Rechtssichere Langzeitarchivierung Elektronischer Dokumente

Bundesministerium
für Wirtschaft
und Technologie

PTB

BundOnline 2005

Ansprechpartner

Ansprechpartner für das Projekt ArchiSafe in der
Physikalisch-Technischen Bundesanstalt Braunschweig:

Herr Dir. u. Prof. Dr. Siegfried Hackel,
Behördenansprechpartner im Rahmen der Initiative BundOnline 2005
Tel.: 0531-592-8400
Fax.: 0531-592-8406
E-Mail: siegfried.hackel@ptb.de

und

Herr Dipl. Wirt.-Inform. Tobias Schäfer,
Projektleiter
Tel.: 0531-592-2456
Fax.: 0531-592-692456
E-Mail: tobias.schaefer@ptb.de

Einführung

Elektronische Dokumente sind aus dem öffentlichen Verwaltungshandeln nicht mehr weg zu denken. Die papiergebundene Aktenführung und das Büro als organisatorische Grundeinheit klassischer Verwaltung verlieren angesichts der flächendeckenden Einführung IT-gestützter Vorgangsbearbeitung zunehmend an Bedeutung. eGovernment, gestern noch Modewort, ist heute unbestritten Inbegriff und Motor vom Modell einer künftigen elektronischen Verwaltung. Dabei steht schon jetzt fest, dass elektronisch hochgerüstete Front-Office-Innovationen auf der Basis des Internet auf die Dauer wenig Sinn machen, wenn im Hintergrund, im so genannten Back-Office, weiterhin Aktenordner, Telefax und Karteikarten vorherrschen. Denn das Back-Office ist die zentrale Produktionsstätte der Verwaltung, die Organisation des „arbeitenden Staates". Ihr zentrales Produkt ist die Akte. Der Aufbau moderner Verwaltungsstrukturen auf elektronischer Grundlage wird also vor allem auch sie betreffen, denn sie soll künftig als nunmehr elektronische Akte den ungehinderten und schnellen Austausch von Daten und Informationen zwischen den Verwaltungseinheiten zum Vorteil der Bürger und der Wirtschaft ermöglichen und unterstützen.

Ihr entscheidender Vorzug aber, als digitalisierte und codierte Information unmittelbar maschinenlesbar zu sein und als einfache Bits und Bytes selbst über große Entfernungen in Sekundenschnelle transportiert werden zu können, erweist sich dabei zugleich auch als ihre entscheidende Schwäche. Digitale Informationen sind nicht nur flüchtig, weil per se virtuell, sondern auch leicht und unbemerkt zu manipulieren. Wer also die elektronische Akte als eine wesentliche Vorraussetzung für weitere Fortschritte im eGovernment will, der muss sich heute schon Gedanken machen, wie er die Authentizität und Integrität, die Vertraulichkeit und Vollständigkeit digitaler Unterlagen nachhaltig, und das bedeutet mindestens für den Zeitraum gesetzlicher vorgeschriebener Aufbewahrungsfristen, gewährleistet. Elektronische Akten haben wie ihre Vorgänger im Papierformat im Rahmen des vom Gesetzgeber vorgeschriebenen ordnungsgemäßen Verwaltungshandelns über die unmittelbare Bearbeitung hinaus auch eine Nachweisfunktion zu erfüllen. Das zentrale Gebot der Aktenmäßigkeit des Verwaltungshandelns gilt zweifellos auch für die elektronische Akte.

Der Wechsel zu einer elektronischen Dokumenteninfrastruktur bleibt ohne ein adäquates elektronisches Archiv unvollständig. Elektronische Dokumente werden zunehmend signiert. Somit kann die langfristige und rechtssichere Aufbewahrung und Verfügbarkeit elektronischer Informationen mit einem adäquaten elektronischen Archiv gewährleistet werden.

Die Entwicklung eines solchen Archivsystems ist das Ziel des Projektes ArchiSafe der Physikalisch-Technischen Bundesanstalt im Rahmen der E-Government-Initiative BundOnline 2005.

