

PDF/A-Format für Langzeitarchivierung, Teil 1

Eine Erfolgsstory par excellence

Wie sicher können Unternehmen und Verwaltungen sein, dass ihre digital gespeicherten Dokumente noch über Jahrzehnte hinweg lesbar sind? Die Antwort auf diese immens wichtige Frage lautet: PDF/A. Die noch relative junge Norm basiert auf einem „abgespeckten“ PDF-Format und hat sich innerhalb von knapp drei Jahren zu einem internationalen Standard für die Langzeitarchivierung von Dokumenten entwickelt. Maßgeblichen Anteil an dieser Erfolgsstory hat der eigens zur Förderung von PDF/A gegründete Verband. BIT sprach mit dem Vorstand des PDF/A Competence Center über die Bedeutung des PDF/A-Standards für die Unternehmen und die Leistungen des Verbandes. Der 2. Teil des Interviews folgt in BIT 3-2009.



Herr Grumser, erklären Sie doch bitte kurz, warum PDF/A sich vor allen anderen Formaten für die Langzeitarchivierung besonders gut eignet.

Harald Grumser: Ich könnte es mir jetzt einfach machen und sagen: Es ist das einzige Format, das zu diesem Zweck als ISO 19005-1:2005 die Weihen einer Normierung erfahren hat. PDF/A, das ja vollständig auf PDF basiert, vereint Vorteile, die kein anderes Dokumentenformat bieten kann. Es ist mächtiger als fast alle anderen Formate, es eignet sich sowohl zum Anzeigen als auch zum Drucken, es bietet mit seinen einzigartigen Metadaten eine hervorragende Grundlage für Archive, und es ist verbreitet wie kein anderes Format.

Dr. Bernd Wild: „Bei öffentlichen Ausschreibungen gehört PDF/A heute schon zu den Grundanforderungen.“



Ein Standard ist erst dann ein Standard, wenn er sich in der Praxis eine breite Akzeptanz verschafft hat. Hat die PDF/A-Norm bereits die kritische Masse erreicht, bei der man von breiter Akzeptanz sprechen kann?

Thomas Zellmann: Wir sind natürlich der Meinung, dass PDF/A sich nach drei Jahren schon hinreichend gut durchgesetzt hat. Bei genauer Betrachtung kann man aber feststellen, dass die Durchsetzung noch regional und in den Branchen unterschiedlich ist. Im deutschsprachigen Raum ist PDF/A sicher schon breit akzeptiert, und wir sehen das seit 2008 auch in Europa. In Asien oder USA ist PDF/A – Stand heute – noch nicht so verbreitet. Auch in den Branchen ver-

läuft die Akzeptanz unterschiedlich schnell. Wir freuen uns über zunehmende Empfehlungen und Vorgaben im Behördenbereich und auch bei Archiven und Bibliotheken wie z. B. der Deutschen und Österreichischen Nationalbibliothek.



Welche Rolle spielt PDF/A heute schon bei Ausschreibungen? In welchen Branchen und Unternehmen ist PDF/A bisher am weitesten verbreitet und warum sind dies die Vorreiter?

Dr. Bernd Wild: Bei öffentlichen Ausschreibungen gehört PDF/A heute schon zu den Grundanforderungen. Das bedeutet zwar nicht, dass bei der Umsetzung nur noch dieses Format verwendet wird, aber der mittelfristige Übergang auf PDF/A als Archivformat ist gesetzt. Für die Industrie spielt PDF/A insbesondere bei echten Langzeitarchivierungen – z. B. im Anlagen- und Flugzeugbau – eine Rolle. Auch im Finanzbereich nimmt insbesondere durch die Öffnung der zentralen Dienstleisterarchive für dieses Format die Verbreitung von PDF/A zu. Hier wird PDF/A in einigen Jahren TIFF abgelöst haben. Lediglich im Healthcare-Markt, d. h. bei der Archivierung von Patienten- und Arztunterlagen, ist noch eine abwartende Haltung erkennbar, da häufig gerade die ersten Schritte in Richtung Digitalisierung unternommen werden und deshalb auf „etablierte“ Technik Wert gelegt wird.



