



Der calibrate  
Prepress Coach

# Radix Translate

Schnittstelle zwischen Sender  
und Empfänger



**RADIX**

Print 4.0 ganz einfach

**TEIL 1**

Radix Map

**TEIL 2**

Radix Project

**TEIL 3**

Radix Preflight

**TEIL 4**

Radix Translate

**TEIL 5**

Radix Prepare

# Inhalt

<b>Management Summary</b>	Seite 03
<b>Einleitung</b>	Seite 04
<b>Wer führt das System?</b>	Seite 04
<b>Wann haben Daten eine hohe Qualität?</b>	Seite 05
<b>Metadaten</b>	Seite 08
<b>XML, JSON und Struktur</b>	Seite 09
<b>Lesbar für Mensch und Maschine</b>	Seite 09
<b>Benefit und Kontakt</b>	Seite 10

Die calibrate Workflow-Consulting GmbH realisiert als Dienstleister weltweit den automatisierten Prozess zur Herstellung von gedruckten und elektronischen Medien. Radix ist das modulare Produktpaket von calibrate, bestehend aus Dienstleistungen und Software. Im Vordergrund steht die Automatisierung der Prozesse. Hierbei geht es um die Sicherheit im Abgleich der Produktionsdaten mit den Metadaten für die Produktion. Ziel ist Digitalisierung – Print 4.0 ganz einfach.

Der calibrate Prepress Coach umfasst fünf Teile. In Teil 4 "Radix Translate – Schnittstelle zwischen Sender und Empfänger" geht es um die Kommunikation im Workflow.

## Management Summary

Vom „führenden System“ sprechen wir häufig bei der Software-Systemintegration. Der Begriff ist sinnvoll, wenn mehrere Systeme betrachtet werden, in denen dieselben Informationen verwendet oder Informationen abgeleitet werden (Beispiel Metadaten). Die Überlegung ist aus der Problematik entstanden, eine Informationsquelle zu bestimmen, die z. B. als richtig oder originär angenommen werden kann.

Nehmen wir an, dass ein Kunde eine Bestellung mit Angabe seiner Adresse in einem Webshop (System A) tätigt. Die Informationen werden in der Datenbank A gespeichert. Im Auftragsbearbeitungssystem (System B) werden diese Daten als ein Auftrag in der Datenbank B erzeugt. Ein Mitarbeiter der Auftragsbearbeitung erkennt einen Fehler in der Adresse, prüft die Adresse und korrigiert sie in System A. Nach der erfolgten Korrektur liegt ein Unterschied der Daten in Datenbank A und Datenbank B vor. Ein weiteres System kann dem Versand dienen (System C). An dieser Stelle muss bereits geklärt sein, welches System (A oder B) führend ist. Bezieht C seine Daten aus A oder B?

In der Praxis definiert man ein führendes System für die Verteilung von Metadaten. Eines der bestehenden Systeme wird als führendes System für eine Stammdatenklasse (= Adresse, Produktbeschreibung und ähnliches) definiert und ist damit Ausgangspunkt für die Verteilung der Informationen. Das erstmalige Anle-

gen der Stammdaten geschieht im führenden System mit den dort vorhandenen Eigenschaften. In der Folge werden die Datensätze mit weiteren Eigenschaften angereichert. Änderungen in anderen Systemen müssen fortan an das führende System gemeldet werden oder aber dürfen nur in diesem erfolgen. Die mit dem führenden System vernetzten weiteren Systeme sind danach für den lokalen Abgleich verantwortlich.

Eine Erschwernis ergibt sich aus heterogenen Systemlandschaften, also verschiedenen Systemen unterschiedlicher Hersteller. Dem Sender muss klar sein, welche Daten der Empfänger erwartet und wie diese aussehen. Häufig sind selbst XML-Daten proprietär, also herstellerspezifisch. Das kann bedeuten, dass die Verwendung der Daten durch die Nichtverfügbarkeit des Quellcodes eingeschränkt ist. Proprietär bedeutet „im Eigentum (des Herstellers) befindlich“. Das heißt: Wenn jemand eine XML-Schnittstelle anbietet, bedeutet das noch lange nicht, dass wir seine Sprache auch verstehen.