Ziel des Projektes ArchiSafe

Die „Einer-für-alle"–Dienstleistung ArchiSafe der Physikalisch-Technischen Bundesanstalt (PTB) unterstützt und fördert die Einführung bundeseinheitlicher Standards für die rechts- und revisionssichere Langzeitspeicherung (Archivierung) elektronischer Dokumente. Mit der Begründung eines standardisierten Datenaustauschformats für die Nutzdaten (Inhaltsdaten, Beschreibungs- und Signaturdaten) im XML-Format und der Implementierung einer Software-Referenzarchitektur schafft ArchiSafe wesentliche Grundlagen für die Einführung und Nutzung sowohl zentraler als auch dezentraler, elektronischer Archive bis hin zum Bundesarchiv.

Im Rahmen des Projektes ArchiSafe sind gemeinsam mit dem Bundesamt für Sicherheit in der Informationstechnik, der Bundesnetzagentur, der KBSt, dem Bundesarchiv und Mitarbeitern aus mehr als 20 verschiedenen Bundes- und Landesbehörden Anforderungen und technische Lösungsmöglichkeiten für eine dauerhafte rechts- und revisionssichere elektronische Ablage rechtsverbindlicher und rechnungsbegründender elektronischer Unterlagen erörtert, bewertet und veröffentlicht worden (siehe http://www.archisafe.de), die die sichere und langfristig verkehrsfähige Aufbewahrung der Unterlagen gemäß den Anforderungen des Signaturgesetzes erfüllen.

Zusammengefasst gilt:

- Die elektronische Dokumenteninfrastruktur muss die langfristige Überlieferung elektronischer Dokumente mindestens unterstützen, d.h. der Zugriff auf elektronische Dokumente muss auch für längere Zeiträume mit einem vertretbaren wirtschaftlichen und zeitlichen Aufwand möglich sein.
- Die elektronische Dokumenteninfrastruktur muss die Rechtssicherheit, insbesondere die Authentizität und Integrität der elektronischen Dokumente dauerhaft, mindestens aber bis zum Erreichen der gesetzlich vorgeschriebenen Aufbewahrungsfristen sicherstellen, d.h. sowohl die bildliche und inhaltliche Übereinstimmung mit den originären Dokumenten gewährleisten als auch, im Falle elektronisch signierter Dokumente, die Signaturerneuerung gemäß § 17 Signaturverordnung ermöglichen.

Zu den bestimmenden Zielen und Vorgaben des Projektes ArchiSafe gehören daher:

- Verwendung eindeutig interpretierbarer, langfristig stabiler und veröffentlichter Nutzdatenformate
- Verwendung eindeutig interpretierbarer, langfristig stabiler und standardisierter Signaturdatenformate
- Berücksichtigung der Sicherheitseignung kryptographischer Algorithmen und Verwendung elektronischer Signaturen mit ausreichend hohem Sicherheitsniveau (qualifizierte elektronische Signatur)
- Archivierung erforderlicher Verifikationsdaten in verkehrsfähiger Form
- Rechtzeitige und beweiskräftige Signaturerneuerung
- Stabile Verfügbarkeit technischer Komponenten
- Sichere Transformation elektronisch signierter Dokumente
- Gewährleistung des Daten- und Geheimnisschutzes
- Erhöhte Sicherheit durch Redundanz bei der Speicherung und Erneuerung elektronisch signierter Dokumente
- Kosteneffizienz durch Nachnutzbarkeit, Skalierbarkeit und den Einsatz standardisierter und wirtschaftlicher Techniken und Technologien

