



(10) **DE 10 2008 012 329 B4** 2011.07.14

(12) **Patentschrift**

(21) Aktenzeichen: **10 2008 012 329.3**  
 (22) Anmeldetag: **03.03.2008**  
 (43) Offenlegungstag: **17.09.2009**  
 (45) Veröffentlichungstag  
 der Patenterteilung: **14.07.2011**

(51) Int Cl.: **G06F 3/12 (2006.01)**

Innerhalb von drei Monaten nach Veröffentlichung der Patenterteilung kann nach § 59 Patentgesetz gegen das Patent Einspruch erhoben werden. Der Einspruch ist schriftlich zu erklären und zu begründen. Innerhalb der Einspruchsfrist ist eine Einspruchsgebühr in Höhe von 200 Euro zu entrichten (§ 6 Patentkostengesetz in Verbindung mit der Anlage zu § 2 Abs. 1 Patentkostengesetz).

(73) Patentinhaber:  
**Océ Printing Systems GmbH, 85586, Poing, DE**

(56) Für die Beurteilung der Patentfähigkeit in Betracht  
 gezogene Druckschriften:

(74) Vertreter:  
**Schaumburg, Thoenes, Thurn, Landskron, Eckert,  
 81679, München, DE**

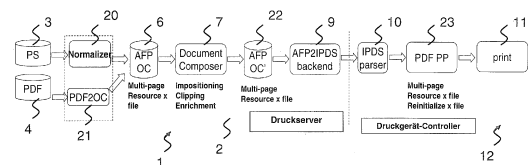
**EP 15 38 534 A2**  
**EP 09 86 003 A2**

(72) Erfinder:  
**Eiler, Göran, 85586, Poing, DE; Wallner, Robert,  
 82061, Neuried, DE; Weiner, Helmut, 81371,  
 München, DE; Matzke, Frank, 82319, Starnberg,  
 DE; Groeger, Hans-Detlef, 85586, Poing, DE; La  
 Rosa Ducato, José, 85435, Erding, DE**

(54) Bezeichnung: **Verfahren, Drucksystem und Computerprogramm zum Erzeugen und Verarbeiten von Dokumentendatenströmen**

(57) Hauptanspruch: Verfahren zum Erzeugen eines seitenweise seriell strukturierten Dokumentendatenstroms eines ersten Formats,

- wobei in dem ersten Format die erste Seite eines mehrseitigen Dokuments vollständig und erfolgreich einen Parsing-Prozess durchlaufen kann, bevor Dokumentendaten der nachfolgenden Seiten den Parsing-Prozess durchlaufen,
- wobei eine Datei mit Dokumentendaten eines zweiten Formats mit dem Dokumentendatenstrom verbunden wird, wobei die Dokumentendaten des zweiten Formats mehrere Seiten eines Dokuments umfassen und wobei für die Seiten gemeinsame Einstellungen und/oder Ressourcen gelten,
- wobei in dem Dokumentendatenstrom des ersten Formats ein gesondertes Typen-Kennfeld vorgesehen wird, das anzeigt, dass die Datei mit den zusammenhängenden Dokumentendaten des zweiten Formats mit dem Dokumentendatenstrom verbunden ist,
- und wobei die Verbindung durch eine Einbettung der Datei mit den Dokumentendaten des zweiten Formats in den Dokumentendatenstrom des ersten Formats erfolgt oder die Datei mit den Dokumentendaten des zweiten Formats separat von den Daten des Dokumentendatenstroms des ersten Formats gespeichert wird...



## Beschreibung

**[0001]** Die vorliegende Erfindung betrifft ein Verfahren, ein Drucksystem und Computerprogramme zum Erzeugen und Verarbeiten von Dokumentendatenströmen. Bei der Erzeugung und Verarbeitung von Dokumentendaten haben sich je nach Anwendungsgebiet verschiedene Datenformate herausgebildet, in denen die Dokumente erstellt werden und zur Weiterverarbeitung auf verschiedenen Ausgabegeräten, wie zum Beispiel Druckgeräten, bereit gestellt werden. Beispielsweise haben sich in der grafischen Industrie die Datenformate Postscript (PS) und Portable Dokument Format (PDF) etabliert, die darauf optimiert sind, komplexe grafische Strukturen zu kodieren. Das Format PDF hat zudem Einzug gefunden in weite Bereiche der Büroanwendungen und ist zwischenzeitlich als ISO-Norm 15930 der International Organization for Standardization angenommen worden. In Büro-Umgebungen ist das Datenformat Page Description Language (PCL) etabliert.

**[0002]** Im Bereich der Hochgeschwindigkeits-Ausgabe digitaler Daten aus Großrechenanlagen, zum Beispiel zum Ausdruck von Rechnungen aus Datenbanken, ist das Datenformat Advanced-Function-Presentation (AFP) einschließlich seiner zahlreichen Weiterentwicklungen, wie zum Beispiel Mixed Object Document Content Architecture (MO:DCA) und das damit verbundene, für die Ausgabe an Druckgeräte optimierte Format Intelligent Printer Data Stream (IPDS) entwickelt worden. Das Format MO:DCA ist beispielsweise in der Spezifikation MO:DCA-Reference der International Business Machines Corporation (IBM), IBM-Dokument Nummer SC31-8602-06 beschrieben. Auf den Seiten 15 und 100 bis 102 dieses Dokuments sind sog. Objekt-Container erläutert, mit denen andere Daten umschlossen werden können. Auf den Seiten 542 und 543 sind Objekttypen aufgeführt, die in der MO:DCA Architektur registriert sind. Dort ist beispielsweise unter der ID Component 25 der Objekt-Typ „PDF Single-page Objekt“ genannt.

**[0003]** In der US 6,538,760 B1 ist ein Verfahren beschrieben, mit dem z. B. PDF Single-page Objekte für einen MO:DCA Datenstrom erzeugt werden können. Ein PDF Datenstrom wird dabei in einem Druckserver analysiert, der Datenstrom in einzelnen Seiten zerlegt und die einzelnen PDF Seitendaten in Objekt-Containern an ein Druckgerät ausgegeben.

