Kernanforderungen für ein digitales Langzeitarchivierungssystem

Einführung

Dieses Dokument enthält Empfehlungen zu zehn übergeordneten funktionalen Anforderungen für ein digitales Langzeitarchivierungssystem. ***Ein solches System wird definiert als die zur Langzeitarchivierung digitaler Inhalte verwendeten Anwendungen und Werkzeuge, die Prozesse wie Übernahme, Speicherung, Erhaltung und Zugriff ausführen.***

Der Schwerpunkt liegt auf den Kernanforderungen für die ***digitale Langzeitarchivierung*** und nicht auf den allgemeinen Anforderungen an ein Informationsmanagementsystem (z. B. Sicherheit, Backup) oder den detaillierten Anforderungen einer bestimmten Instanz eines Langzeitarchivierungssystems (z. B. Anzahl der aufzubewahrenden Kopien). Dieses Dokument ist kein Versuch, Referenzanforderungen für jedes digitale Langzeitarchivierungssystem oder jeden Kontext aufzustellen, sondern vielmehr jene Merkmale zu erfassen, die unserer Ansicht nach ein digitales Langzeitarchivierungssystem typischerweise von anderen Anwendungen und Werkzeugen unterscheiden, die zur Verwaltung digitaler Inhalte verwendet werden.

Dieses Dokument ist Teil des Procurement Toolkit der DPC, das Ratschläge zu verschiedenen Ansätzen für die Beschaffung von Systemen und Diensten von Drittanbietern enthält. Es wird empfohlen, dieses Dokument in Verbindung mit den anderen Komponenten des Toolkits zu verwenden, einschließlich Lessons Learned und Common Requirements.

Nutzen

Dieses Dokument zielt darauf ab, die Beschaffung digitaler Langzeitarchivierungssysteme sowohl für Beschaffende als auch für Dritte, die an Beschaffungsverfahren teilnehmen, zu vereinfachen und zu verbessern.[1] Organisationen, die digitale Langzeitarchivierungssysteme beschaffen, können diese Kernanforderungen als Ausgangspunkt übernehmen und sich dann darauf konzentrieren, zusätzliche Funktionen zu identifizieren, die für sie wichtig sind (möglicherweise spezifisch für ihren eigenen organisatorischen Kontext, Arten von zu verwaltenden Inhalten, bestehende technologische Landschaft oder Zertifizierungswege).

Dieses Dokument kann auch als Lehrmittel verwendet werden, zum Beispiel wenn eine Praktikerin oder ein Praktiker erklären muss, warum ein typisches IT-System möglicherweise nicht den Anforderungen der Langzeitarchivierung entspricht.

Verwendung

Die Anforderungen werden in diesem Dokument durch die Wahl der Wörter „muss“, „sollte“ und „könnte“ gewichtet. **Praktikerinnen und Praktiker sind aufgefordert, diese Gewichtungen und die Aussagen selbst zu modifizieren und eigene Anforderungen hinzuzufügen, um ihrem eigenen Bedarf gerecht zu werden.** Zusätzliche Anforderungen können sich auf die Integration mit anderen Systemen, Workflows für bestimmte Inhaltstypen oder zusätzliche Aufgaben des Lösungsanbieters (z. B. Datenmigration) beziehen. Ein separates Dokument ist verfügbar, in dem die Verwendung dieser Anforderungen zusammen mit dem Rapid Assessment Model der DPC erläutert wird.

[1] Herausforderungen rund um diesen Prozess (für alle Parteien) wurden zuvor in einem Workshop identifiziert, an dem DPC-Mitglieder und DPC-Unterstützer teilnahmen. https://www.dpconline.org/events/past-events/guide-to-dp-procurement-event

Ein Glossar mit Begriffen zur digitalen Langzeitarchivierung ist im Digital Preservation Handbook verfügbar.

Danksagungen und Feedback

Diese Anforderungen wurden in Zusammenarbeit mit der UK Nuclear Decommissioning Authority erstellt. Wir sind auch den DPC-Mitgliedern und DPC-Unterstützern dankbar, die ihre Einblicke und ihr Feedback zu dieser Arbeit beigesteuert haben. Diese Kernanforderungen für die digitale Langzeitarchivierung werden sich als Reaktion auf Feedback weiterentwickeln und dabei eng mit dem Procurement Toolkit gekoppelt bleiben.