# Einleitung

In der Druckindustrie reklamieren die ERP-Systeme für sich, das führende System zu sein. Das mag für klassische Druckereien mit Außendienst und mit etwa 3.500 Aufträgen pro Jahr in den meisten Fällen auch richtig sein. Wie aber sieht es aus, wenn 35.000 Aufträge oder mehr im Jahr zu bewältigen sind? Die Mengen sind initiale Auslöser für das Bedürfnis einer Automatisierung der Prozesse.

Ein großes Hindernis in der Kommunikation sind Formate und Strukturen – vergleichbar mit Sprachen und Dialekten, welche eine Übersetzung zwischen Datensender und Datenempfänger notwendig machen. Dabei kann **Radix Translate** mehr leisten als eine schlichte Übersetzung, die definierten Regeln folgt. Aus der Information erschließen sich Aktionen, die nachfolgende Prozesse anstoßen.

## Wer führt das System?

ERP-Systeme führen, basierend auf Informationen über Ressourcen, eine rechtzeitige und bedarfsge- rechte Planung und Steuerung der Auftragsabwick- lung durch. Darüber hinaus ermöglichen Stamm- und Bewegungsdaten dem ERP-System eine transpa- rente und effiziente Erstellung von Arbeitsaufträgen.

Die Webshop-Entwickler bildeten anfänglich sehr einfache Auftragsstrukturen, meist Produkte eines Warenkorbes von einem Besteller an einer Lieferad- resse, ab. Das funktioniert, wie Amazon unter Beweis stellt, für unveränderliche Consumables sehr gut. Doch für die Druckindustrie mit den Möglichkeiten der individuellen Gestaltung war das zu wenig. Gestaltung und komplexere Auftragsstrukturen sind heute mög- lich. Gleich den Webshops liefern auch andere Systeme wie PIM, MAM oder DAM Daten für die Produktion.

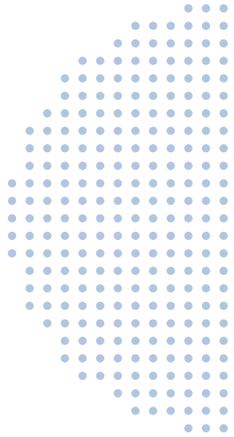
Als die Dritten im Bunde können die Anbieter der Workflows für die Produktion genannt werden. Praktisch hat jeder der großen Anbieter wenigstens ein Upload-Portal im Portfolio.

Mit dem Ziel, in der Gesamtinfrastruktur eine höhere Abhängigkeit vom eigenen System zu schaffen, erwei- tern die Software-Anbieter stetig die Kompetenzen der eigenen Systeme, wodurch Überschneidungen bei

der Abwicklung von Aufgaben entstehen. So brachten die ERP-Anbieter einen Webshop auf den Weg, wäh- rend die Webshop-Entwickler ein Auftragsmanage- ment anbieten, und auch die Workflow-Spezialisten der Produktion stehen auf dem Feld der All-in-One- Lösung. Man kann von einer Konkurrenzsituation zwischen den Systemwelten sprechen. Da alle Systeme notwendig sind, kommt es für die Anwender auf eine saubere Synchronisation an. Grundvoraus- setzung hierfür ist die verlustfreie Kommunikation der Systemlandschaften.

### An dieser Stelle gibt es zwei gute Nachrichten:

- 1) Es gibt kein grundsätzliches Richtig oder Falsch. Die Eignung der Systeme ist vom Anwendungsfall abhängig und bedarf immer eines Proof of Concept, also einer Machbarkeitsstudie, welche die prinzipielle Durchführbarkeit der Vernetzung belegt.
- 2) Die meisten Systeme sind in der Lage, strukturierte Daten zu exportieren. Diese Daten lassen sich in der Regel vom Sender zum Empfänger übersetzen. Welches System also tatsächlich das „führende System“ ist, hängt von unterschiedlichen Faktoren ab. Eine wichtige Frage ist immer: Was soll mit all den gesammelten Daten am Ende geschehen? Müssen diese gespeichert werden, oder können wir nach Auftragsende den Datensatz „vergessen“?