**[0004]** Zur Ausgabe grafischer Informationen wurde weiterhin das Datenformat Personalized Print Markup Language (PPML) entwickelt, das im Dokument PPML 2.2, November 2006, der Organisation PODi, Rochester, NY, USA beschrieben ist.

**[0005]** Im Vergleich der oben genannten Datenströme zeigt sich, dass Datenströme, die zur Ausga-

be auf Hochgeschwindigkeits-Druckgeräten optimiert sind, wie zum Beispiel AFP oder PPML, seitenweise abschließend strukturiert sind. Der Datenstrom ist dabei so aufgebaut, dass für die Seiten jeweils ein Seitenanfangskommando und ein Seitenendkommando vorgesehen sind, und zwischen diesen Kommandos die jeweiligen Informationen für diese Seite stehen. Alle für diese Seite notwendigen Informationen, neben den Darstellungsobjekten als solchen auch die entsprechenden Steuerinformationen und Zusatzinformationen sind dabei entweder zwischen diesen beiden Seitenkommandos enthalten oder im Datenstrom bereits in vorherigen Seiten oder seitenunabhängigen einleitenden Informationen enthalten. Dadurch kann jede Dokumentenseite nach Abschluss ihres Seitenendkommandos von einer Bearbeitungsstufe, insbesondere in einem Parsing-Prozess, in dem die Daten des Dokumentendatenstroms gelesen, analysiert und entsprechend dem Sprach-Syntax interpretiert werden, sofort vollständig verarbeitet werden.

**[0006]** Der Dokumentendatenstrom kann dadurch für einen nachfolgenden Prozess sofort aufbreitet werden, beispielsweise für einen Rasterprozess, in dem aus den kodierten Seitendaten seitenweise Rasterbilder erzeugt werden. Dadurch ist es insbesondere bei der Verarbeitung von großen Datenströmen, die zum Beispiel einige zehn- oder hunderttausend Seiten enthalten, möglich, die ersten Seiten des Datenstroms bereits an einem Druckgerät auszugeben, während die nachfolgenden Seiten noch vorherige Verarbeitungsstufen, wie den Parsing-Prozess und den Raster Image Prozessor (RIP) durchlaufen.

**[0007]** Bei anderen Datenformaten, wie zum Beispiel dem PDF-Format, sind die Daten demgegenüber nicht seitenweise abschließend strukturiert, so dass zuerst alle Seiten bzw. Daten einer PDF-Datei gelesen werden müssen, bevor ein Parsing-Prozess die Seitenstruktur vollständig auflösen und die Daten zum Beispiel an einen Raster Image Prozessor zur Durchführung des Rasterprozesses übergeben kann. Im PDF-Format ist zum Beispiel am Dateiende eine Referenz-Tabelle vorgesehen, die gelesen werden muß, bevor die Seiten des Dokuments aufgelöst werden können.

**[0008]** In der Prozesskette von dem Parsing-Prozess des Dokuments bis zur Ausgabe auf einem Druckgerät entsteht dadurch ein hoher Zeitbedarf, insbesondere bei der Verarbeitung von Datenströmen mit einer großen Seitenzahl, weil das Druckgerät mit dem Druckvorgang erst beginnen kann, nachdem der vollständige Datenstrom den Parsing-Prozess durchlaufen hat.

**[0009]** Aus der EP 1 538 534 A2 ist ein Verfahren bekannt, mit dem aus einem PDF-Dokument ein PPML-Template erzeugt wird, welches dann wiederum mit

Daten eines PPML-Druckauftrags verbunden und an eine Hochgeschwindigkeitsdruckmaschine gesandt werden kann.

**[0010]** In [Fig. 2](#) ist ein Verfahrensablauf gezeigt, der beispielsweise in einem von der Anmelderin vertriebenen Drucksystem **1**, bei dem beispielsweise ein Druckserver **2** der Serie Océ PRISMA Production und ein elektrofotografisches Farb-Druckgerät **12** des Typs Océ VarioStream 9240 Anwendung findet.

**[0011]** Dabei werden beispielsweise Postscript-Dateien **3** oder PDF-Dateien **4** zur Erzeugung eines AFP-Druckdatenstroms **13** verwendet. Die Daten werden in den Druckserver **2** eingelesen und dort mit einem Rasterbildprogramm (RIP) **5** bildpunktweise gerastert. Die so gerasterten Daten werden dann in ein AFP-spezifisches Bildformat IOCA (Image Object Content) umgewandelt und die entsprechenden IOCA-Objekte **6** einem Ausschießprogramm **7** zugeführt. Eingehende Postscript- bzw. PDF-Dokumente, die eine Vielzahl von Seiten umfassen, werden dabei in IOCA-Objekte umgewandelt, bei denen pro Objekt genau eine gerasterte Seite des Dokuments enthalten ist. Die so erzeugten einseitigen IOCA-Objekte werden dann einzeln dem Ausschießprogramm **7** zugeführt, in dem die einzelnen Seiten bzw. IOCA-Objekte entsprechend einer Seitenreihenfolge ausgeschossen werden (impositioning). Details eines derartigen Ausschießvorgangs sind beispielsweise in der WO 00/68877 A1 zu entnehmen. Dabei können einzelne Seiteninformationen auch verändert bzw. mit weiteren Daten angereichert werden, beispielsweise einem Clipping-Vorgang unterzogen werden. Das Ausschießprogramm **7** gibt dann AFP-Daten aus, bei denen die einzelnen Seiten als AFP IOCA-Daten in neuer Reihenfolge und gegebenenfalls mit modifizierten Informationen ausgegeben werden. Dieser AFP-Druckdatenstrom **13** wird in einem Konverter **9** zu einem IPDS-Druckdatenstrom **14** umgewandelt und dann vom Druckserver **2** an ein Druckgerät übertragen. In dem Druckgerät **12** ist ein Datenkontroller enthalten, der Druckdaten in einer oder mehreren Sprachen interpretieren kann. Ein IPDS Parsing-Programm **10** interpretiert entsprechende IPDS-Daten und bereitet diese zur weiteren Verarbeitung im Druckgerät auf, das dann letztlich die Daten auf einen Aufzeichnungsträger druckt, beispielsweise in einem elektrofotografischen Prozess auf Papier mittels eines Zeichengenerators **11**.

