



## Efficiency in Automation – SAP AG Uses Informatica Technology to Integrate Unstructured Data in Travel Management

“The SAP Conversion Agent by Informatica is a tool that allows us to make data available for an automated process that wouldn’t otherwise be accessible. What particularly impresses me about Informatica technology is the visual development of its parsing capability. This facilitates our work. All we need to do is simply select the information from a PDF invoice that is required.”

—Rainer Krefl

E-Commerce/SAP IT Contracts & Billing  
SAP AG

SAP was founded in 1972 and today the world’s leading provider of business software offering applications and services that enable companies of all sizes and in more than 25 industries to become best-run businesses. With more than 47,800 customers (excludes customers from the acquisition of Business Objects) in over 120 countries, the company is listed on several exchanges, including the Frankfurt stock exchange and NYSE, under the symbol “SAP” (For more information, visit [www.sap.com](http://www.sap.com))

### The Challenge

The highly specialized field employees at SAP are frequently on the move; the number of trips they make each day can top several hundred on average. SAP has entered into a cooperative relationship with a travel management service provider to address its travel booking requirements. Field employees at SAP can now make all travel arrangements directly from their desks, with all charges settled via a credit card provider. To ensure that the client using the services of the field employees also incurs the travel costs, each trip is created in the internal SAP system as its own object under its own individual travel number and is assigned to a travel date and the personnel number of the individual making the trip. The travel agency creates a pro forma invoice for each trip, which was previously sent by e-mail to an internal SAP Shared Service Center as a PDF document.

“The high degree of manual effort associated with pulling out the information needed from the pro forma invoices in the Shared Service Center was one of several sources of error in the travel management process,” says Stefan Kraus, Project Manager SAP IT, SAP AG. It would take two person-days per week to pull out the information needed and resolve issues in this manual process. Issues that would arise in this manual process included lost invoices, invoices attached incorrectly, and increased workload for the travel agency because reprints were frequently requested.

“This is why we initiated a project to automate our travel management processes on a large scale,” explains Kraus. A solution was required that would automate the extraction of relevant information (including trip date, employee number, and trip number) from ticket invoices in an unstructured PDF format.

Figure 1: Here is an example of an invoice from which trip date, employee number, and trip number were once manually extracted—a process that added extra effort and errors. Today, the information is automatically extracted from a PDF file using the SAP Conversion Agent by Informatica.

**BCD travel**

SAP SYSTEMS INTEGRATION AG  
GRUENDELBACHSTRASSE 10  
71691 FREIBERG A.N.

GLOBUS REISEDIENST GMBH  
DIETMAR-HOPP-ALLEE 20  
69190 69190 WALLDORF

TELEFON 06227-748200  
TELEFAX 06227-742107

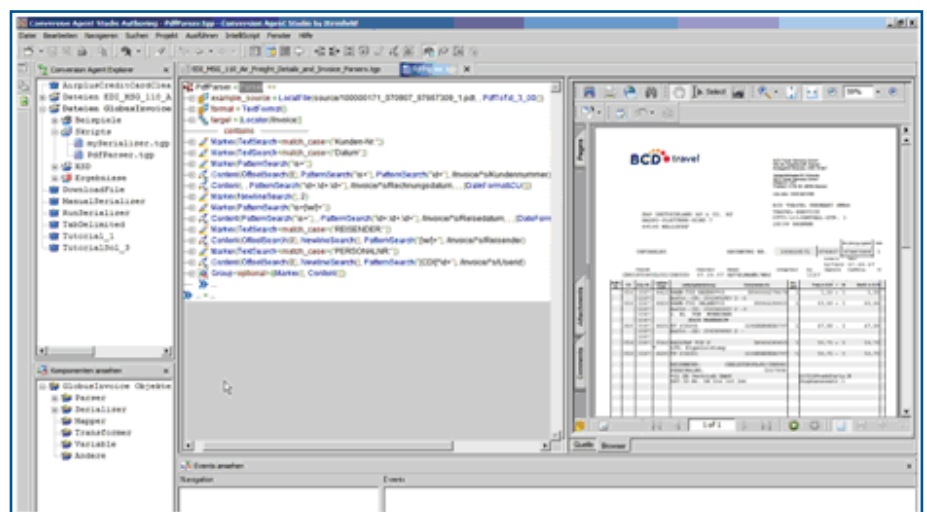
INFOBELEG RECHNUNG NR. 100003690 070928 97081448 1

Kunden-Nr. 527469 Datum 28.09.07  
Gegenkonto 3693

Reisender: RAUSCH/URSULA MRS Reise-Datum: 01.10.07 Reiseziel: RAUSCH/MRS Auftrags-Datum: 6391 Exp: 3693