# Wann haben Daten eine hohe Qualität?

Grundsätzlich sind Daten dann von einer hohen Qualität, wenn sie den vom Nutzer festgelegten Zweck erfüllen. Insofern wird Datenqualität auch häufig mit dem Begriff „Fitness for Use“ beschrieben. Die konkretere Bestimmung der Qualität sollte verschiedene Kriterien berücksichtigen. Diese Kriterien sind Vollständigkeit, Aktualität, Konsistenz, Zugänglichkeit und Interpretierbarkeit.

- Daten sind dann vollständig, wenn alle definierten Kriterien erfüllt sind, also kein Anteil an Null-Werten mehr besteht. In der Fertigung müssen z. B. vor Anlauf sämtliche Eigenschaften des Produkts (z. B. Format, Farbe, Papier etc.) vorliegen, bevor mit der Fertigung begonnen werden kann.
- Die Aktualität beschreibt, wie zeitnah Datenwerte bei einer Änderung des realen Zustands angepasst werden. Anforderungen an Aktualität können unterschiedlich sein: So ist beispielsweise die Angabe von Liefertermin und -adresse während der Produktion ausreichend, wohingegen die Aktualität von Produktdaten bei Just-in-Time-Produktion in Echtzeit abrufbar sein muss.

- Konsistente Datenbestände zeichnen sich dadurch aus, dass sie frei von Widersprüchen sind. Lieferantenstammdaten sind z. B. konsistent, wenn für jeden Lieferanten eine eindeutige Kennzeichnung existiert, also keine Dubletten vorhanden sind.
- Der Umfang, in dem Daten für den Nutzer verfügbar bzw. auffindbar sind, wird durch das Kriterium der Zugänglichkeit bestimmt. Bei Online-Verkaufskanälen müssen Artikelinformationen für den Konsumenten in Echtzeit zugänglich sein.
- Mit Interpretierbarkeit meint man, inwieweit Daten in geeigneten Formaten vorliegen, und dabei klare Regeln und Definitionen eingehalten werden. So müssen Daten für die Bezahlung (z. B. ZUGFeRD) gewissen syntaktischen Vorgaben entsprechen, damit eine automatische Verarbeitung erfolgen kann.

Die genannten Kriterien können die Qualität der Daten konkretisieren und „messbar“ machen, denn erst über die Messung ist die Ableitung von Maßnahmen zur Verbesserung der Datenqualität möglich.



Abb. 1: Die 15 Dimensionen der Datenqualität: Definition der deutschen Gesellschaft für Informations- und Datenqualität (DGIQ) auf Basis der Forschungsergebnisse von Richard Wang und Diane Strong (Quelle: DGIQ IQ-Definition – 2007).

Aufgrund der steigenden Anforderungen wachsen sowohl die Bedeutung der Stammdatenqualität in Unternehmen als auch die Anstrengungen, Qualitätsprobleme zu beheben. Ursachen für derartige Probleme sind laut einer Studie von Pricewaterhouse-Coopers häufig folgende Faktoren:

■ **Fehler bei der Dateneingabe**

Irgendjemand gibt die Daten am Anfang jedes Prozesses ein. Manuelle Eingaben beherbergen die Gefahr von Fehlern. Typische Fehler sind Rechtschreibfehler, die Eingabe von unpassenden Angaben in Pflichtfelder, die nicht automatisch auf die Verwendung erlaubter Werte überprüft werden, oder die fehlerhafte Auswahl von Werten aus vorgegebenen Listen.

■ **Widersprüchliche Definitionen für übergreifend verwendete Datenobjekte**

Gerade wenn zu einem bestehenden System ein neues hinzukommt, kann dieses Problem zum Tragen kommen. Ist das Datenmanagement historisch unabhängig voneinander in unterschiedlichen Systemen entstanden, dann sind widersprüchliche Definitionen kaum vermeidbar. Das wird klar, wenn

wir mehrere Personen nach der Definition für Formate fragen – offen, geschlossen, beschnitten und so weiter.