**[0012]** Nachteilig bei dem in [Fig. 2](#) gezeigten Verfahren ist, dass bereits zu einem sehr frühen Zeitpunkt der Verarbeitungskette durch das Rasterbildprogramm **5** eine Rasterung von unter Umständen großen PDF-Dateien erfolgen muss, was zu einer relativ langsamen Geschwindigkeit des Gesamtprozesses führt. Insbesondere ist dabei ein Problem, dass am Druckgerät solange keine Ausgabe erfolgen

kann, bis die vollständigen PDF-Dateien den Rasterprozess durchlaufen haben.

**[0013]** Die eingangs genannten Dokumente werden hiermit durch Bezugnahme in die vorliegende Beschreibung aufgenommen.

**[0014]** Es ist Aufgabe der Erfindung, ein Verfahren zum Erzeugen und Verarbeiten eines seitenweise seriell strukturierten Dokumentendatenstroms anzugeben, bei dem Dokumentendaten eines zweiten Formats so eingebettet werden können, dass eine hohe Verarbeitungsgeschwindigkeit bei der Ausgabe des zusammengesetzten Datenstroms erreichbar ist.

**[0015]** Diese Aufgabe wird durch die in den unabhängigen Ansprüchen angegebene Erfindung gelöst. Vorteilhafte Weiterbildungen der Erfindung sind in den abhängigen Ansprüchen beschrieben.

**[0016]** Erfindungsgemäß wird ein seitenweise seriell strukturierter Dokumentendatenstrom eines ersten Formats erzeugt und/oder verarbeitet. In dem Format kann die erste Seite eines mehrseitigen Dokuments vollständig und erfolgreich einen Parsing-Prozess durchlaufen, bevor Dokumentendaten der nachfolgenden Seiten den Parsing-Prozess durchlaufen. Dokumentendaten eines zweiten Formats, die mehrere Seiten eines Dokuments umfassen, wobei für die Seiten gemeinsame Einstellungen und/oder Ressourcen gelten, werden mit dem Dokumentendatenstrom verbunden. In dem Dokumentendatenstrom des ersten Formats wird ein gesondertes Typen-Kennfeld vorgesehen, das anzeigt, dass zusammenhängende Dokumentendaten, die mehrere Seiten umfassen, mit dem Dokumentendatenstrom verbunden sind.

**[0017]** Erfindungsgemäß wurde erkannt, dass in der Verarbeitungskette von seitenweise seriell strukturierten Dokumentendatenströmen eine Einbettung von Dokumentendaten, die eine Vielzahl von Seiten umfassen, aber in einem anderen Format vorliegen, möglich ist, wenn gemeinsame Einstellungen gelten, wie zum Beispiel in einem PDF-Datenstrom oder wenn gemeinsame Ressourcen verwendet werden, beispielsweise ein gemeinsamer Zeichensatz, die in den entsprechenden Dokumentendaten mit enthalten sein können. Dann kann eine Beschleunigung des Verfahrens dadurch erreicht werden, dass derartige Daten des zweiten Formats in den Datenstrom des ersten Formats kompakt eingebettet werden. Durch eine derartige Einbettung ist es möglich, diese in nachfolgenden Prozessketten, insbesondere in einem Datenkontroller eines Druckgerätes, bei der Verarbeitung des Datenstroms einmalig zu laden und die geladenen Daten bei späteren Aufrufen von Seiten desselben ursprünglichen Dokumentendatensatzes wieder zu verwenden. Solche wieder verwendbaren Daten sind insbesondere externe Ressourcen,

die getrennt von den Dokumentendaten speicherbar sind, wie zum Beispiel Zeichensätze oder wiederkehrende Standard-Bilder. Insbesondere in dem Datenstrom AFP gibt es einen Mechanismus um derartige Ressourcen einmalig zu laden und sie mehrfach als sogenannte „sekundäre Ressourcen“ zu verwenden. Mit der Erfindung wurde erkannt, dass derartige Ressourcenhandhabung auch mit Daten anderer Formate, insbesondere PDF, möglich ist und sich durch die erfindungsgemäßen Maßnahmen die Verarbeitung derartiger Ressourcen in einem Datenstrom, in dem Daten beider Formate gemischt sind, korrelieren lässt.

**[0018]** Weiterhin ist mit der Erfindung der Vorteil erreichbar, dass sogenannte Seitenprogrammierungsvorgänge, wie zum Beispiel das Ausschließen, die Auswahl von Druckmedien oder die gemischte Ausgabe auf verschiedene Medienarten (sogenanntes Mixplexing) als auch die Anreicherung von zum Beispiel AFP-Objekten, die typisch für Druckproduktionsumgebungen sind, wie zum Beispiel BCOCA-Barcodes oder Kommandos, die zwischen Nachverarbeitungsgeräten, wie Druckgeräten, Schneidegeräten und Kuvertiergeräten über ein geeignetes Protokoll, beispielsweise über das standardisierte UP<sup>3</sup>I-Protokoll, möglich sind.

**[0019]** Mit der Erfindung wird insbesondere erreicht, dass gemeinsame Einstellparameter oder Ressourcen der verschiedenen Seiten eines Dokuments des zweiten Formats, zum Beispiel PDF, im Ausgabeprozess mehrfach wieder verwendbar sind. Dadurch können spezielle Maßnahmen zur Bereitstellung dieser Einstellungen und Ressourcen an den einzelnen Seitenpositionen eingespart werden und insbesondere ist eine Reinitialisierung des Parsing-Prozesses im ausgabeseitigen Verfahren vielfach nicht mehr nötig.