Die ArchiSafe Architektur

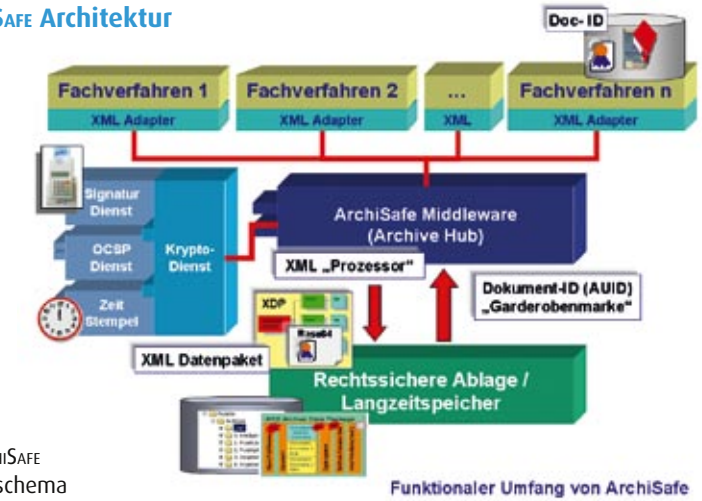

Abb. 1: ArchiSafe Architekturschema

Technisch gesehen basiert das ArchiSafe-Konzept auf einer service-orientierten, geschichteten und mandantenfähigen Softwarearchitektur.

Ein Fachverfahren (z.B. ein Dokumentenmanagement- oder Vorgangsbearbeitungssystem) dient als Plattform und führendes System zur Dokumentenverwaltung, Vorgangsbildung und Schnittstelle zur Langzeitspeicherung. Das Fachverfahren initiiert die Ablage elektronischer Dokumente im elektronischen Archiv und verwaltet die vom Archivsystem erzeugten Dokumentkennungen für die im Langzeitspeicher abgelegten Dokumente. Hierzu gehört, die Dokumentkennungen gespeicherter Dokumente mit den für die operative Vorgangsbearbeitung vorgehaltenen Dokumentinstanzen und Prozessdaten zu verknüpfen. (s. Abb. 1)

Das Fachverfahren kommuniziert mit dem Langzeitspeicher über eine einheitliche Archiv-Schnittstelle (Archivservice) zur Übergabe der zu archivierenden Objekte (Dokumente, Vorgänge, Akten) an die Langzeitspeicherung. Der Archivservice für die Übergabe von elektronischem Schriftgut aus den Fachanwendungen an das Langzeitspeichersystem wird in einer systemunabhängigen Middleware-Komponente (Archive-Hub) abgebildet (s. Abb. 1). Der Archivservice prüft und verarbeitet die Archivobjekte auf der Grundlage standardisierter Zeichensätze und Datenformate, sowie syntaktischer und semantischer Vereinbarungen für die Strukturen der im Langzeitspeicher abzulegenden Datenobjekte. Der Archivservice fordert darüber hinaus gegebenenfalls kryptografische Funktionen wie Signaturen, Zertifikatsprüfungen und Zeitstempel an und verarbeitet XML-Formate auf der Basis definierter XML-Schemata.

Die Erzeugung der standardisierten Archivobjekte und die Kommunikationseröffnung mit dem Archivservice erfolgt in fachspezifischen Serviceschnittstellen (Service-Adaptern).

Im Kern besteht der Archivservice aus einem XML Prozessor mit definierten Schnittstellen (Kommunikationskanälen) zu den Fachverfahren und zum elektronischen Langzeitspeicher. Darüber hinaus soll der Archivservice die Anbindung zusätzlicher Dienste, wie beispielsweise einem Signaturdienst (zur Erzeugung und/oder Prüfung elektronischer Signaturen) und einem Zeitstempeldienst ermöglichen.

Die kryptografischen Dienste signieren auf Anforderung durch das Fachverfahren die in den elektronischen Langzeitspeicher einzustellenden Dokumente oder versehen sie mit einem Zeitstempel. Sie verifizieren zudem auf Anforderung durch das Fachverfahren die Signaturen und Zertifikate signierter Dokumente und stellen dem Archivservice die Verifikationsdaten zur Verfügung, der diese wiederum für eine spätere Nachweisfunktion in einem standardisierten Format in die Archivobjekte einbettet.