10 Kernanforderungen für ein digitales Langzeitarchivierungssystem

1. **Das System muss die Integrität und Authentizität der digitalen Inhalte und Metadaten, die in seinem Bestand sind, präzise verzeichnen und verwalten. Es muss anhand von Prüfsummen verifizieren, dass digitale Inhalte im Laufe der Zeit nicht versehentlich oder böswillig geändert wurden, und mit Metadaten alle Maßnahmen aufzeichnen, die an den Inhalten vorgenommen wurden.**
   - Begründung: Um auf transparente und nachprüfbare Weise zu belegen, dass digitale Inhalte im Laufe der Zeit unverändert bleiben, und um sicherzustellen, dass geplante Änderungen vollständig dokumentiert werden.

2. **Das System muss über ein umfassendes Datenmodell verfügen, das es ermöglicht, die komplexe Struktur digitaler Objekte bei der Übernahme zu erfassen und über die Zeit hinweg genau wiederzugeben, während sie verwaltet und aufbewahrt werden.**
   - Begründung: Um es dem System zu ermöglichen, jeden beliebigen digitalen Inhalt zu speichern und die Beziehungen zwischen den Elementen komplexer digitaler Objekte aufrechtzuerhalten, damit sie ohne Verlust aufbewahrt werden können.

3. **Das System muss eine klare Ausstiegsstrategie zu anderen Systemen bieten und darf keine Anbieterbindung (vendor lock-in) aufweisen. Dies muss gewährleisten, dass die unvermeidliche Migration zu einem zukünftigen digitalen Langzeitarchivierungssystem möglich ist und sollte den Aufwand und das Risiko minimieren, die mit einer solchen Migration verbunden sind.**
   - Begründung: Um sicherzustellen, dass digitale Inhalte und alle zugehörigen Metadaten bei Bedarf aus dem System zur digitalen Langzeitarchivierung extrahiert werden können.

4. **Das System muss die Übernahme authentischer digitaler Inhalte und Metadaten ermöglichen.**
   - Begründung: Um sicherzustellen, dass Inhalte und Metadaten ohne Verlust oder Beschädigung übernommen und beschrieben werden können.

5. **Das System muss über die Möglichkeit verfügen, die Merkmale übernommener digitaler Inhalte zu bewerten und sie in zugehörigen Metadaten aufzuzeichnen.**
   - Begründung: Es ist wahrscheinlicher, dass die Erhaltungsverwaltung digitaler Inhalte möglich und erfolgreich ist, wenn das System Informationen über die Art des Inhalts hat.

6. **Das System muss Replikation und Speicherverwaltung unterstützen. Es muss in der Lage sein, mehrere Kopien übernommener digitaler Inhalte in unterschiedlichen Speichersystemen an unterschiedlichen geografischen Standorten zu speichern.**
   - Begründung: Eine effektive Speicherverwaltung mindert das Risiko einer Beschädigung oder des Verlusts von Inhalten.

7. **Das System sollte eine Erhaltungsplanung einschließlich Risikobewertung und den Entwurf und Test von Erhaltungsplänen für digitale Inhalte unterstützen.**
   - Begründung: Um verlässliche Pläne für den langfristigen Schutz digitaler Inhalte zu erstellen.

8. **Das System sollte Erhaltungsmaßnahmen unterstützen, die Erhaltungspläne zur Minimierung identifizierter Erhaltungsrisiken umsetzen.**
   - Begründung: Um die notwendigen Schritte zum Schutz digitaler Inhalte zu ermöglichen und den fortwährenden Zugriff sicherzustellen.

9. **Das System muss die Verwaltung digitaler Inhalte und Metadaten über die Zeit hinweg unterstützen.**
   - Begründung: Die Verwaltung und Erhaltung digitaler Inhalte über längere Zeiträume ist ein aktiver Prozess.

10. **Das System muss das geregelte Auffinden von und Zugreifen auf digitale Inhalte und Metadaten ermöglichen.**
    - Begründung: Digitale Inhalte werden so aufbewahrt, dass sie von anderen gefunden und genutzt werden können.

Anforderungen im Detail

1. **Das System muss die Integrität und Authentizität der digitalen Inhalte und Metadaten, die in seinem Bestand sind, präzise verzeichnen und verwalten. Es muss anhand von Prüfsummen verifizieren, dass digitale Inhalte im Laufe der Zeit nicht versehentlich oder böswillig geändert wurden, und mit Metadaten alle Maßnahmen aufzeichnen, die an den Inhalten vorgenommen wurden.**
   - 1.1 Das System muss für jede Datei Prüfsummen speichern.
   - 1.2 Das System muss in der Lage sein, Prüfsummen gegen die mit den Inhalten mitgelieferten Prüfsummen zu validieren.
   - 1.3 Das System muss regelmäßige Integritätsprüfungen unterstützen und beschädigte oder fehlende Dateien in diesem Zuge melden.
   - 1.4 Das System muss Schritte zum Reparieren oder Ersetzen beschädigter Dateien aus replizierten Kopien unterstützen und über durchgeführte Maßnahmen berichten.
   - 1.5 Das System muss in der Lage sein, ein Prüfprotokoll zu generieren und Ereignismetadaten (etwa die von PREMIS[2] geforderten) aufzuzeichnen, die alle an digitalen Inhalten vorgenommenen Maßnahmen beschreiben.