**[0020]** Mit der Erfindung wird insbesondere vorgeschlagen, ein entsprechendes Verfahren bei der Erzeugung von AFP-Datenströmen auszuführen. Sie ist jedoch auch für andere seitenweise seriell strukturierte Dokumentendatenströme, wie zum Beispiel PPML, anwendbar. Weiterhin sieht die Erfindung insbesondere vor, als eingebettete Dokumentendaten des zweiten Formats PDF-Daten zu verwenden.

**[0021]** Gemäß der Erfindung können entsprechende Dateien des zweiten Formats, zum Beispiel PDF-Dateien, die eine Vielzahl von Seiten enthalten, von einem Anwenderprogramm an den Druckserver übertragen werden. Sie können so im Druckserver in einen AFP-Dokumentendatenstrom vorteilhaft eingebettet werden, insbesondere als sogenannte inline-Resource gemeinsam mit dem übrigen Datenstrom, d. h. den Daten des ersten Formats in einer Datei, deren übergeordnete Hauptstruktur dem ersten Format entspricht.

**[0022]** Die Datei des zweiten Formats kann weiterhin vorteilhaft innerhalb eines einzigen AFP-Containers übertragen werden, entweder als inline-AFP-Resource oder auch als sog. externe Resource in einer vom übrigen Dokumentendatenstrom unabhängigen Datei.

**[0023]** Die Dokumentendaten des zweiten Formats können auch in einer von den Daten des Dokumentendatenstroms separaten Datei gespeichert werden und die Verbindung durch Referenzierung erfolgen. Die Datei kann dabei als sog. externe AFP-Resource verarbeitet werden, wobei die Verbindung zum Dokumentendatenstrom des ersten Formats logisch und über Referenzierungs-Daten erfolgt. Die Ressourcen-Datei entspricht dabei insbesondere in einer den Daten des zweiten Formats übergeordneten Hauptstruktur dem ersten Format.

**[0024]** Die Datei mit den Daten des zweiten Formats kann jeweils in zugeordneten Seitenbereichen des AFP-Dokumentendatenstroms lediglich über Referenzen auf diesen AFP-Multipage-Container aufgerufen werden, beispielsweise über das aus der AFP-Spezifikation bekannte Include-Object-Feld (IOB). Die entsprechenden Multipage-AFP-Object-Container werden dabei systematisch wie andere, bereits bekannte AFP-Object-Container behandelt, hinsichtlich Erzeugung, Verwaltung, Identifizierung, Ladevorgängen, Zwischenspeicherungsvorgängen und Löschvorgängen. Mit der Erfindung wird also eine Erweiterung vorgeschlagen, die sich ohne großen architektonischen Aufwand bzw. ohne Änderung der AFP-Spezifikation einführen lässt. Andererseits kann mit der Erfindung die Geschwindigkeit bei der Einbettung und Verarbeitung von PDF-Daten in einen AFP-Datenstrom erreicht werden, bei dem die PDF-Daten eine Vielzahl von Seiten umfassen.

**[0025]** Mit der Erfindung kann vorteilhaft vorgesehen werden:

- dass jeder Multipage bzw. Multi-Image-Container gehandhabt wird wie eine Standard-AFP-Resource.
- Die Seiten innerhalb des Containers werden individuell aufgerufen, wobei ein Indizierungsmechanismus (Seiten-Index) zu einem Include-Object-Aufruf (IOB) hinzugefügt wird. Der Indizierungsmechanismus kann insbesondere bereits in den Dokumentendaten des zweiten Formats (z. B. PDF) vorhandene seitenspezifische Daten, wie zum Beispiel die Seitennummer, verwenden. Dadurch kann in dem IOB-Aufruf beispielsweise seitenspezifisch eine Seite anhand ihrer Nummer aufgerufen werden.
- Die IOB-Aufrufe werden in eine Standardsequenz von Seitenanfang- und Seitenendkommandos eingefügt.
- Andere, bereits bekannte AFP-Kontroll-Kommandos, die auf Seiten- bzw. Blattstrukturebene

an sich bekannt sind, wie zum Beispiel Medium-Auswahl, Finishing-Befehle für Geräte, die das Ausgabemedium vor oder nach dem Druckprozess bearbeiten, Bild-Rotation usw., bleiben weiterhin verfügbar.

– Auch andere AFP-Objekttypen, wie zum Beispiel BCOCA, PTOCA, GOCA können ebenfalls weiterhin in die AFP-Seiten eingefügt werden und zusammen mit den eingebetteten Dokumentendaten des anderen Formats ausgegeben werden.

– Innerhalb einer in AFP dargestellten Seite können auch mehrer Seiten aus verschiedenen Multi-PDF-Containern oder aus einem PDF-Container eingefügt werden. Die entsprechenden Seiten können überlagert oder nebeneinander innerhalb der AFP-Seitendarstellung wiedergegeben werden.

– Weiterhin kann vorgesehen werden, dass verschiedene Seiten eines Multipage-AFP-Containers, die nicht das gleiche Set gemeinsamer PDF-Eigenschaften teilen, eine Neuinitialisierung des PDF-Parsers und/oder PDF-Raster-Image-Prozessors im ausgabeseitigen Verarbeitungsprozess, beispielsweise im Druckgerät zur Folge haben. Die jeweiligen PDF-Seiten werden dann jeweils individuell mit den entsprechenden PDF-Parametern und/oder Ressourcen verarbeitet.

**[0026]** Obwohl hinsichtlich der einzubettenden Daten vorstehend hauptsächlich auf PDF-Dateien Bezug genommen wurde, können die jeweiligen Prozesse und Einstellungen auch für Daten anderer Datenformate, z. B. auch für PCL Daten angewandt werden.

