experiment durchführen („Run experiment“)*

Die betrachteten Alternativen werden nun in einem kontrollierten Experiment auf die gewählten Beispielobjekte angewandt. Das heißt, die Objekte werden mit den ausgewählten Tools migriert oder in den jeweiligen Emulatoren geöffnet. Dabei anfallende Fehlermeldungen bzw. Zeitmessungen sowie Ausgaben in Protokolldateien werden erfasst. Auch dieser Schritt kann durch die Verwendung von in zentralen Registries erfassten Tools, die über Webservices standardisiert aufgerufen werden können, drastisch vereinfacht werden.

#### *Experimente evaluieren („Evaluate experiments“)*

Um festzustellen, zu welchem Grad die Anforderungen im Kriterienbaum von den einzelnen Alternativen erfüllt werden, werden die Ergebnisse der Experimente evaluiert. Hierfür wird jedes einzelne Blatt im Kriterienbaum für jedes Objekt evaluiert. Die Evaluierung kann zum Teil automatisiert durch Analysetools unterstützt werden, welche die signifikanten Eigenschaften der Objekte vor und nach der Anwendung der Tools vergleichen und die Ergebnisse dokumentieren.

*Umwandeln/ Gleichsetzung der gemessenen Werte („Transform measured values“)*

Nach der Evaluierung der Kriterien im Kriterienbaum sind diese in unterschiedlichen Skalen (z.B. EURO, Sekunden, Farbe: ja/ nein) definiert. Damit die Kriterien vergleichbar und aggregierbar werden, wird pro Skala eine Transformationstabelle spezifiziert, welche die Werte der Messskala auf eine einheitliche Zielskala, den sogenannten Nutzwert abbildet. Die Zielskala ist üblicherweise eine Zahl zwischen 0 und 5, wobei 5 der beste Wert ist, während 0 ein nicht akzeptables Ergebnis darstellt.<sup>31</sup>

Das Kriterium „Proprietäres Dateiformat“ mit einer Boolean Skala „Yes/ No“ kann je nach Szenario unterschiedlich transformiert werden. Bei einer Transformation von „No“ auf den Wert „eins“ und „Yes“ auf den Wert „fünf“, wäre ein proprietäres Dateiformat zwar akzeptabel aber niedrig bewertet. Jedoch bei einer Transformation von „No“ auf den Wert „null“ (und „Yes“ auf den Wert „fünf“) wäre ein proprietäres Dateiformat ein Ausschlusskriterium für die gesamte Alternative.

Alternative	Total Score Weighted Sum	Total Score Weighted Multiplication
PDF/A (Adobe Acrobat 7 prof.)	4.52	<b>4.31</b>
PDF (unchanged)	<b>4.53</b>	<b>0.00</b>
TIFF (Document Converter 4.1)	4.26	3.93
EPS (Adobe Acrobat 7 prof.)	4.22	3.99
JPEG 2000 (Adobe Acrobat 7 prof.)	4.17	3.77
RTF (Adobe Acrobat 7 prof.)	3.43	<b>0.00</b>
RTF (ConvertDoc 4.1)	3.38	<b>0.00</b>
TXT (Adobe Acrobat 7 prof.)	3.28	<b>0.00</b>

Abbildung 4: Evaluierungsergebnisse elektronischer Dokumente

*Wertigkeiten festlegen („Set importance factors“)*

Die Kriterien, die im Kriterienbaum festgelegt worden sind, haben nicht alle die gleiche Wertigkeit für den Planenden. In diesem Schritt wird daher eine relative Gewichtung der Kriterien auf allen Ebenen durchgeführt, um der unterschiedlichen Bedeutung der einzelnen Ziele Rechnung zu tragen. Sind beispielsweise für eine Institution die Kosten sehr wichtig, werden sie in der Gewichtung höher gestuft als beispielsweise bestimmte Objekteigenschaften. Eine Institution, die beispielsweise eine sehr große Anzahl an Objekten migrieren muss, wird auf der höchsten Ebene des Kriterienbaums die Prozesseigenschaften etwas höher

31 Becker (2007)

gewichten als die übrigen. Folgende Gewichtung wäre denkbar: Objekteigenschaften (20%), Datensatzeigenschaften (20%), Prozesseigenschaften (40%) und Kosten (20%). Damit haben gute bzw. schlechte Prozesseigenschaften einen größeren Einfluss auf das Endergebnis.