Im Projekt ArchiSafe werden die kryptografischen Dienste über das Kernsystem der Virtuellen Poststelle des Bundes abgebildet.

Die Übergabe an den elektronischen Langzeitspeicher kann optional mit einem Zeitstempel für die zu speichernden Dokumente kombiniert werden. Darüber hinaus wird ein Zeitstempeldienst für die Signaturerneuerung elektronisch signierter Dokumente nach § 17 SigV benötigt. Das Projekt ArchiSafe setzt hierbei auf das als rechtssicher eingestufte und bewährte ArchiSig-Verfahren auf (http://www.archisig.de). Im Backend liegt schließlich das eigentliche Langzeitspeichersystem, in das grundsätzlich nur „Original"-Dokumente nebst den zugehörigen Vorgangsmetadaten abgelegt werden sollen. Parallel hierzu werden gegebenenfalls Kopien der Dokumente und Metadaten zum Zwecke der Vorgangsbildung und Vorgangsbearbeitung weiterhin im Fachverfahren vorgehalten. Über die Vergabe eindeutiger Dokumentkennungen (Dokument-ID, „Garderobenmarken") stellt das Langzeitspeichersystem sicher, dass zu jedem Zeitpunkt aus dem Fachverfahren heraus in wirtschaftlich vertretbaren Zeiträumen auf die abgelegten „Originale" zugegriffen werden kann. Dieses Vorgehen garantiert die rechtssichere Ablage von Originaldokumenten, ohne das Langzeitspeichersystem mit vorgangsspezifischen Logiken zu überfrachten. Für die Realisierung einer mandantenfähigen Lösung kann die Dokument-ID zusätzlich mit einer eindeutigen Kennung für das zuständige Fachverfahren verknüpft werden. Auf diese Weise lässt sich zudem über ein Berechtigungskonzept im Fachverfahren sicherstellen, dass unzulässige Zugriffe auf die archivierten Daten verhindert werden.

Um einen vom Fachverfahren unabhängigen Zugriff auf den Langzeitspeicher zu ermöglichen, kann – auf Anforderung – ergänzend ein Such- und Darstellungsdienst sinnvoll sein. Insbesondere dann, wenn eine mandantenfähige Lösung angestrebt wird. Über diesen Dienst, der die im Langzeitspeicher vorhandenen Metadaten datenbankgestützt redundant vorhält, kann im Fall des Ausfalls des führenden Systems eine Rekonstruktion der Vorgänge oder Akten erfolgen. Bei Bedarf können die Daten und Dokumente in einer weiterverwendbaren Form präsentiert (Viewer) oder exportiert werden. Die Implementierung eines solchen Dienstes darf jedoch auf keinen Fall die Möglichkeit einräumen, die gesetzlich vorgeschriebenen Datenschutzbestimmungen zu unterlaufen.

Dokumentformate und Datenstrukturen im Projekt ArchiSafe

Dokumentformate

Für die dauerhafte Speicherung von Dokumenten sollten gemäß den Empfehlungen des DOMEA-Organisationskonzeptes in der Version 2.0 nur einige wenige Formate zur Anwendung kommen. Das Nebeneinander unterschiedlichster Formate im Bereich der Langzeitspeicherung erhöht das Risiko, das einzelne Datentypen im Verlaufe der Aufbewahrungsfrist nicht mehr originalgetreu reproduziert werden können und somit die Authentizität der abgelegten Dokumente verloren geht.

ArchiSafe empfiehlt daher auf der Grundlage des DOMEA Organisationskonzeptes in Abhängigkeit vom Format der zu archivierenden Daten folgende Dokumentformate für die Langzeitspeicherung:

- TXT (ASCII 7-bit) für einfache Textinformationen, Metadaten und Stammdaten aus Fachsystemen
- PDF-Format (vorzugsweise PDF-A) für kodierte Dokumente (CI).