**[0027]** In einem vorteilhaften Ausführungsbeispiel der Erfindung werden die Dokumentendaten des zweiten Formats spätestens mit den Daten der ersten Seite des Datenstroms eingebettet, in denen eine Seite des zweiten Formats aufgerufen wird. Dadurch kann gewährleistet werden, dass bei der ausgabeseitigen Verarbeitung des Datenstroms sämtliche Informationen für den Parsing- und Rasterprozess von der ersten Seite ab vollständig zur Verfügung stehen, die im Format des zweiten Formats vorliegen.

**[0028]** Insbesondere werden die Daten des kompletten Dokuments des zweiten Formats vollständig übertragen, bevor deren erste Seite aufgerufen wird. Dies kann innerhalb der ersten Seitendarstellung im Dokumentendatenstrom erfolgen oder auch in einem Bereich mit allgemeinen, seitenunabhängigen Kopf-Ressourcen, die unabhängig von den übrigen Daten des Dokumentendatenstroms übertragen wird.

**[0029]** Die Multipage-Object-Container können innerhalb des AFP-Datenstroms übertragen oder auch unabhängig von den übrigen Daten des Datenstroms an das Ausgabegerät übertragen werden.

**[0030]** Die Erfindung lässt sich insbesondere für Dokumentendaten des zweiten Formats vorteilhaft anwenden, bei denen formatbedingt für den Abschluss eines Parsing-Prozesses, bei dem die Daten gelesen, analysiert und interpretiert werden, der ersten Seite die gesamten Dokumentendaten gelesen und interpretiert werden müssen. Wenn der Datenstrom ein AFP-Datenstrom ist, dann kann insbesondere vorgesehen werden, dass das Typen-Datenfeld vom Ressourcen-Typ „Object-Container“ ist, wobei es insbesondere als Multipage-Object-Container von anderen Object-Containern unterscheidbar ist.

**[0031]** Für andere seitenweise strukturierte Dokumentendatenformate wie z. B. PPML kann ebenfalls jeweils ein gesondertes, Multipage spezifisches Strukturelement vorgesehen werden.

**[0032]** Zum Weiterverarbeiten eines Dokumentendatenstroms, der erfindungsgemäß erzeugt ist, kann insbesondere vorgesehen sein, zumindest den Teil der Dokumentendaten des zweiten Formats einen vollständigen Parsing-Prozess durchlaufen zu lassen, der mehrfach verwendet wird, bevor die Daten der ersten Seite des Dokuments, auf die Daten des zweiten Dokuments referenziert werden, weiterverarbeitet werden.

**[0033]** Die geparsten Dokumentendaten des zweiten Formats können dabei mindestens solange gespeichert und bei einem Aufruf durch Daten nachfolgender Seiten wiederverwendet werden, bis alle Seiten des Dokuments, in denen Dokumentendaten des zweiten Formats referenziert werden, ebenfalls geparst sind.

**[0034]** Erfindungsgemäß sind auch ein dem erfindungsgemäßen Verfahren entsprechendes Computerprogrammprodukt, ein Druckserver und ein Drucksystem vorgesehen.

**[0035]** Weitere Einzelheiten und Vorteile der Erfindung werden anhand der nachfolgenden Beschreibung von Ausführungsbeispielen deutlich.

**[0036]** Es zeigen:

**[0037]** [Fig. 1](#): ein Verfahren zur Erzeugung eines erfindungsgemäßen Druckdatenstroms,

**[0038]** [Fig. 2](#): ein Verfahren zum Erzeugen eines AFP-Datenstroms gemäß dem Stand der Technik und

**[0039]** [Fig. 3](#): eine beispielhafte Datenstruktur für einen AFP-Druckdatenstrom.

**[0040]** In [Fig. 1](#) werden, soweit es sich um gleiche Elemente handelt, dieselben Bezugszeichen ver-

wendet, wie bereits in der vorstehend beschriebenen [Fig. 2](#).

**[0041]** Bei dem in [Fig. 1](#) gezeigten Verfahrensablauf werden Postscript-Daten mittels eines Normalizer-Programms **20**, beispielsweise dem von der Firma Adobe Systems Inc. bekannten Normalizer-Programms verarbeitet und der dabei erzeugte Datenstrom in einen Multipage-Object-Container **6** im Datenformat AFP eingebettet. Dasselbe kann mit PDF-Daten **4** mittels eines Konversionsprogramms **21** erfolgen. Der so erzeugte AFP-Object-Container **6** stellt eine Ressource dar. Diese Ressource kann mit dem Ausschließprogramm **7** wiederum so verarbeitet werden, dass eine Multipage-Object Content Ressource **22** entsteht, mittels derer Seiten in einer geänderten Reihenfolge gemäß der im Ausschließprogramm **7** vorgenommenen Umsortierung ausgegeben werden können. Im IPDS-Parsing-Programm **10** bzw. dem damit verbundenen Parsing-Vorgang werden die eingebetteten Postscript bzw. PDF-Daten sukzessive entpackt, beispielsweise mittels eines PDF-Parsing-Programms **23**. Dabei werden im Zuge der ersten verarbeiteten PDF-Daten die jeweiligen übergeordneten bzw. mehrfach genutzten PDF-Einstellungen oder PDF-Ressourcen geladen, verarbeitet und zwischengespeichert. Die zwischengespeicherten Daten können dann sukzessive beim Aufruf der jeweiligen Seiten in Verbindung mit den dort vorkommenden individuellen Seitendaten verbunden werden, bevor sie an den Zeichengenerator **11** ausgegeben werden. In diesem Zusammenhang ist es auch möglich, dass vorgegebene Bildobjekte, wie zum Beispiel Logos, die als Ressourcen zur Verfügung stehen, einmalig gerastert und auf den jeweiligen Seiten mehrfach wiederverwendet werden, ohne dass für die nachfolgenden Seiten jeweils ein neuer Rastervorgang ausgeführt werden muss.