#### *Evaluierungsergebnisse analysieren („Analyse evaluation results“)*

Im abschließenden Schritt werden die Ergebnisse aller Alternativen berechnet und aggregiert, um eine Kennzahl zu schaffen, die zum Vergleich der Alternativen herangezogen werden kann. Dabei können verschiedene Aggregationsmechanismen verwendet werden. Die wichtigsten Aggregationsmechanismen sind die Aufsummierung und die Multiplikation. Bei der Aufsummierung werden die transformierten Ergebniswerte jeder Alternative mit dem relativen Gewicht des entsprechenden Kriteriums multipliziert und über die Hierarchie des Baumes hinweg aufsummiert. Dadurch ergibt sich auf jeder Ebene eine Kennzahl zwischen null und fünf, die dem Erfüllungsgrad der entsprechenden Anforderung durch die betrachtete Alternative entspricht. Bei der Multiplikation dagegen werden die transformierten Werte mit dem relativen Gewicht potenziert und über die Hierarchie des Baumes hinweg multipliziert. Wiederum ergibt sich auf jeder Ebene eine Kennzahl zwischen null und fünf. Der wesentliche Unterschied zur Aufsummierung besteht darin, dass ein einzelnes nicht-akzeptiertes Kriterium zu einem Totalausfall der Alternative führt, da durch die Multiplikation der Wert „null“ bis in den Wurzelknoten durchschlägt. Das Ergebnis sind aggregierte Ergebniswerte für jeden Teilbaum des Kriterienbaumes und für jede Alternative. Eine erste Reihung der Alternativen kann auf den aufsummierten und multiplizierten Kennzahlen geschehen. Abbildung 4 zeigt die Bewertung von verschiedenen Alternativen mit Hilfe der zwei Aggregationsmethoden „Gewichtete Summe“ und „Gewichtete Multiplikation“. Der Hauptunterschied dieser zwei Aggregationsmethoden liegt in der Einflussnahme von nicht erfüllten Kriterien auf das Bewertungsergebnis der Alternative. Bei der Multiplikation scheidet Alternativen aus, d.h. sie werden mit 0 bewertet, falls ein oder mehrere Mindestkriterien nicht erfüllt werden. Die Alternativen RTF und TXT scheidet beispielsweise aus, weil sie große Nachteile in der Erhaltung der Struktur des Dokuments aufweisen. Die Alternative PDF („unchanged“) scheidet bei der Aggregationsmethode Multiplikation aus, da das essentielle Kriterium der Verhinderung von eingebetteten Skripten nicht erfüllt wird. Bei Aufsummierung wird die Alternative PDF („unchanged“) mit 4.53 knapp am höchsten bewertet, da nicht erfüllte Mindestkriterien kein Ausscheiden der Alternative verursachen, sondern normal in die Berechnung einfließen. Unter Berücksichtigung der Ergebnisse der beiden Aggregationsmethoden kann eine genaue Analyse der Stärken und Schwächen

jeder Alternative durchgeführt werden.

Das Ergebnis dieses Planungsprozesses ist eine konzise, objektive und dokumentierte Reihung in Frage kommender Alternativen für ein betrachtetes Archivierungsproblem unter Berücksichtigung der spezifischen situationsbedingten Anforderungen. Welche Lösung tatsächlich umgesetzt wird, hängt von den begleitenden Umständen ab. Aus der Nutzwertanalyse lässt sich jedoch eine klare Empfehlung ableiten, die mit direkt sichtbaren Argumenten hinterlegt und sorgfältig abgewogen ist und sich daher sehr gut als Entscheidungsgrundlage eignet. Durch die Darstellung sowohl allgemeiner als auch detaillierter Ergebniszahlen aus standardisierten und reproduzierbaren Testbedingungen wird eine solide Basis geschaffen, auf der wohlüberlegte und dokumentierte Entscheidungen getroffen werden können.

In der vierten Phase („Build preservation plan“) wird auf Basis der empfohlenen Alternative der Langzeitarchivierungsplan erstellt. Dieser Plan entspricht der „Develop Packaging Designs & Migration Plans“ Funktion im OAIIS-Modell (Abbildung 1).

## Das Planungstool Plato

Das EU-Projekt PLANETS entwickelt eine verteilte, serviceorientierte Architektur mit anwendbaren Services und Tools für die digitale Langzeitarchivierung<sup>32</sup>. Plato (PLANETS Preservation Planning Tool) (vgl. Kapitel 13.2) ist ein in PLANETS entwickeltes Planungstool, das den oben beschriebenen, in drei Phasen unterteilten Workflow implementiert und zusätzlich externe Services integriert, um den Prozess zu automatisieren.<sup>33</sup>

Eines dieser Services ist DROID (Digital Record Object Identification) von den National Archives UK. Damit kann automatisch die Bezeichnung des Dateiformats, die Version, der MIME-Type (Multipurpose Internet Mail Extensions) und der PUID (PRONOM Persistent Unique Identifier) ermittelt werden. Ein weiteres integriertes Service ist die Beschreibung des digitalen Objektes im XCDL-Format. Dieses Service wurde von der Universität Köln entwickelt und wandelt die ausgewählten Objekte in ein XCDL-Format um, welches für die spätere Evaluierung notwendig ist [5]. Zudem integriert Plato mehrere Registries, aus denen zu den Beispielobjekten passende Erhaltungsmaßnahmen ausgewählt und automatisch auf die Beispielobjekte angewendet werden können. Bestimmte Objekteigenschaften können automatisch gemessen und evaluiert werden.

