

DAK migriert Informationsarchiv mit LuraTech nach PDF/A

Die DAK, Deutschlands zweitgrößte Krankenkasse, hat ihr internes Informationsarchiv mit Hilfe von LuraTech auf den neuesten ISO-Standard PDF/A migriert. Zuvor bestand dies aus digitalisierten Mikrofilmen, die im TIFF-Format vorlagen. Das hatte nicht nur den Nachteil, viel Speicherplatz zu verbrauchen, sondern vor allem waren die Dateien nicht volltextfähig. Jetzt, nach der Konvertierung und Kompression von ursprünglich rund 300.000 Textseiten, haben die Nutzer des DAK-INFO-Dienstes besten Zugriff auf alle gespeicherten Informationen. Außerdem konnte die DAK in diesem Startprojekt frühzeitig Erfahrungen mit dem neuen PDF/A Format sammeln. Diese werden in weitere Projekte einfließen.



Die DAK als zweitgrößte Krankenkasse Deutschlands betreut unter dem Motto „Unternehmen Leben“ rund 4,7 Millionen Mitglieder und 6,2 Millionen Versicherte. Circa 12.000 Mitarbeiter arbeiten bundesweit bei dem Gesundheitsdienstleister. Mit Blick auf eine über 230-jährige Tradition und Know-how im Bereich der Sozialversicherung hilft die DAK nicht nur im Krankheitsfall. Das Unternehmen gilt als besonders

innovativ innerhalb der gesetzlichen Krankenversicherung: Mit modernen Angeboten wie Bonusprogrammen, Integrationsverträgen oder Gesundheitsreisen will die DAK vor allem gesundes Verhalten belohnen. Das bestätigt auch die aktuelle Sonderausgabe des Verbrauchermagazins ÖKO-Test, in der die DAK als beste Krankenkasse abschnitt.

Antworten auf alle Fragen: der DAK INFO-Dienst

Der in der Hamburger Zentrale ansässige interne INFO-Dienst der DAK ist die Informationsplattform für die Mitarbeiter in den Geschäftsstellen. Hier fließt der Input aus allen für die DAK relevanten Fachpublikationen ein, die dafür eingescannt werden. In der Regel sind dies juristisch oder medizinisch geprägte Texte wie Fachaufsätze zu rechtlichen Aspekten der Gesundheitsversorgung, zu Patientenrechten, über Kunstfehler, neue Behandlungsmethoden etc. Alle diese täglich eingehenden Informationen werden in die Literatur-Datenbank der DAK DOR!*Online* (DOKumentations Retrieval und Informations System) eingepflegt, wo sie

dann dem hausinternen INFO-Dienst zur Recherche zur Verfügung stehen. Die Mitarbeiter des INFO-Dienstes bearbeiten Anfragen aus den Geschäftsstellen zu bestimmten Problemstellungen rechtlicher oder medizinischer Natur und überstellen ihnen recherchierte Informationen aus dem Archiv. „Außerdem haben wir eine Informationspflicht unseren Mitgliedern gegenüber, was neue gesetzliche Regelungen oder sonstige, für sie relevante Entwicklungen betrifft“, erklärt Hans-Hinrich Harms, Dokumentar des INFO-Dienstes bei der DAK. „Auch diese Informationen beziehen wir aus unserer Datenbank.“ Nicht wenige der relevanten Texte – es sind rund 3000 pro Jahr – umfassen 50 Seiten oder mehr. Diese werden dann an den externen Scandienstleister Alphacom in Hamburg gegeben, der sie digitalisiert an die DAK zurücksendet. Kleinere Artikel werden mit Etagen-Scannern von Hewlett Packard im Haus erfasst und in das Archiv abgelegt.

Volltextfähigkeit gefragt

Bisher wurden die eingescannten Informationen im TIFF-Format archiviert. Außerdem befanden sich in dem DAK-Archiv vom Mikrofilm nach TIFF digitalisierte Dateien aus den Jahren 1988 bis 1995. Später wurde zwar nicht mehr im TIFF-Format archiviert, sondern in PDF, doch ist letzteres kein standardisiertes Format für die Langzeitarchivierung. Problematisch war, dass der große Bestand der TIFF-Dateien nicht volltextfähig und damit nur schwer recherchierbar war. Daher entschied man sich zu Beginn des Jahres, ein modernes Format und eine entsprechende Lösung für die Konvertierung des Archivbestands zu suchen. Auf der CeBIT 2006 informierte sich die DAK über das neue, ISO-zertifizierte PDF/A-Format für die Langzeitarchivierung. PDF/A hat im Vergleich zu anderen Formaten entscheidende Vorteile: Es ist im Gegensatz zu TIFF volltextfähig, gibt Farben original wieder, und die Dateien sind kleiner. Im Vergleich zu nativen Formaten, die Anwendungsprogramme wie Office oder CAD-Systeme erzeugen, ist es langfristig lesbar und benötigt nur einen Viewer – den Adobe Reader.