Wie ist die Akzeptanz gegenüber dem Standard in der Industrie – in der Dokumentenarchivierung im Umfeld der industriellen Fertigung?

Johannes Hesel: In der Industrie, insbesondere in Branchen mit hohen Sicherheitsstandards und Compliance-Anforderungen, wird

PDF/A zunehmend eingesetzt. In der Luftfahrt- oder Pharmaindustrie lassen sich die Anforderungen von FAA oder FDA an die elektronische Langzeitarchivierung von Dokumenten mit PDF/A erfüllen. Da technische Unterlagen häufig 30 Jahre und länger archiviert werden müssen, brauchen Industrieunternehmen Datenformate, die einen gesicherten Zugriff in diesem Zeitraum garantieren. PDF/A erfüllt diese Anforderung und bietet als ISO-Norm den Vorteil, dass es auch von vielen Kontrollinstanzen und Behörden als Datenformat für die elektronische Archivierung akzeptiert ist. Vor diesem Hintergrund ist es nicht verwunderlich, dass immer mehr Industrieunternehmen auf PDF/A als Datenformat für die Dokumentenarchivierung setzen.

? Ein wesentlicher Vorteil von PDF/A ist die Konsolidierung auf ein Format. Was bedeutet das in der Praxis?

Thomas Zellmann: Der Vorteil der Formatvereinheitlichung wird unseres Erachtens fast etwas stiefmütterlich behandelt, ist aber in der Praxis fast genauso wichtig wie ein

sicheres Langzeitformat. Im Laufe der Zeit ist bei vielen Anwendern ein regelrechter „Format-Zoo“ in den Archiven entstanden. So habe ich kürzlich beispielsweise von einem Fall beim Bundesarchiv gehört, wo eine Behörden-Anwendung zur Übergabe an das Bundesarchiv geprüft wurde. In diesem Archiv lagen mehr als 30 bekannte und obendrein auch noch unbekannte Formate vor, von denen keiner mehr genau wusste, wofür die eigentlich mal vorgesehen waren.

In kommerziellen Firmen ist die alte Papierakte ja meist schon durch digitale Akten abgelöst worden, aber jetzt liegen Dokumente einer logischen Kundenakte meist verteilt in mehreren Systemen. Der Sachbearbeiter muss dann bei vielen Formaten mit vielen Viewern arbeiten, wo er im Extremfall ohne Schulung nicht mal wissen kann, wie man in den einzelnen Viewern blättert oder einfach nur druckt.

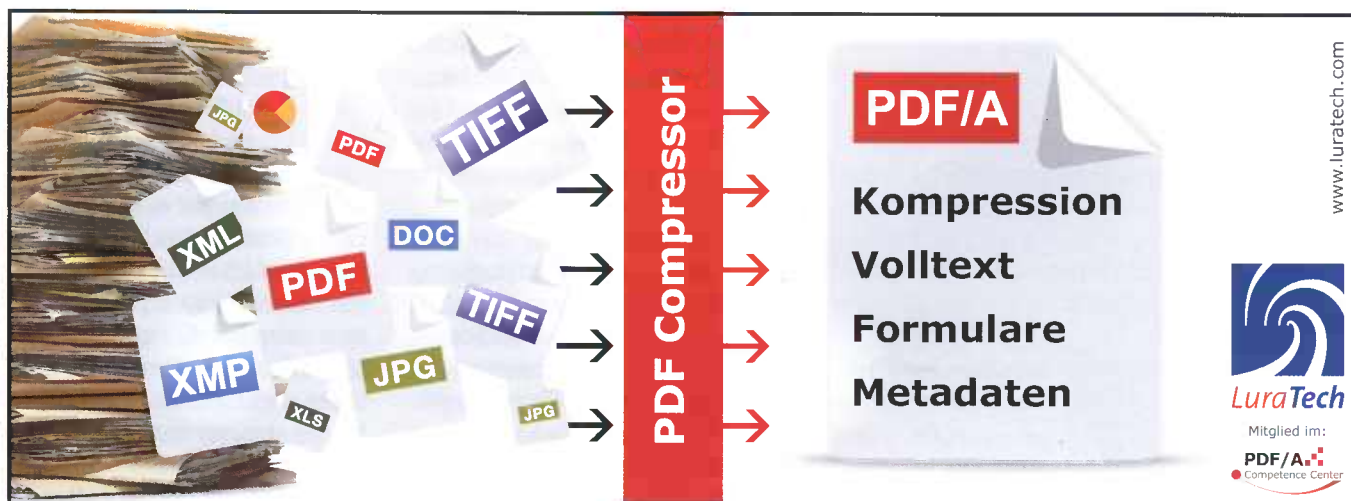
? Bevor PDF/A-Dokumente in das Archiv überführt werden, müssen sie validiert, also überprüft, werden. Wie erfolgt die Validierung?