[2] https://www.loc.gov/standards/premis/v3/index.html

2. **Das System muss über ein umfassendes Datenmodell verfügen, das es ermöglicht, die komplexe Struktur digitaler Objekte bei der Übernahme zu erfassen und über die Zeit hinweg genau wiederzugeben, während sie verwaltet und aufbewahrt werden.**
   - 2.1 Das Datenmodell muss in der Lage sein, digitale Objekte zu erfassen und darzustellen, die aus mehreren hierarchischen Komponenten bestehen, wie z. B. Dateien, Entwürfe, veröffentlichte Versionen oder nachträglich für die Erhaltung oder den Zugriff erstellte Kopien. Digitalen Objekten sollten eindeutige und dauerhafte Identifikatoren zugewiesen werden.

3. **Das System muss eine klare Ausstiegsstrategie zu anderen Systemen bieten und darf keine Anbieterbindung (vendor lock-in) aufweisen. Dies muss sicherstellen, dass die unvermeidliche Migration zu einem zukünftigen digitalen Langzeitarchivierungssystem möglich ist und sollte den Aufwand und das Risiko minimieren, die mit einer solchen Migration verbunden sind.**
   - 3.1 Die Struktur von gespeicherten digitalen Inhalten kann ohne die Langzeitarchivierungssoftware verstanden/interpretiert werden.
   - 3.2 Das System muss die Fähigkeit zum Massenexport digitaler Inhalte und aller damit verbundenen Metadaten in handhabbarem Format und handhabbarer Struktur für die Übernahme in ein anderes System besitzen.
   - 3.3 Die Systemsoftware könnte treuhänderisch hinterlegt werden, um im Falle des Ausfalls des Anbieters eine gewisse Rückversicherung zu bieten.

4. **Das System muss die Übernahme authentischer digitaler Inhalte und Metadaten ermöglichen.**
   - 4.1 Das System muss die Übernahme digitaler Inhalte und zugehöriger Metadaten in großem Umfang ermöglichen.
   - 4.2 Inhalte sollten vor Übernahme auf Viren überprüft werden und geeignete Quarantäneeinrichtungen sollten vorhanden sein.
   - 4.3 Das System muss in der Lage sein, nicht nur den ursprünglichen Bitstream aufzubewahren, sondern auch andere Merkmale, die für die Erhaltung und den Zugriff auf den Inhalt erforderlich sind, etwa die ursprüngliche Ordnerstruktur, Dateiinformationen wie Datumsstempel, zugehörige Dokumentation und zugehörige Metadaten.
   - 4.4 Das System muss in der Lage sein, Management-Berichte über den Erfolg oder Misserfolg von Übernahmeaktivitäten bereitzustellen und sollte auch in der Lage sein, potenzielle Probleme mit übernommenen Inhalten zu melden (z. B. veraltete Dateiformate, nicht erkannte Dateiformate, fehlende Metadaten, passwortgeschützte oder verschlüsselte Dateien usw.).

5. **Das System muss die Merkmale übernommener digitaler Inhalte bewerten und sie in zugehörigen Metadaten aufzeichnen.**
   - 5.1 Das System muss Dateiformate bis zur Ebene einer bestimmten Dateiformatversion identifizieren und auf geeignete Verzeichnisse mit weiteren Informationen wie PRONOM und/oder Wikidata verweisen.
   - 5.2 Das System muss technische Merkmale (wie Größe, Bildabmessungen, Videocodec, Audiolaufzeit, Erstellungssoftware) extrahieren.
   - 5.3 Das System sollte Inhalte identifizieren, die nicht wiedergegeben werden können, wie z. B. defekte, schlecht konstruierte oder verschlüsselte Inhalte.

5.4 Das System sollte Dateiformate anhand von Dateiformatspezifikationen oder benutzerdefinierten Profilen validieren.

5.5 Das System sollte externe Abhängigkeiten festhalten, bei denen nicht im digitalen Objekt vorhandene Inhalte (oder Software) wesentlich für die Wiedergabe oder Verwendung sind (z. B. nicht eingebettete Schriftarten, nicht eingebettete Medien wie YouTube-Videos oder Softwarebibliotheken).

6. Das System muss Replikation und Speicherverwaltung unterstützen. Es muss in der Lage sein, mehrere Kopien übernommener digitaler Inhalte in unterschiedlichen Speichersystemen an unterschiedlichen geografischen Standorten zu speichern.