■ **Fehler bei der Migration von Daten in neue Systeme**

Bei der Einführung oder Zusammenführung von betrieblichen Informationssystemen (z. B. ERP-Systeme) müssen Daten von Altsystemen in die neue Systemumgebung überführt werden. Liegen die Daten nicht konsistent und syntaktisch einheitlich interpretierbar vor, kommt es bei der – zumeist automatisierten – Datenübernahme zu falschen Einträgen.

■ **Fehler bei der Integration von Daten aus mehreren Systemen oder aus externen Quellen aufgrund unterschiedlicher Datenmodelle**

Häufig müssen Daten aus unterschiedlichen Quellsystemen zusammengeführt und aggregiert werden. Besitzt ein Unternehmen mehrere Systeme (ERP, Webshop, Portal, etc.), möchte aber eine Analyse über die Aufträge durchführen, müssen beispielsweise die Materialstammdaten in einem einheitlichen Modell vorliegen oder mindestens in ein einheitliches Modell überführt werden können.

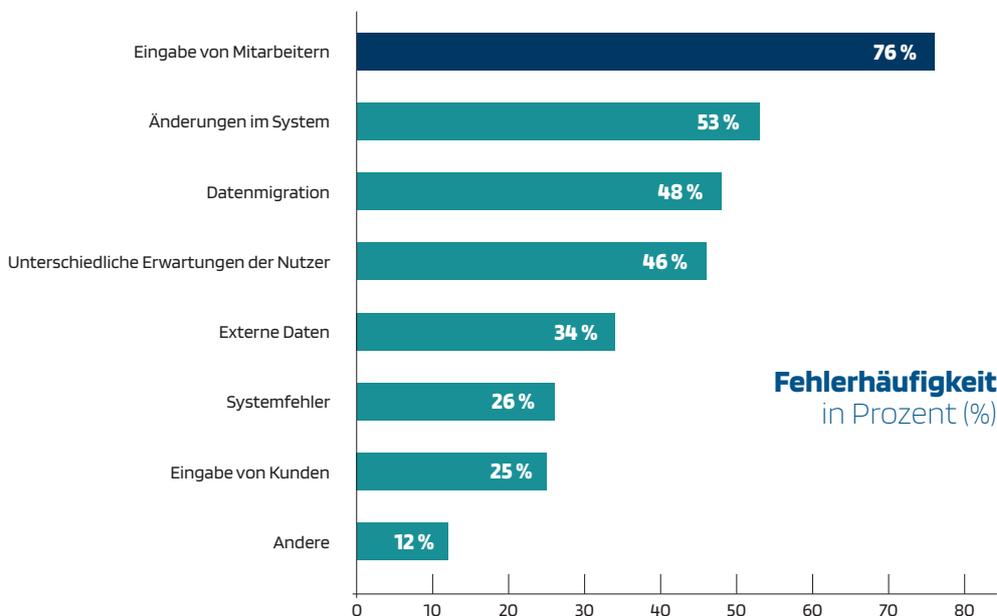


Abb. 2: Die Datenqualität ist von unterschiedlichen Faktoren abhängig. Allem voran steht zumeist jedoch die Eingabe durch den Mensch. (Quelle: The Data Warehousing Institute).



Um die unterschiedlichen Anforderungen an die Übersetzung von Daten besser zu verstehen, lohnt sich ein Blick auf die unterschiedlichen Stammdatenarchitekturen. Zur Umsetzung des Stammdatenmanagements in einer verteilten Systemlandschaft stehen vier idealtypische Architekturansätze zur Verfügung: Führendes Stammdatensystem, zentrales Stammdatensystem, Verbindung über ein Verzeichnis (Registry) und Standards. Die Ansätze lassen sich anhand der Dimensionen „Stammdatenmodell“ und „Stammdatenpflege und -haltung“ unterscheiden. In einem harmonisierten Stammdatenmodell sind die Stammdatenattribute (z. B. Name und Feldlänge) und deren Inhalte (z. B. eindeutige Identifikatoren oder Wertebereiche) global, d. h. je nach Reichweite unternehmensweit oder unternehmensübergreifend, festgelegt. Bei Stammdatenpflege und -haltung unterscheidet man, inwieweit Stammdaten zentral angelegt und gespeichert werden. Die zentrale Erfassung und Pflege der Stammdaten sichert hingegen die Konsistenz von Stammdaten, da mit dem zentralen System ein „Single Point of Truth“ existiert. Allerdings gilt auch: je zentraler der Ansatz, desto geringer die Flexibilität.