Auf dem CeBIT-Stand des Berliner Anbieters LuraTech wurde die DAK fündig. Schließlich hatte der Kompressions-Spezialist gerade sein Produktportfolio rundum für den Einsatz von PDF/A erweitert. Nicht nur gescannte, sondern auch digital erstellte Dokumente lassen sich mit den Lösungen von LuraTech in PDF/A umwandeln und auf Konformität überprüfen. Letzteres ist notwendig, um sicherzustellen, dass ein zu archivierendes Dokument alle Anforderungen des Standards erfüllt. Denn die Syntax von PDF/A, einer eingegrenzten Variante von PDF 1.4, schreibt detailliert vor, welche Inhalte erlaubt sind und welche nicht. So muss beispielsweise der gesamte Inhalt in der Datei enthalten sein. Fremabhängigkeiten oder Referenzen zu externen Quellen, welche die langfristige Lesbarkeit eventuell beeinträchtigen, sind ausgeschlossen. Die genutzten Schriften müssen komplett eingebunden sein. Eine Verschlüsselung ist nicht zulässig.

Schlechte Lesbarkeit kein Problem für den PDF Compressor

Mit dem LuraTech PDF Compressor in der neuen Version 4, der gescannte Dokumente in hoch komprimierte und im Volltext (OCR) durchsuchbare PDF/A-Dokumente verwandelt, migrierte LuraTech innerhalb weniger Wochen die Dokumente aus dem alten Archiv in PDF/A. Die auf Mikrofilm gebannten Daten waren jedoch teilweise sehr schlecht lesbar gewesen und hatten zu entsprechend wenig qualitätvollen TIFFs geführt. Durch die ausgereifte OCR-Technologie, mit der der PDF-Compressor arbeitet, konnte trotzdem eine sehr hohe Erkennungsrate erzielt werden. Um eine unkomplizierte Zu-

ordnung der gewandelten Dateien zu ermöglichen, übernahm LuraTech die alten Dateibezeichnungen. Diese bestanden aus der Jahreszahl des Entstehungszeitpunkts und einer fünfstelligen laufenden Nummer. Lediglich die Extension TIFF veränderte sich.

Zufriedenheit auf allen Seiten

Die Mitarbeiter des DAK INFO-Dienstes profitieren jetzt von der komfortablen und schnellen Volltextrecherche. Geplant ist außerdem, die Informationen aus der DORIS Datenbank langfristig über ein Portal auch direkt für sämtliche DAK-Mitarbeiter abrufbar zu machen. Dafür ist jetzt ein erster Schritt getan. Gleichzeitig stellte das Projekt eine Art Test für das noch taufische PDF/A Format dar, an dessen Möglichkeiten die DAK sehr interessiert ist.

Hans-Hinrich Harms fasst zusammen: „Mit PDF/A bot sich ein modernes Format als Nachfolger für unser TIFF-Archiv an, das wesentlich mehr Funktionen bietet und außerdem noch ein ISO-Standard ist. Für unser Info-System war insbesondere die Volltextfähigkeit von PDF/A der Hauptvorteil. Die Zusammenarbeit und Kommunikation mit der LuraTech gestaltete sich auch sehr angenehm. Darüber hinaus haben wir im Verlauf des Projekts gesehen, dass Farb-Scanning mit PDF/A heute möglich ist und hier prüfen wir die Möglichkeiten, in Zukunft auch Farb-Dokumente in unserem Info-System anzubieten.“

PDF/A ist der ISO Standard 19005 für die Langzeitarchivierung im PDF-Format. Es stellt eine eingegrenzte Variante von PDF dar, ein standardisiertes Profil zur Verwendung von PDF in der Langzeitarchivierung. Der Standard schreibt detailliert vor, welche Inhalte erlaubt sind und welche nicht. Durch diese und andere Vorschriften soll eine langfristige Lesbarkeit der Dokumente garantiert sein – und zwar unabhängig davon, mit welcher Anwendungssoftware und auf welchem Betriebssystem sie ursprünglich erstellt wurden. Anerkannte Experten gehen davon aus, dass durch die ISO-Standardisierung das PDF/A-Format sehr schnell an Bedeutung gewinnt und das TIF-Format allein schon aufgrund der Volltextrecherchemöglichkeit kurzfristig ablöst.

LuraTech Europe GmbH

Kantstr. 21

10623 Berlin

Telefon: +49 30 394050-0

Telefax: +49 30 394050-99

info@luratech.com

<http://www.luratech.com>