Dieses Dokumentenformat ist plattformuabhängig einsetzbar und erlaubt neben der grafischen Information auch die Textinformationen zu speichern, so dass auch nach der Konvertierung eine Volltextrecherche möglich bleibt. Darüber hinaus verfügt das PDF-Format über weitere nützliche Funktionalitäten wie die Einbettung elektronischer Signaturen, so dass PDF ausdrücklich auch von der KBSt für die Archivierung von in CI-Formaten bereitgestellten Textdokumenten empfohlen wird. Dies wird zusätzlich durch die Veröffentlichung der ISO 19005-1 Norm „Document management - Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1)" unterstrichen.

- TIFF- oder / und PDF-Format für Dokumente, die in einem NCI-Format vorliegen und
- XML als Auszeichnungssprache für zu archivierende Metadaten oder Datensätze.

Metadatenstrukturen und Schnittstellen

Die von der KBSt veröffentlichten „Standards und Architekturen für E-Government-Anwendungen (Vers. 2.0, Schriftenreihe der KBSt, Bd. 59, vom Dez. 2003) empfehlen Metadaten und Datenschnittstellen zu Drittsystemen grundsätzlich über XML und entsprechende Schemadefinitionen zu beschreiben und zu realisieren.

ArchiSafe setzt daher auch für die Kommunikation zwischen Fachverfahren und Archiv auf die Verwendung von XML als Beschreibungssprache für in sich abgeschlossene Archivobjekte, die sich über ein vereinbartes XML-Schema selbst beschreiben und so alle wichtigen und hinreichenden Informationen enthalten, die man für einen späteren Zugriff benötigt (s. Abb. 2). Die Beschreibung durch ein gültiges XML-Schema verspricht vor allem folgende Vorteile:

- Das Archivobjekt kann vor der Übergabe an den elektronischen Langzeitspeicher auf syntaktische Richtigkeit evaluiert werden.
- Fachverfahrensspezifische oder Behördenspezifische Erweiterungen der Metadaten sind mit wenig Aufwand durch Erweiterung und/oder Einschluss zusätzlicher XML-Schemata möglich.

Im einfachsten Fall besteht ein solches Archivobjekt neben einer Versionsangabe und der Angabe der zugeordneten XML-Schemadatei, aus einem Block, der die Inhaltsdaten enthält (Objektblock) und gegebenenfalls aus einem

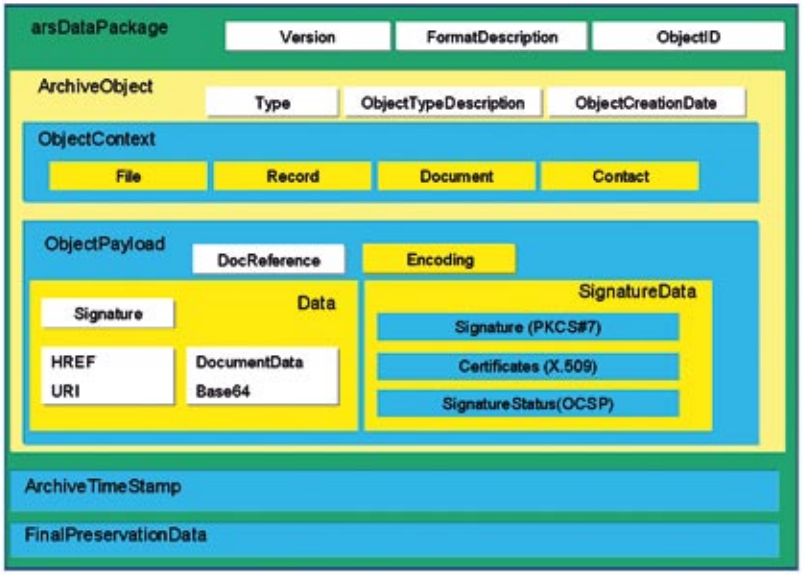

Abb. 2: ArchiSafe Metadatenschema (grafisch)

oder mehreren Signaturblöcken. Der Objektblock kann selbst wieder ein oder mehrere in XML eingebettete Dokumente enthalten. Jeder Block enthält als Einleitung Metadaten, in denen beispielsweise eine Dokumentkennung (Dokument-ID), eine Beschreibung des Dokumentes und seiner Herkunft abgelegt werden können. Optional wird im ArchiSafe-Container zudem ein Block für Beschreibungsdaten für die Übergabe an das Bundesarchiv vorgehalten