Von einem zentralen Stammdatensystem ist die Rede, wenn die Führung der Stammdaten in einem separaten Stammdatensystem erfolgt, das diese an die lokalen Systeme verteilt. Die Erfassung und Pflege

findet grundsätzlich in einem zentralen System statt, allerdings auf Basis eines harmonisierten Stammdatenmodells.

„Standards“ führen nicht zu einer zentralen Speicherung und Verteilung von Stammdaten, es werden nur unternehmensweit einheitliche Strukturen definiert. Ein harmonisiertes Stammdatenmodell gewährleistet, dass Aufbau und Befüllung eines Stammdatensatzes über verschiedene Systeme hinweg gleich sind.

Ein „führendes System“ ist das am häufigsten verwendete Verfahren für die Stammdatenverteilung: Eines der bestehenden Systeme wird als das führende System für eine Stammdatenklasse definiert und ist damit Ausgangspunkt für die Verteilung an die anderen Systeme.

„Verbindung über ein Verzeichnis“ (Registry) ist der Weg, ein übergreifendes Verzeichnis zu implementieren, das Zuordnungen der verschiedenen Stammdatensätze zu den verschiedenen Quellsystemen enthält. Benötigt z. B. ein System Daten zu einem bestimmten Kunden, so startet es eine Anfrage an die Registry und erhält eine Antwort darüber, in welchem System die Daten zu dem jeweiligen Kunden abgelegt sind. In einem weiteren Schritt werden die Daten dann direkt aus dem entsprechenden System abgerufen.