6.1 Das System muss die Replikation digitaler Inhalte an mehreren Speicherorten (potentiell an verschiedenen geografischen Standorten) automatisch verwalten.

6.2 Das System sollte regelmäßige Systemsicherungen durchführen.

6.3 Das System sollte in der Lage sein, seine Sicherungs- und Wiederherstellungsfähigkeiten regelmäßig zu testen und zu melden.

6.4 Das System sollte Management-Berichte über Replikations-, Speicherverwaltungs-, Sicherungs- und Wiederherstellungsaktivitäten erstellen und aufbewahren.

7. Das System sollte eine Erhaltungsplanung einschließlich Risikobewertung und den Entwurf und Test von Erhaltungsplänen für digitale Inhalte unterstützen.

7.1 Das System sollte die Identifizierung, Management und Analyse von Erhaltungsrisiken unterstützen (z. B., wenn digitale Inhalte in einem Dateiformat vorliegen, das nicht mehr unterstützt wird).

7.2 Das System sollte die Gestaltung, Entwicklung und Verwaltung von Plänen zur Minderung identifizierter Erhaltungsrisiken ermöglichen.

7.3 Das System sollte Berichte liefern, um die effektive Verwaltung digitaler Inhalte zu ermöglichen. Dies muss ein breites Spektrum an funktionalen, betrieblichen und statistischen Berichten und Analysen umfassen.

8. Das System sollte Erhaltungsmaßnahmen ermöglichen, die Erhaltungspläne zur Minimierung identifizierter Erhaltungsrisiken umsetzen.

8.1 Das System sollte die Migration von Dateien von einem Dateiformat in ein anderes ermöglichen.

8.2 Das System sollte die Wiedergabe digitaler Inhalte durch die Anwendung von Emulations- und/oder anderen Spezialwerkzeugen ermöglichen.

8.3 Das System sollte die Qualität der Ergebnisse jeder Erhaltungsmaßnahme sichern.

8.4 Das System muss Erhaltungsmaßnahmen (und deren Ergebnisse) in den zugehörigen Metadaten aufzeichnen.

9. Das System muss die Verwaltung digitaler Inhalte und Metadaten über die Zeit hinweg unterstützen.

9.1 Das System muss Kontrollen bereitstellen, um das Risiko einer versehentlichen oder böswilligen Löschung oder Inhaltsänderung zu minimieren.

9.2 Das System muss in der Lage sein, alle erforderlichen Metadaten aufzunehmen und zu verwalten, einschließlich für bestimmte Inhaltstypen angemessene Metadaten (z. B. Geoinformationen, audiovisuelle Inhalte).

9.3 Das System muss die Verbesserung von Metadaten und Dokumentation während der gesamten Lebensdauer des digitalen Inhalts ermöglichen.

9.4 Das System muss die geregelte Entfernung von Inhalten ermöglichen und die Aufbewahrung und Pflege von Metadaten zulassen, selbst wenn die zugehörigen digitalen Inhalte aus dem System entfernt wurden.

9.5 Das System muss alle Maßnahmen zur Verwaltung digitaler Inhalte auf Basis konfigurierbarer Benutzerrollen einschränken.

10. Das System muss das geregelte Auffinden von und Zugreifen auf digitale Inhalte und Metadaten ermöglichen.

10.1 Das System muss sicherstellen, dass alle digitalen Inhalte und alle zugehörigen Metadaten nur für autorisierte Benutzer auffindbar und zugänglich sind.

10.2 Das System könnte Benutzer warnen, wenn sie versuchen, auf möglicherweise problematische Inhalte zuzugreifen (z. B. digitale Inhalte, die in einem nicht unterstützten Dateiformat gespeichert sind, unvollständige Metadaten aufweisen, zusätzliche Berechtigungen erfordern usw.).

10.3 Das System könnte Emulationsmöglichkeiten, spezialisierte Viewer, On-the-Fly-Migration oder Hilfsanwendungen bereitstellen, um die Nutzung digitaler Inhalte in veralteten oder nicht unterstützten Formaten zu ermöglichen.

10.4 Das System muss eine Zugriffsschnittstelle für Benutzer und/oder eine ausreichend leistungsfähige API bieten, um die Integration mit anderen Benutzerzugriffssystemen oder Erkennungstools zu ermöglichen.

10.5 Das System sollte Informationen über den digitalen Inhalt beim Zugriff offenlegen, einschließlich Erhaltungsmetadaten und Prüfsummen, um eine Authentizitätskette nachzuweisen.

10.6 Metadaten und/oder digitale Inhalte sollten automatisiert abrufbar sein.