Für das Dokument selbst ist als Standard PDF-A vorgesehen, das um in XML eingebettet werden zu können, zunächst in ein Textformat (Base64) konvertiert wird. Für speicherintensive Binärdaten empfiehlt sich, nicht zuletzt auch aus Performancegründen bei häufigen Zugriffen auf den Langzeitspeicher, die Binärdaten als Anhang (attachment) im XML-Datenstrom zu referenzieren. In diesem Fall muss der Objektblock eine Referenz auf die dann zusätzlich archivierte Binär-Datei erhalten. Darüber hinaus können nach dem ArchiSafe-Konzept die eigentlichen Inhaltsdaten (Dokumentinhalte) auch in mehreren unterschiedlichen Dokumentformaten innerhalb oder außerhalb der XML-Datei abgelegt werden. Der Entwurf für ein solches ArchiSafe konformes XML-Schema ist unter dem Arbeitstitel ARS (für ArchiSafe Record-Keeping Strategy) 05.02 „ARS XML Datenpakete und Metadaten“ spezifiziert und wird noch in 2005 auf www.archisafe.de veröffentlicht.

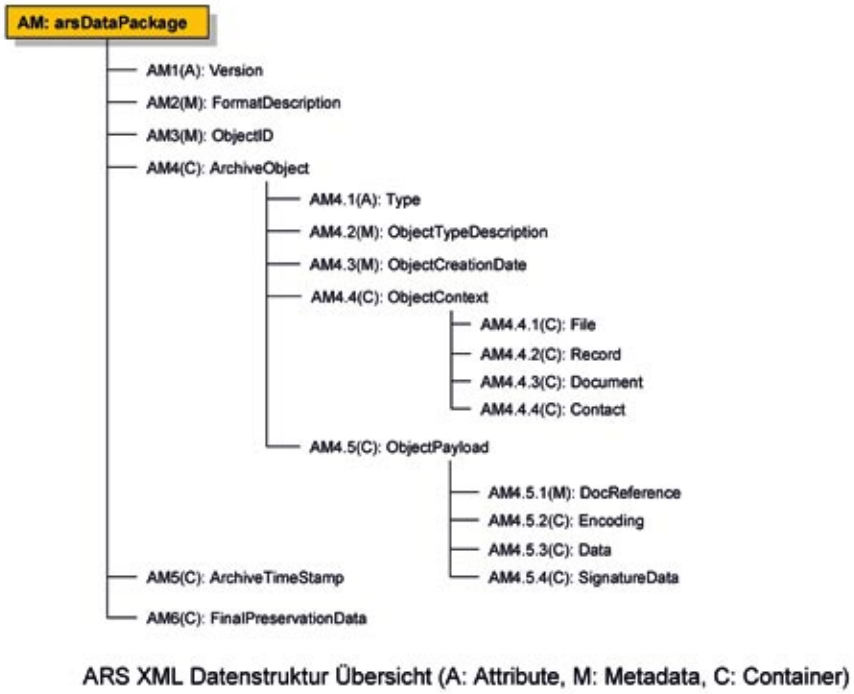

Abb. 3: ArchiSafe Metadatenbaum

Im Projekt ArchiSafe wird die Ablage eines Dokumentes inklusive der oben beschriebenen Metadaten angestrebt. An dieser Stelle sei darauf verwiesen, dass die im DOMEA-Konzept dargestellten Beziehungen zwischen einer Akte, Vorgängen und Dokumenten ausschließlich über die Metadaten eines „Archivobjektes“, sprich eines Dokumentes, abzubilden sind. Dies bedeutet, dass der Aufbau einer Akte oder eines Vorganges zunächst außerhalb der ArchiSafe-Lösung, also im Fachverfahren, zu realisieren und vorzuhalten ist. An Hand der Metadaten der „archivierten“ Objekte kann jederzeit dynamisch (z.B. über eine Suche) wieder eine Liste aller zu einem Vorgang oder einer Akte gehörenden Dokumente im Langzeitspeicher erstellt werden. Informationen über den Aufbau und den Werdegang einer Akte oder eines Vorganges können zudem über ein separates Dokument (z.B. auf Basis des XDOMEA-Standards) in den Langzeitspeicher eingestellt werden. Die angedachte „Single“-Dokument-/Objekt-Lösung ist somit auch für „Nicht“-DOMEA- konforme Systeme, d.h. beliebige Fachverfahren, nutzbar. Darüber hinaus kann jeder Nachnutzer eigene Akten- und Vor-

gangsstrukturen definieren und ggf. notwendige Metadaten dafür in ARS einpflegen. Hier verweist das Projekt ArchiSafe auf die in Niedersachsen beim IZN gesammelten Erfahrungen.