**[0042]** In [Fig. 3](#) ist eine Datenstruktur gezeigt, die beispielsweise in einem AFP-Dokumentendatenstrom verwendet werden kann. In einer ersten Daten-Gruppe **25** werden Object-Container des Typs Multipage bzw. Multiimage definiert. In einem zweiten Bereich **26** werden Ressourcen definiert. In einem dritten Bereich **27** werden die für eine individuelle Seite notwendigen Daten aufgerufen. Bei dem seriellen Abarbeiten des Datenstroms wirken die Object-Containerdaten des Bereichs **25**, die Pre-Process-Presentation-Object-Daten des Bereichs **26** und die Include-Object-Daten des Bereichs **27** so zusammen, dass in der aufgerufenen Seite die entsprechende Seite der ursprünglichen PDF-Datei aufgerufen wird, wobei die PDF-Einstellungen und Ressourcen durch die vorherige Abarbeitung, beispielsweise in den Bereichen **25**, **26** wieder verwendbar sind.

**[0043]** Für AFP-Dokumentendatenströme wird vorgeschlagen, folgende Object Container vorzusehen:

- PDF Document Object
- PDF Document Object with transparency

**[0044]** Diese Datenobjekte enthalten ein PDF-Dokument bzw. eine PDF-Datei, die mehrere Seiten enthalten kann. Die Object Container werden von einem Druckserver an ein Druckgerät gesandt, insbesondere im Grundzustand des Druckgerätes, z. B. im sog. Homestate-Zustand eines IPDS-fähigen Druckers.

**[0045]** Die Datenobjekte werden durch Identifier (HAID) identifiziert und können gespeichert und aktiviert werden durch ein AFP-Activate-Ressource-Kommando. Innerhalb einer in AFP kodierten Seite oder eines Overlays können sie durch das IPDS-Kommando IDO eingefügt werden.

**[0046]** Durch ein AFP Activate-Ressource (AR) Kommando wird das gesamte Multipage-Dokument aktiviert. Einzelne Seiten oder Bereiche des Dokuments werden dabei nicht gesondert aktiviert.

**[0047]** Mittels eines optionalen AFP Rasterized-Presentation-Object (RPO) Kommandos kann entweder eine Vorab-Rasterung des kompletten Dokuments erfolgen oder einzelne Seiten des PDF-Dokuments bzw. der PDF-Datei erfolgen. Durch eine derartige Vorab-Rasterung können insbesondere sehr große oder anderweitig raster-aufwendige, kritische Anwendungen die Verarbeitungsgeschwindigkeit bei der Ausgabe des Dokuments insgesamt erhöhen.

**[0048]** Durch eine optionale Seitensequenznummer innerhalb der RPO-Einträge ist es möglich, eine bestimmte Seite eines PDF-Dokuments vorab dem Rasterprozess zu unterziehen. Ist in dem RPO-Eintrag keine Seitennummer angegeben, so wird das gesamte Dokument, das heißt alle Seiten bzw. Bilder gerastert.

**[0049]** Wenn das komplette Dokument gerastert wird, so kann ein gemeinsames Set von Steuerungsparametern aus dem RPO-Kommando für alle Seiten des Dokuments verwendet werden. Wenn Seiten bzw. Bilder separat gerastert werden sollen, so kann jedes entsprechende Objekt ein eigenes Set von Steuerungsparametern haben. Jedes Objekt enthält dann einen eigenen Eintrag innerhalb des RPO-Kommandos.

**[0050]** Das AFP Include-Data-Objekt-Kommando (IDO) wird benutzt, um eine ausgewählte Seite des PDF-Dokuments in eine Seitenstruktur eines IPDS-Datenstroms oder in eine entsprechende Overlay-Struktur einzubinden. Innerhalb eines IDO Data-Object-Data-Distriktors kann dann die Nummer der gewünschten Seite bzw. des gewünschten Bildes angegeben werden.

**[0051]** Nachfolgend wird beschrieben, wie eine große PDF-Datei, wie zum Beispiel ein Buch in den AFP- bzw. IPDS-Datenstrom eingebunden wird, bei der keine weitere Bearbeitung oder Änderung vorgenommen wird.

– Schritt 1: Die PDF-Dokumenten-Datei wird an den Drucker als Homestate-Object-Container, HAID#0001 gesandt.

– Schritt 2: Das optionale RPO-Kommando wird benutzt um alle Seiten des Dokuments zu rastern. Wenn das RPO-Kommando nicht benutzt wird, werden die Seiten erst im nächsten Schritt gerastert.

– Schritt 3: Es wird eine Schleife durchlaufen mit der Kommandosequenz

Beginn Page

Include-Data-Objekt (HAID#0001, Page N)

End Page

Dadurch werden sukzessive alle Seiten des Dokuments seriell abgearbeitet und ausgegeben.

**[0052]** Mit den oben beschriebenen Maßnahmen ist eine beträchtliche Beschleunigung bei der Verarbeitung von mehrseitigen PDF-Dateien möglich. Um diese weiter zu verbessern ist auch denkbar, sehr große Dateien in kleinere Dateien zu segmentieren. Dadurch kann zum Beispiel die Zeit bis zum ersten Druck weiter verkürzt werden. In diesem Fall müssten jedoch Maßnahmen ergriffen werden, mit denen für spätere Multipage-Container angezeigt wird, dass für diese dieselben Parameter und gegebenenfalls Ressourcen gelten wie für die früheren Multipage-Container. Dies könnte zum Beispiel durch eine gemeinsame Identifikationsnummer (ID) erfolgen. Eine weitere Identifikationsnummer könnte dazu vorgesehen werden, um anzuzeigen, dass Multipage-PDF-Dateien in eine gemeinsame AFP-Seite zusammengeführt werden sollen.