Olaf Drümmer: Seit der Veröffentlichung des PDF/A-Standards haben mehrere Hersteller Werkzeuge zur PDF/A-Validierung herausgebracht – das bekannteste dürfte sicher die von Callas entwickelte Preflight-Funktion in Adobe Acrobat Pro sein. Allen professionellen PDF/A-Validierern gemeinsam ist, dass sie ein PDF detailliert analysieren, eventuelle Verletzungen der PDF/A-Norm-Vorschriften identifizieren und eine entsprechende Rückmeldung an den Benutzer bereitstellen.

? Wie Sie sagen, gibt es zur Validierung eine Vielzahl von Produkten. Wer garantiert, dass diese Software-Produkte korrekt arbeiten?

Thomas Merz: Das PDF/A Competence Center hat großen Aufwand in die Erstellung einer Test-Suite für PDF/A-Validierungs-Software gesteckt. Diese Isartor-Test-Suite steht auf www.pdfa.org zum freien Download zur Verfügung und dient zur Überprüfung von PDF/A-Validierern. Da die Isartor-Test-Suite alle denkbaren Verletzungen des PDF/A-Standards provoziert, kön-

Thomas Zellmann: „Im Laufe der Zeit ist bei vielen Anwendern ein regelrechter ‚Format-Zoo‘ in den Archiven entstanden.“





Harald Grumser: „Die Forderung von PDF/A, alle Ressourcen und damit auch alle Fonts einbetten zu müssen, ist eine wesentliche Voraussetzung, die Dokumente auch in zehn oder 50 Jahren noch anzeigen zu können.“



Thomas Zellmann: „Der Sachbearbeiter muss bei vielen Formaten mit vielen Viewern arbeiten, wo er im Extremfall ohne Schulung nicht mal wissen kann, wie man in den einzelnen Viewern blättert oder einfach nur druckt.“



Olaf Drümmer: „Allen professionellen PDF/A-Validierern gemeinsam ist, dass sie ein PDF analysieren, Verletzungen der PDF/A-Norm-Vorschriften identifizieren und eine Rückmeldung an den Benutzer bereitstellen.“

Dr. Bernd Wild: „Durch die Schema-Erweiterungen können nahezu beliebige Metadaten in ein PDF/A-Dokument integriert werden. Dadurch müssen aufwendigere Metadaten nicht mehr außerhalb des Dokuments produktspezifisch in Archiv- oder DMS-Datenbanken abgelegt werden, sondern bilden mit dem Dokument eine Einheit.“

nen Integratoren und Anwender damit überprüfen, ob diese Standardverletzungen von einem Validierer auch erkannt werden.



Das PDF/A-Format ist in zwei Varianten verfügbar, PDF/A-1a und -1b. Was ist der Unterschied?