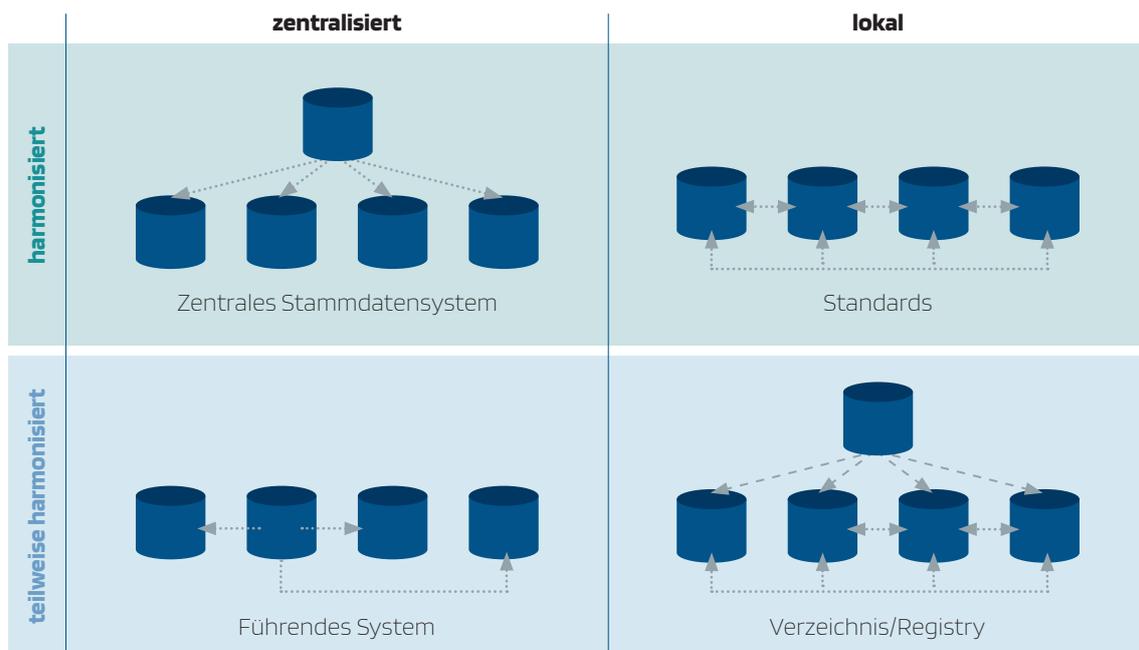


Abb. 3: Stammdatenpflege und -haltung

# Metadaten

Welche Daten kann der Sender liefern? Ein erster Schritt auf der Reise durch die PDF-Eigenschaften und Metadaten kann erstmal in einem in Adobe Acrobat geöffneten PDF beginnen. Dort findet sich unter „Datei/Eigenschaften“ schon eine ganze Menge an Hinweisen zu den Dokumenteigenschaften. Die Beschreibung zeigt grundlegende Informationen zum Dokument an. Titel, Verfasser, Thema und Stichwörter können vom Ersteller des Dokuments in der Ausgangsanwendung festgelegt worden sein. Der Parameter „Sicherheit“ gibt an, welche Änderungen und Funktionen im PDF-Dokument zugelassen sind. Wurde für das PDF-Dokument ein Kennwort, ein Zertifikat oder eine Sicherheitsrichtlinie eingerichtet, findet man die Methode hier aufgeführt. Ein großes Thema sind die Schriftarten. Auch hier finden sich die Informationen darüber, ob die Schriften eingebunden sind oder nicht. Richtig spannend wird es unter dem Register „Benutzerdefiniert“: Hier wurde das Einfügen von Dokumenteigenschaften ermöglicht. Es lassen sich also zusätzliche Metadaten, die mit dem eigentlichen Druckbild nichts zu tun haben, in dem PDF speichern. Hier finden sich unter „Erweitert“ eine Liste der PDF-Einstellungen, Druckvorgaben und Leseoptionen für das Dokument.

PDF-Dokumente, die mit Acrobat 5.0 oder höher erstellt wurden, enthalten Dokument-Metadaten im XML-Format. Grundsätzlich kann XML zum Beschreiben, Speichern und Austauschen von Daten genutzt werden. Die wichtigsten Vorteile von XML sind die große Verbreitung und der geringe Lernaufwand. Außerdem kann XML leicht von Menschen und Maschinen interpretiert werden. Metadaten enthalten Informationen über das Dokument und seinen Inhalt. Mittels der Extensible Metadata Platform (XMP) wird eine gemeinsame XML-Grundstruktur erzeugt, die hilft, die Erstellung, die Verarbeitung und den Austausch von Dokument-Metadaten zwischen Publishing-Arbeitsabläufen zu standardisieren.

Die Metadaten, die wir für ein Gespräch mit dem Sachbearbeiter einer Druckerei benötigen, sind Format, Umfang, Farbigkeit, Bedruckstoff und für das Angebot die Auflage. Das ist beim digitalen Kollegen genauso. Bestellen wir bei einer Druckerei in den USA, müssen wir das auf Englisch tun. Können wir kein Englisch, brauchen wir einen Übersetzer. Auch diese Analogie lässt sich auf den automatisierten Ablauf übertragen – wobei die digitalen Prozesse zusätzlich berücksichtigen, ob die Daten überhaupt verarbeitet werden können.



Abb. 4: Der Datenaustausch erfolgt über Kommunikation zwischen Systemen. Die Automatisierung im Prozess sieht hier Maschinen als Sender und Empfänger vor. Interpretationen und Eingabefehler lassen sich per Definition vermeiden. Jede Definition setzt jedoch eine Entscheidung voraus.

# XML, JSON und Struktur

Im Wesentlichen hat ein XML-Dokument die Komponenten Layout, Daten und Struktur. Adobe hat schon vor rund 20 Jahren das PDF um XML-Funktionen erweitert und damit die Grundlage geschaffen, Formulare in PDF-Dateien einzubinden. Allerdings ist XML nicht typenspezifisch. Das bedeutet, wird `height="210.00"` übergeben, ist nicht definiert, ob es hier um die Zahl 210 geht oder den Text 210.00. Im Datenformat JSON hingegen verwendet man `height:210.0` – dadurch ist `height` eindeutig als Zahl definiert, mit der man rechnen kann.