**[0053]** Schließlich kann es vorteilhaft sein, innerhalb des AFP-Dokumentendatenstroms Maßnahmen vorzusehen, durch die ein automatischer Rasterprozess in einem Druckserver oder in einem Druckgerät aktiviert oder deaktiviert wird. Dies könnte zum Beispiel durch eine neue AFP Medium-Map-Anweisung erfolgen.

#### Bezugszeichenliste

1	Drucksystem
2	Druckserver
3	Postscript-Datei
4	PDF-Datei
5	Rasterbildprogramm (RIP)
6	IOCA-Objekt
7	Ausschießprogramm
8	IOCA-Objekt
9	Konverter
10	IPDS Parser
11	Zeichengenerator

12	Druckgerät
13	AFP-Druckdatenstrom
14	IPDS-Druckdatenstrom
20	Normalizer
21	Konversionsprogramm
22	Object-Container
23	PDF Parser
25	Object-Container-Definition
26	Ressource-Definition
27	Seitendefinition

#### Patentansprüche

1. Verfahren zum Erzeugen eines seitenweise seriell strukturierten Dokumentendatenstroms eines ersten Formats,

– wobei in dem ersten Format die erste Seite eines mehrseitigen Dokuments vollständig und erfolgreich einen Parsing-Prozess durchlaufen kann, bevor Dokumentendaten der nachfolgenden Seiten den Parsing-Prozess durchlaufen,

– wobei eine Datei mit Dokumentendaten eines zweiten Formats mit dem Dokumentendatenstrom verbunden wird, wobei die Dokumentendaten des zweiten Formats mehrere Seiten eines Dokuments umfassen und wobei für die Seiten gemeinsame Einstellungen und/oder Ressourcen gelten,

– wobei in dem Dokumentendatenstrom des ersten Formats ein gesondertes Typen-Kennfeld vorgesehen wird, das anzeigt, dass die Datei mit den zusammenhängenden Dokumentendaten des zweiten Formats mit dem Dokumentendatenstrom verbunden ist, – und wobei die Verbindung durch eine Einbettung der Datei mit den Dokumentendaten des zweiten Formats in den Dokumentendatenstrom des ersten Formats erfolgt oder die Datei mit den Dokumentendaten des zweiten Formats separat von den Daten des Dokumentendatenstroms des ersten Formats gespeichert wird und die Verbindung durch Referenzierung erfolgt.

2. Verfahren nach Anspruch 1, wobei die Dokumentendaten des zweiten Formats spätestens mit den Daten der ersten Seite des Dokumentendatenstroms des ersten Formats eingebettet werden, in denen eine Seite des zweiten Formats aufgerufen wird.

3. Verfahren nach einem der Ansprüche 1 oder 2, wobei bei den Dokumentendaten des zweiten Formats formatbedingt zum vollständigen Parsing der ersten Seite die gesamten Dokumentendaten des zweiten Formats gelesen und interpretiert werden müssen.

4. Verfahren nach einem der vorhergehenden Ansprüche, wobei das zweite Format das Portable Document Format (PDF) ist.

5. Verfahren nach einem der Ansprüche 1 bis 3, wobei das zweite Format Page Description Language (PCL) ist.

6. Verfahren nach einem der vorhergehenden Ansprüche, wobei die Dokumentendaten des zweiten Formats seitenweise indiziert werden und ein Seiten-Index gebildet wird.

7. Verfahren nach Anspruch 6, wobei bei der Weiterverarbeitung des Dokumentendatenstroms die Seiten individuell unter Verwendung des Seiten-Index aufgerufen werden.

8. Verfahren nach einem der vorhergehenden Ansprüche, wobei das erste Format Advanced Function Presentation (AFP), Mixed Object Content Architecture (MO:DCA) oder Intelligent Printer Data Stream (IPDS) ist.

9. Verfahren nach Anspruch 8, wobei das Typen-Datenfeld vom Ressourcen-Typ Object Container ist und insbesondere ein Multipage Object Container ist.

10. Verfahren nach Anspruch 9, wobei bei der Weiterverarbeitung des Dokumentendatenstroms Seiten innerhalb des Multipage Object Containers über eine Include Object Anweisung aufgerufen werden.

11. Verfahren nach einem der Ansprüche 1 bis 7, wobei das erste Format Personalized Print Markup Language (PPML) ist.

12. Verfahren nach einem der vorhergehenden Ansprüche, wobei eine Datei mit den Dokumentendaten des zweiten Formats in kleinere Dateien segmentiert wird.

13. Verfahren zum Weiterverarbeiten eines Dokumentendatenstroms, der nach einem der vorhergehenden Ansprüche erzeugt ist, wobei der Dokumentendatenstrom geparkt wird.

14. Verfahren nach Anspruch 12, wobei die Dokumentendaten des zweiten Formats vollständig geparkt werden, bevor die Daten der ersten Seite des Dokuments, auf die Daten des zweiten Dokuments referenziert werden, weiterverarbeitet werden.

15. Verfahren nach Anspruch 14, wobei geparkte Dokumentendaten des zweiten Formats, mindestens so lange gespeichert und bei einem Aufruf durch Daten nachfolgender Seiten wieder verwendet werden, bis alle Seiten des Dokuments, in denen Dokumentendaten des zweiten Formats referenziert werden, geparkt sind.

16. Computerprogramm, durch das beim Laden und Ausführen auf einem Computer ein Verfahren

nach einem der vorhergehenden Ansprüche ausgeführt wird.