Hans Bärffuss: PDF/A-1a (Level A Conformance) bezeichnet die vollständige Übereinstimmung mit dem PDF/A-Standard. Konkret muss ein PDF/A-1a-konformes Dokument maschinenlesbaren Text, Angaben zur Leserichtung usw. enthalten. Damit ist es für behinderte Menschen zugänglich und erfüllt die Anforderungen gemäß Section 508 in den USA. PDF/A-1b ist eine weniger strenge Auslegung des Standards und definiert die Minimalanforderungen an PDF/A-konforme Dokumente. In der Praxis ist PDF/A-1b in den meisten Fällen genügend.



Seit seiner Veröffentlichung vor drei Jahren hat sich PDF/A weiterentwickelt. Was sind die wesentlichen Neuerungen, z. B. bei der elektronischen Signatur und den Metadaten?

Dr. Bernd Wild: Echte Neuerungen gab es keine. Das würde auch dem Charakter als stabiler Standard widersprechen. Vielmehr erkennt man im Markt erst langsam das Potenzial und die Möglichkeiten, die bereits im PDF/A-1-Standard liegen. Hierzu leistet ja gerade das PDF/A Competence Center die notwendige Aufklärungsarbeit. Das Thema „elektronische Signaturen in PDF/A“ gehört dazu, da die Zulässigkeit von Signaturen in PDF/A-Dokumenten anfänglich nicht für jeden Kunden ersichtlich war. Gerade in Verbindung mit der Möglichkeit, Dokumente standardkonform qualifiziert zu signieren oder mit Zeitstempeln zu versehen, und dies noch eingebettet im Dokument selbst, eröffnet PDF/A weite Anwendungsfelder.

Daneben sind die durch den ISO-Standard definierten Möglichkeiten der Ablage von Metadaten im Dokument noch bei weitem nicht ausgeschöpft. Hier bieten sich für alle Anwender, die komplexe und recherchierbare Dokumente archivieren müssen, eine Fülle von

neuen Möglichkeiten, die es zu erschließen gilt. Durch die Schema-Erweiterungen können nahezu beliebige Metadaten in ein PDF/A-Dokument integriert werden. Dadurch müssen aufwendigere Metadaten nicht mehr außerhalb des Dokuments produktspezifisch in Archiv- oder DMS-Datenbanken abgelegt werden, sondern bilden mit dem Dokument eine Einheit.



PDF/A-Dokumente enthalten Metadaten zur Definition der Inhalte des Dokuments. Welche Vorteile bieten firmen- oder branchenspezifische Metadaten in PDF/A-Dokumenten außer den genannten noch und um welches Metadatenformat handelt es sich?

Olaf Drümmer: Das Wiederauffinden und inhaltliche Zuordnen von Dokumenten – insbesondere nach vielen Jahren – wird naturgemäß durch begleitende Informationen erleichtert, die eine Identifikation gestatten oder wichtige Aspekte des Dokuments beschreiben. PDF/A gibt hier das XML-



Thomas Merz: „Da die Isartor-Test-Suite alle denkbaren Verletzungen des PDF/A-Standards provoziert, können Integratoren und Anwender damit überprüfen, ob diese Standardverletzungen von einem Validierer auch erkannt werden.“



Hans Bärffuss: „PDF/A-1b ist eine weniger strenge Auslegung des Standards und definiert die Minimalanforderungen an PDF/A-konforme Dokumente. In der Praxis ist PDF/A-1b in den meisten Fällen genügend.“



Dr. Bernd Wild: „In Verbindung mit der Möglichkeit, Dokumente standardkonform qualifiziert zu signieren oder mit Zeitstempeln zu versehen, und dies noch eingebettet im Dokument selbst, eröffnet PDF/A weite Anwendungsfelder.“

Harald Grumser: „Wenn man es mit unterschiedlichen Schriften nicht übertreibt, ist PDF/A immer noch kleiner als selbst eine entsprechende TIFF-Schwarzweiß-Datei.“



Johannes Hesel: „In der Industrie, insbesondere in Branchen mit hohen Sicherheitsstandards und Compliance-Anforderungen, wird PDF/A zunehmend eingesetzt.“

basierte XMP-Metadatenformat vor, das als integraler Bestandteil in PDF und damit auch PDF/A eingebettet werden kann. Durch die breite Unterstützung – neben Adobe selbst etwa auch andere große Unternehmen wie Canon, Nokia, Microsoft, Apple oder Sony – kristallisiert sich XMP inzwischen als Konvergenzformat für Metadaten heraus.