Die JavaScript Object Notation (JSON) ist ein kompaktes Datenformat in einer einfach lesbaren Textform und dient dem Zweck des Datenaustausches zwischen Anwendungen. JSON ist von der Programmiersprache unabhängig. Das ist ein Grund, warum **Radix Translate** auf dieses Format setzt. XML-Dateien werden in das JSON Format konvertiert. Die Validierung der Daten erfolgt mit JSON-Schemadateien.

## Lesbar für Mensch und Maschine

Informationen müssen unterschiedlich aufbereitet werden – z. B. wird ein Druckprodukt mit „Flyer A4, farbig gedruckt auf weißem Papier“ beschrieben. Das ist vom Menschen lesbar. Für die Automatisierung benötigen wir jedoch strukturierte Daten mit einem definierten Aufbau und festgelegter Bedeutung. Der Produktions-Workflow braucht es maschinenlesbar, wie im nachfolgenden Beispiel dargestellt:

```
{
  "dimensions": {
    "width": 297,
    "height": 210,
    "unit": "mm",
    "bleed": 3
  },
  "media": {
    "surface": "Coated",
    "paperclass": "PS1"
  }
}
```

Abb. 5: Maschinen brauchen mehr Struktur, um Informationen „lesen“ zu können.

Zusätzliche Komplexität erhält die Kommunikation durch moderne Webservices. Ein Webservice ist im weitesten Sinne eine Kommunikationsmethode zwischen zwei Anwendungen über ein Netzwerk, in der Regel das World Wide Web. Webservices gibt es in zwei unterschiedlichen Arten: Als Simple Object Access Protocol (SOAP) und als Representational State Transfer (REST). Zwischen SOAP- und REST-Webservices gibt es erhebliche Unterschiede, auf die wir hier nicht näher eingehen. Jedoch können Befürworter von REST oder SOAP leidenschaftlich für ihre Webservicearchitektur eintreten. Sowohl SOAP- als auch REST-Architekturen haben sich als zuverlässig, erfolgreich und skalierbar erwiesen. Die Entscheidung für REST oder SOAP hat daher weniger mit ihrer Effizienz zu tun als vielmehr damit, wie sich beide Ansätze in die Softwareentwicklungskultur und die Projektanforderungen eines Unternehmens einfügen. **Radix Translate** übernimmt in vielen Prozessphasen die Übersetzung und ermöglicht so die fehlerfreie Kommunikation unabhängig von Protokollen und Dateiformaten. **Radix Translate** macht Informationen für Menschen und Maschinen lesbar.

# Benefit

**Radix Translate** sorgt an den verschiedenen Schnittstellen für eine sichere Kommunikation. Diese kann direkt zwischen Systemen erfolgen, also vom Webshop hin zum ERP-System oder direkt zur Digitaldruck-Maschine oder einem anderen digitalen Frontend, wie etwa einem Plotter. Gleiches gilt für den Austausch zwischen ERP-System und den Systemen für die Produktion.

Ein weiterer Anwendungsfall ist die Übersetzung von Daten aus dem ERP-System in JDF oder xJDF. xJDF basiert, ebenso wie JDF, auf XML. Während man mit JDF noch versuchte, eine vollständig elektronische Auftragstasche mit allen dazugehörigen Arbeitsschritten zu beschreiben, liegt bei xJDF das Hauptaugenmerk auf der Beschreibung eines einzelnen Arbeitsschrittes bzw. eines Druckproduktes. Längst sind nicht alle ERP-Systeme in der Lage, einen CIP4-konformen JDF/xJDF-Export hinzubekommen. Hier zeigt sich ebenfalls das „Sprachtalent“ von **Radix Translate**.



## Kontakt ▶

### **calibrate Workflow-Consulting GmbH**

St. Pöltnerstraße 26  
A-3130 Herzogenburg  
Tel. +43 / 732 / 68 26 60  
office@calibrate.at  
[www.calibrate.at](http://www.calibrate.at)

Geschäftsführer: Peter Kleinheider