Das Projekt ArchiSafe wird in drei Stufen erarbeitet. In der ersten Stufe wurde ein Fachkonzept und ein DV-technisches Konzept entwickelt, auf dessen Grundlage in der zweiten Stufe ein funktional zwar noch eingeschränkter, aber voll funktionsfähiger Pilot in der PTB installiert wurde, der es erlaubt, Erfahrungen vor allem hinsichtlich der Realisierung und Einbettung in die elektronische Dokumenteninfrastruktur der PTB erfolgen und die ArchiSafe-Lösung auch anderen Behörden als Dienstleistung angeboten werden.

Aktueller Stand des Projektes Weiterführende Informationen

Für 2006 sind zudem u. a. folgende Weiterentwicklungen geplant:

- Anschluss der Virtuellen Poststelle des Bundes (VPS) an die ArchiSafe Middleware
  - Zur Einholung von Signaturen, Zertifikaten und Signaturprüfinformationen
  - Zur Einholung von Zeitstempeln
- Konzeption und Implementierung der Mandantenfähigkeit von ArchiSafe
- Konzeption und Implementierung eines sicheren Zugangskanals zum ArchiSafe-Service (manueller und automatisierter Aufruf)
- Fertigstellung der ArchiSafe Spezifikationen
- Prüfung der Möglichkeit der Zertifizierung der im Rahmen von ArchiSafe verwendeten Komponenten
- Erstellung einer Verfahrensdokumentation im Sinne der Grundsätze ordnungsmäßiger DV-gestützter Buchführungssysteme (GoBS)
- Erarbeitung eines Erweiterungsmoduls „ArchiSafe“ für das DOMEA-Organisationskonzept in Zusammenarbeit mit der KBSt
- Aufbau und Betriebsführung eines ArchiSafe-Services für externe Kunden

Die Konzepte, Spezifikationen, Schnittstellendefinitionen und Erfahrungsberichte sind zur Nachnutzung und Diskussion auf einem extra eingerichteten Online-Forum veröffentlicht (http://www.archisafe.de) und frei zugänglich. Die Einbeziehung von mehr als 20 Bundesbehörden, darunter das Bundesarchiv, die KBSt, das BSI und die Bundesnetzagentur unterstützen die Nachnutzungsfähigkeit im Sinne der von BundOnline 2005 geförderten „Einer-für-alle“ Dienstleistungen.

ArchiSafe Partner

Bundesamt für Sicherheit in der Informationstechnik

Informatikzentrum Niedersachsen – Landesbetrieb –

Das Bundesarchiv

CSC

Bundesnetzagentur

OpenLimit

Adobe

KBSt

BearingPoint

MICUS Management Consulting GmbH

OPEN TEXT CORPORATION

secunet

Diese Broschüre wurde ermöglicht durch die Unterstützung von
Micus Management Consulting GmbH, Düsseldorf, Berlin, www.micus.de