Ein wesentliches Merkmal von PDF/A ist die Einbettung der Schriften. Diese benötigt jedoch Speicherkapazität. Wie stellt sich dieses Problem z. B. bei Massenrechnungen dar, die sowohl die Rechnungsinformationen wie auch die statischen Informationen des Absenders beinhalten?

Harald Grumser: Die Forderung von PDF/A, alle Ressourcen und damit auch alle Fonts einbetten zu müssen, ist eine wesentliche Voraussetzung, die Dokumente auch in zehn oder 50 Jahren noch anzeigen zu können. Wenn Massenrechnungen einzeln gespeichert werden, bedeutet das in der Tat eine gewisse Redundanz und höheren Speicherbedarf gegenüber herkömmlichem PDF. Wir haben Messungen gemacht und vernünftig erstellte PDF/A-Dateien mit TIFFs verglichen. Wenn man es mit unterschiedlichen Schriften nicht übertreibt, ist PDF/A immer noch kleiner als selbst eine entsprechende TIFF-Schwarzweiß-Datei. Bei Farbe ist der Unterschied zugunsten von

**Frustrierte Kunden.
Verpasste Gelegenheiten.
Verschwendete Ressourcen.**

Mit der Customer Communication Management Lösung von Group 1 Software decken Sie den gesamten Prozess der bedarfsorientierten Kundenansprache ab – von der Datenaufbereitung, über die Dokumentenformatierung bis hin zum Outputmanagement.

Mehr unter: www.group1.de

Harald Grumser:
„Die PDF/A-Verbreitung ist ein sich selbst verstärkender Prozess, das sehen wir bei jeder Konferenz aufs Neue.“

PDF/A dann noch gewaltiger. Daneben kann man mit intelligenten Lösungen den Effekt reduzieren und z. B. mehrere Rechnungen in einer PDF/A-Datei zusammenfassen und erst beim Abfragen wieder auftrennen.



Mit welchen Verfahren lassen sich die Speicherressourcen schonen?

Hans Bärffuss: Aus der Projekterfahrung heraus haben sich bereits Lösungen abgezeichnet, die jedoch eine Erweiterung des PDF/A-Standards voraussetzen. Das PDF/A Competence Center setzt sich aktiv bei der ISO dafür ein, dass spätere Teile des Standards diese Lösungen beinhalten.



Wie können sich Interessenten über PDF/A informieren und auf dem Laufenden halten, welche Angebote gibt es dazu vom PDF/A Competence Center?

Thomas Zellmann: Der Verband bietet viele Möglichkeiten, um sich über PDF/A zu informieren. Neu in diesem Jahr sind die Webinare, in denen Mitglieder Fachinformationen zu speziellen Themen rund um PDF/A per Internet anbieten. Auf unserer Web-Seite www.pdfa.org, die sich als zentraler Anlaufpunkt für Informationen versteht, kann man sich gerne für den regelmäßig erscheinenden Newsletter registrieren und sich in die Mailing-Listen eintragen. Nicht zuletzt kann man sich natürlich jederzeit direkt und persönlich an das PDF/A Competence Center wenden, wo der Verband Fragen entweder direkt beantwortet oder Kontakte zu den passenden Mitgliedern vermittelt.



Höhepunkt der Aktivitäten des PDF/A Competence Centers wird die internationale Kon-

ferenz im Juni in Berlin sein. Welche Auswirkungen versprechen Sie sich von der Konferenz?