Durch die Zuhilfenahme von frei verfügbaren Frameworks wie z.B. Java Ser-

---

32 Becker (2008b)

33 Becker (2009) Strodl, (2007)

ver Faces und AJAX wurde Plato als eine J2EE-Web-Applikation entwickelt, die frei verfügbar für Planungsvorhaben zur digitalen Langzeitarchivierung genutzt werden kann.<sup>34</sup>

## Zusammenfassung

In diesem Kapitel wurde der Planets Workflow zur Planung digitaler Langzeitarchivierungsvorhaben vorgestellt. Dieser Workflow ist die konkrete Ausarbeitung der Kerneinheit „Preservation Planning“ des mit dem ISO Standard 14721 verabschiedeten OAIS-Modells. Der Workflow erfüllt nach derzeitigem Wissenstand in den entsprechenden Bereichen die Anforderungen von Initiativen zur Zertifizierung und Validierung von vertrauenswürdigen Archiven, insbesondere nach TRAC<sup>35</sup> und dem nestor - Kriterienkatalog für vertrauenswürdige digitale Langzeitarchive<sup>36</sup>.

## Literaturverzeichnis

- Becker, Christoph / Rauber, Andreas (2007): *Langfristige Archivierung digitaler Fotografien*. Wien.
- Becker, Christoph / Kulovits, Hannes / Rauber, Andreas / Hofman, Hans. (2008b): *Plato: a service-oriented decision support system for preservation planning*. In: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries. 2008. S. 367-370.
- Becker, Christoph / Rauber, Andreas / Heydegger, Volker / Schnasse, Jan / Thaller, Manfred. (2008c): *A Generic XML Language for Characterising Objects to Support Digital Preservation*. In: Proceedings of the 2008 ACM symposium on Applied computing. 2008. S. 402-406
- Becker, Christoph / Kulovits, Hannes / Guttenbrunner Mark / Strodl Stephan / Rauber An-dreas / Hofman, Hans (2009) Systematic planning for digital preservation: Evaluating potential strategies and building preservation plans. In: International Journal on Digital Libraries (IJDL)
- CCDS Consultative Committee for Space Data Systems (Hrsg.) (2002): *Reference model for an open archival information system (OAIS)* / Consultative Committee for Space Data Systems. [public.ccsds.org/publications/archive/650x0b1.pdf](http://public.ccsds.org/publications/archive/650x0b1.pdf)

---

34 <http://www.ifs.tuwien.ac.at/dp/plato> (12.02.2010)

35 OCLC (2007)

36 nestor (2006)

- Farquhar, Adam. / Hockx-Yu, Helen (2007) *Planets: Integrated services for digital preservation*. In: International Journal of Digital Curation, 2. (2007). S. 88-99.
- National Library of Australia, Unesco. Information Society Division (Hrsg.) (2005): *Guidelines for the preservation of digital heritage. Prepared by the National Library of Australia*. <http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>
- nestor-Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung (Hrsg.) (2006): *Kriterienkatalog vertrauenswürdige digitale Langzeitarhive*. Version 2. (nestor-Materialien 8). Frankfurt am Main: nestor. [www.langzeitarchivierung.de/downloads/mat/nestor\\_mat\\_08.pdf](http://www.langzeitarchivierung.de/downloads/mat/nestor_mat_08.pdf)
- OCLC Online Computer Library Center, CRL The Center for Research Libraries (Hrsg.) (2007): *Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist*. Chicago, Dublin: Center for Research Libraries, OCLC Online Computer Library Center. <http://www.crl.edu/PDF/trac.pdf>
- Rauch, Carl / Rauber, Andreas (2004): *Preserving digital media: Towards a preservation solution evaluation metric*. In: Chen, Zhaoneng et al.: Proceedings of the 7<sup>th</sup> International Conference on Asian Digital Libraries (ICADL 2004). Berlin: Springer. S. 203-212.
- Strodl, Stephan / Rauch, Carl / Rauber, Andreas / Hofman, Hans / Debole, Franca / Amato, Guiseppe (2006): *The DELOS Testbed for Choosing a Digital Preservation Strategy*. In: Lecture Notes in Computer Science: Proceedings of the 9th International Conference on Asian Digital Libraries (ICADL 2006). Berlin, Heidelberg: Springer. S. 323-332.
- Strodl, Stephan / Becker, Christoph / Neumayer, Robert / Rauber, Andreas (2007): *How to Choose a Digital Preservation Strategy: Evaluating a Preservation Planning Procedure*. In: Proceedings of the ACM IEEE Joint Conference on Digital Libraries. 2007. S. 29 - 38.