17. Druckserver auf dem ein Computerprogramm nach Anspruch 16 geladen ist.

18. Drucksystem umfassend einen Computer, auf dem ein Computerprogramm nach Anspruch 16 geladen ist.

19. Drucksystem nach Anspruch 18 wobei der Computer als Datencontroller eines Druckgeräts vorgesehen ist.

Es folgen 3 Blatt Zeichnungen



Anhängende Zeichnungen

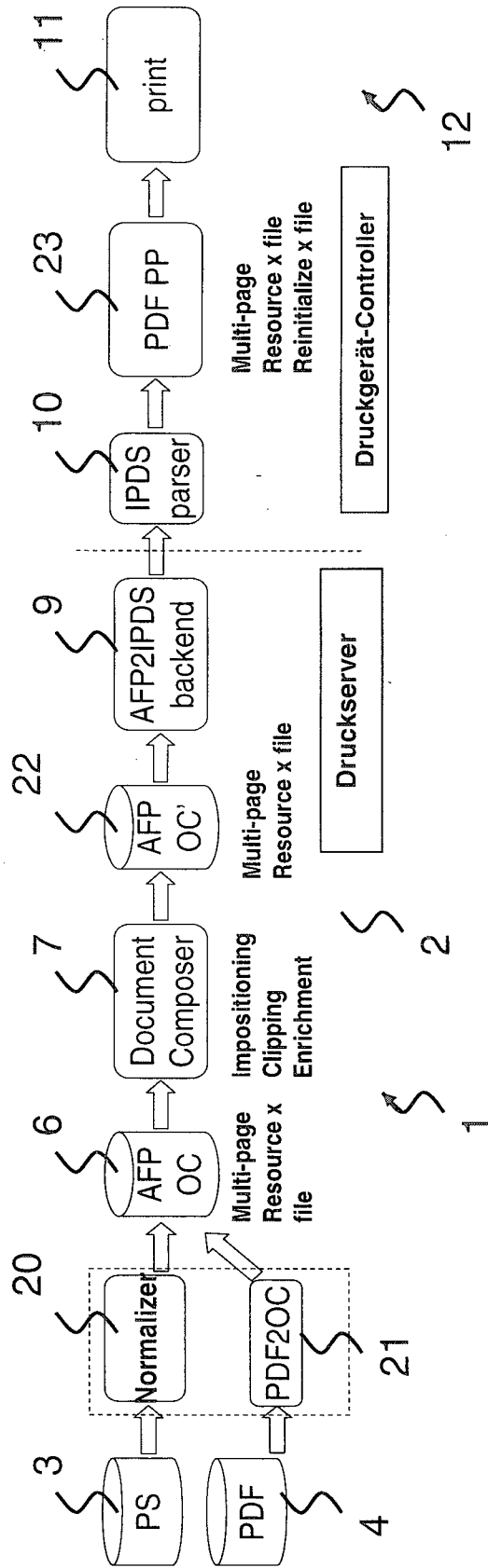


Fig. 1

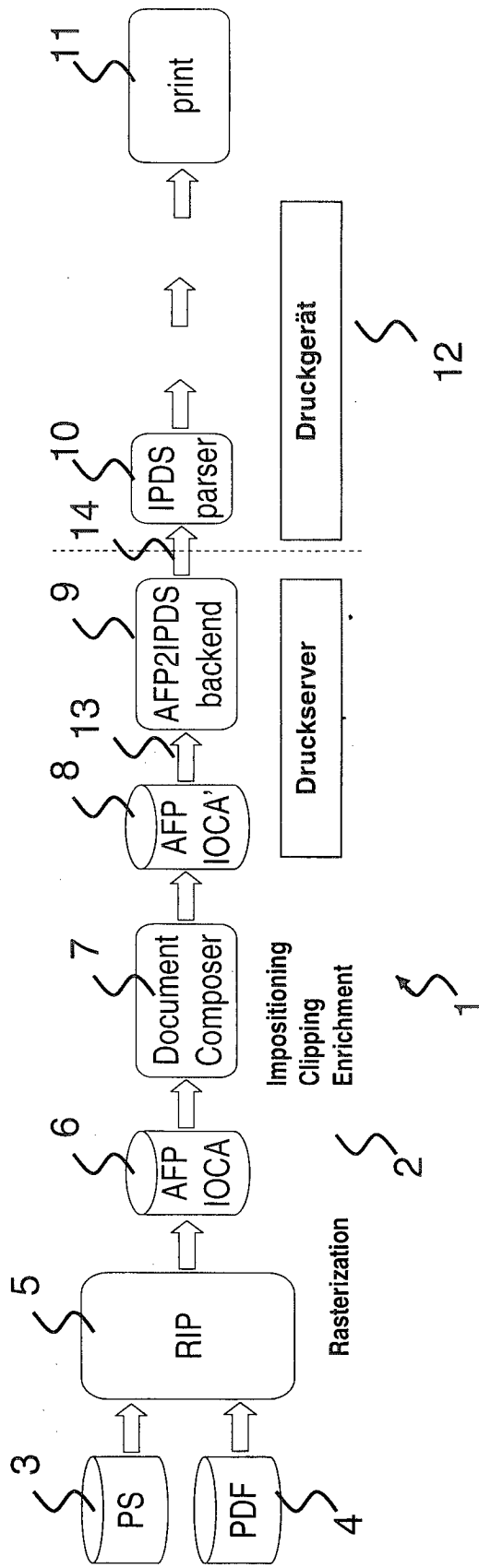
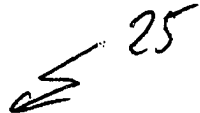
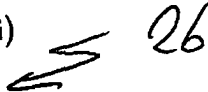


Fig. 2  
Stand der Technik

Begin Object Container (BOC)  25  
  Object Environment Group  
  **Object Container Data (Multi- page/image container)**  
End Object Container

Begin Resource Environment Group (BSG)  26  
  Map Data Resource (MDR)  
  Map Page Overlay (MPO)  
  **Preprocess Presentation Object (PPO)**  
End Resource Environment Group (ESG)

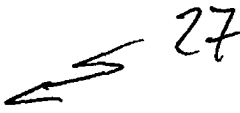
Begin Page (BPG)  27  
  Page Environment Group  
  ...  
  Include Object (IOB)  
    Object area position, rotation, Mapping  
    Object name / type: multi-page PDF container  
    **Object offset triplet (0x5A)**  
  ...  
End Page (EPG)

Fig. 3