Harald Grumser: Wir erreichen mit einer solchen Konferenz, die wir jetzt zum dritten Mal durchführen, bedeutend mehr als die 200 bis 300 Teilnehmer. Wir haben z. B. letztes Jahr in Amsterdam viele neue Mitglieder gewonnen, die wir als die eigentlichen Multiplikatoren für die Verbreitung von PDF/A sehen. Mit weit über 30 Fachvorträgen, darunter einigen Anwenderberichten, klären wir ja

nicht nur auf, sondern machen auch Mut zum Umstieg. Die PDF/A-Verbreitung ist ein selbst verstärkender Prozess, das sehen wir bei jeder Konferenz aufs Neue.

Vielen Dank für das Gespräch, meine Herren.

Der 2. Teil des Interviews folgt in BIT 3-2009, die am 30. März erscheint. Darin geht es um die weitere Entwicklung des PDF/A-Standards, die Bedeutung von PDF/A im SAP-Umfeld und die Unterstützung des Standards durch Microsoft.

Compart Systemhaus

Umwandlung in AG

Das beständige starke Wachstum des Output-Spezialisten Compart machte eine Umstrukturierung der Vertriebs-, Service- und Entwicklungsaktivitäten sowie der eigenständigen Ländergesellschaften und Integrationspartnerschaften erforderlich. Der bereits vor einem Jahr begonnene Prozess wurde im Dezember 2008 mit der Wandlung der Compart Systemhaus GmbH in eine AG abgeschlossen. Die Compart AG ist die Muttergesellschaft, die sich auf Produktentwicklung und Support konzentriert. Die zu 100 Prozent der Muttergesellschaft gehörenden Töchter Compart Deutschland GmbH, Compart North America Inc. und Compart Iberia S. L. sind Vertriebs- und Service-Gesellschaften in den zugehörigen Ländern. Ziel des Rechtsformwechsels ist eine höhere Flexibilität des Unternehmens sowie eine klarere Gliederung und Aufgabenteilung der regionalen Töchter.

Für den Aufsichtsrat konnten Reinhard Hamburger (Vorsitzender), ehemaliger Geschäftsführer der Agilent Technologies Deutsch-

land GmbH, der Unternehmensberater Hans-Jürgen Berger und der Rechtsanwalt Dr. Rene-Alexander Hirth, Spezialist für internationales Recht, gewonnen werden.

„Mit der AG-Umwandlung ändert sich nichts an den Eigentumsverhältnissen der Gesellschaft“, ließ das Unternehmen verlauten. Die Gesellschaft bleibe ein Unternehmen, dessen Anteile vollständig von dem Vorstandsvorsitzenden Harald Grumser (CEO) und dem für die Entwicklung verantwortlichen Vorstand Christof Mayer (CTO) gehalten werden. Mit Peter Hopf als Chief Financial Officer und Jörg Palmer als Chief Services Officer wurden zwei langjährige Führungskräfte ebenfalls in den Vorstand berufen.

„Die AG wurde nicht zu Kapitalbeschaffungszwecken gegründet“, betont Harald Grumser. Compart finanziere sein starkes Wachstum als profitables Unternehmen aus Eigenmitteln und habe in den letzten Jahren gleichzeitig die Eigenkapitalquote wesentlich verbessert.

(www.compart.net)

Das PDF/A Competence Center informiert

PDF/A für Eingangsdokumente

Täglich gehen in Unternehmen Dokumente per Post und Telefax oder als E-Mail mit Attachments ein. Und immer mehr Unternehmen archivieren den Posteingang elektronisch. Um zu vermeiden, dass diese in unterschiedlichen Formaten abgelegt werden, empfiehlt sich die Archivierung in PDF/A. Dazu werden die gescannten Dokumente mittels Konverter ins PDF/A-Format umgewandelt. Der Markt bietet aufgrund der großen Nachfrage entsprechende Werkzeuge an, die eine effiziente Farb-

kompression ermöglichen und volltextfähige PDF/A-Dateien erzeugen können – aus jeglichem Dokument, das auch der Scanner verarbeiten kann. Die Konverter gibt es für jede Unternehmensgröße und entsprechendes Dokumentaufkommen: Von der gelegentlichen manuellen Konvertierung von gescannten Dokumenten bis hin zu serverbasierten Tools, die automatisch Eingangsverzeichnisse überwachen, dort befindliche Dateien komprimieren und in definierte Ausgangsverzeichnisse ablegen.

Neue Mitglieder

MPI Tech, Dänemark, und die Pegasus Imaging Corporation, USA, sind neue Mitglieder des PDF/A Competence Center.

MPI bietet Lösungen zum Drucken und Konvertieren die in Mainframe-, AS/400-, Unix- aber auch Windows-

Umgebungen zum Einsatz kommen.

Pegasus Imaging bietet anspruchsvolle Imaging-Technologie, leistungsfähige Komponenten für digitales Imaging und optimierte Bildkomprimierungs-SDKs für die Anwendungsentwicklung.

Workshops während der PDF/A Konferenz

Das PDF/A Competence Center veranstaltet seine dritte internationale PDF/A-Konferenz vom 18. bis 19. Juni in Berlin. Die Konferenz bietet Fachvorträge internationaler Experten, Tracks für unterschiedliche PDF/A-Schwerpunkte und Interessensgebiete sowie Workshops zum praktischen Einsatz von PDF/A.

Bereits am Vortag der Konferenz, dem 17. Juni, finden die Workshops statt. The-

men, wie gescannte Dokumente bzw. elektronische gewandelt werden können, die Verwaltung von Schriften und XMP-Metadaten, der Einsatz von elektronischen Signaturen oder Aspekte, die beim Output-Management zu beachten sind, stehen auf der Agenda. Darüber hinaus findet ein Workshop speziell für Sparkassen statt. Die kleinen Gruppen ermöglichen detaillierte Diskussionen sowie einen tiefgreifenden

PDF/A Competence Center auf der CeBIT Neueste Trends

Das PDF/A Competence Center präsentiert neueste Trends und Entwicklungen für die Langzeitarchivierung auf der CeBIT 2009 auf einem Partnerstand beim VOI (Verband Organisations- und Informationssysteme) in der Halle 3. Als Ansprechpartner stehen der neue Geschäftsführer Rainer Lemke sowie Vorstandsmitglieder und Experten des PDF/A Competence Center zur Verfügung.

Olaf Drümmer, Vorstandsmitglied im PDF/A Competence Center: „Unser Messestand steht Interessenten während der CeBIT als eine zentrale Anlaufstelle zur Verfügung, um sich umfassend über das PDF/A-Format zu informieren.“



Folgende Mitglieder des PDF/A Competence Center präsentieren auf der CeBIT ihre Lösungen:

- Compart AG, Halle 3, Stand A29,
- Luratech Europe GmbH, Halle 3, Stand A29,
- Zöller & Partner GmbH, Halle 3, Stand A29,
- Satz-Rechen-Zentrum, Halle 3, Stand B45,
- Janich & Klass, Halle 3, Stand B55,
- Optimal Systems GmbH, Halle 3, Stand C 55,
- Adobe Systems, Halle 9, Stand D22 u. Pavillon P33,
- Verlag für Landesamtswesen, Halle 9, Stand C42/1,
- Openlimit Signcubes, Halle 9, Stand C60,
- Secrypt GmbH, Halle 11, Stand B11.

Erfahrungs- und Wissensaustausch.

Die halbtägigen Workshops kosten 250 Euro. Die Konferenz-Teilnahme kostet

980 Euro. Bis zum 31. März gilt der Early-Bird-Preis in Höhe von 870 Euro. Weitere Informationen und Anmeldung unter www.pdfa.org.

Neue Online-Communities im Web

Das PDF/A Competence Center hat im Internet unter Xing.de (<https://www.xing.com/net/pdfacc/>) und Linkedin.com jeweils eine Com-

munity-Plattform zu PDF/A eingerichtet.

Interessenten sind herzlich eingeladen, sich an den Foren zu beteiligen.