Auftr.-Nr.	KS	Exp.-Nr.	Leistungs-Code	Leistungsbezeichnung	Dokumenten-Nr.	An-zahl	Preis in EUR + M	MwSt. in EUR
010	6391	9415	E	TICKET INLAND	7451771186078	1	99,17 + 1	99,17
	6391			REISENDER:	RAUSCH/URSULA MRS			
010	6391	8355	AIR PLUS (FLUG)		220XXXXXXXX3392	1	99,17 - 1	99,17
	6391			STUTTGART /BERLIN				
	6391			TKTR				
010	6391	9415	E	TICKET INLAND	2201771186079	1	135,80 + 1	135,80
	6391			REISENDER:	RAUSCH/URSULA MRS			
010	6391	8355	AIR PLUS (FLUG)		220XXXXXXXX3392	1	135,80 - 1	135,80
	6391			BERLIN /STUTTGART				
	6391			TKTR				
010	6391	3105	AIRTAF	RTX D TEL	7451771186078	1	8,33 + 1	8,33
				LTR: Eigenleistung				
010	6391	3105	AIRTAF	RTX D TEL	2201771186079	1	8,33 + 1	8,33
				LTR: Eigenleistung				
010	6391	9605	TP	262103	220XXXXXXXX3392	1	8,33 - 1	8,33
010	6391	9605	TP	262104	220XXXXXXXX3392	1	8,33 - 1	8,33
				REISENDER:	RAUSCH/URSULA MRS			
				PERSONALNR:	D001196			
				BEMERKUNGEN:	0210621243			

Figure 2: Here is the trip invoice pulled into SAP Conversion Agent. The right-hand pane contains the PDF view of the invoice. The middle pane contains the same data, automatically parsed by SAP Conversion Agent so that the desired data elements can be extracted. The addition of Informatica® technology contained in Informatica B2B Data Transformation eliminated errors and days of extra work.



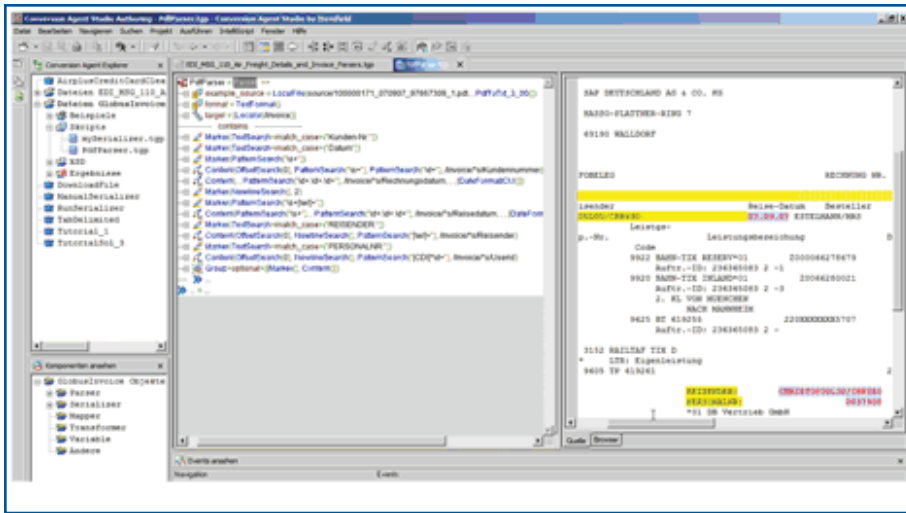


Figure 3: The process is simple and virtually error free. As shown in the right-hand pane, the user marks the trip date and employee number in SAP Conversion Agent.

```

<?xml version="1.0" encoding="UTF-8"?>
<Invoice>
  <Kundennummer>527460</Kundennummer>
  <Rechnungsdatum>20070907</Rechnungsdatum>
  <Reisedatum>20070907</Reisedatum>
  <Reisender>CHRISTOPOULOU/CHRYSO</Reisender>
  <Userid>D037908</Userid>
  <Reisnummer></Reisnummer>
</Invoice>

```

Figure 4: The result is an XML file containing only the information needed from the trip invoice that is used to assign the PDF document in the SAP system. The trip information is passed to an SAP system handling billing of the customer and expenses.

“The greatest innovation is the fact that a manual process based on e-mail was successfully converted to an automated, fully structured, and organized process.”

— Stefan Kraus, Project Manager SAP IT, SAP AG

## The Solution

Working with unstructured data—for example, PDF files, Excel spreadsheets, Word documents, presentations, and e-mails—can prove extremely difficult in a modern company. These data sources, however, represent an estimated 85 percent of all information within a company and cannot be ignored.

SAP sought to streamline its travel management by automating the extraction of invoice data delivered by the travel agency in the form of PDF documents. It adopted Informatica B2B Data Transformation™ to tackle the process. Informatica is a pioneer in the integration of unstructured data and B2B industry standard data, and Informatica B2B Data Transformation is a component of the Informatica B2B Data Exchange™ product family.

The software product enables users to access and transform a wealth of data that previously could only be viewed through time-consuming and expensive special programming. The solution can also be used to record and document metadata useful for compliance with auditing processes. Common metadata questions include: When was the data received? Where did it come from? Has the file name been altered? Has all data been received in full during integration? Without Informatica, SAP Travel Management was spending eight days per month manually pulling out the invoice information needed and resolving the issues resulting from the error-prone manual process.

SAP licensed Informatica B2B Data Transformation as part of an OEM partnership, adding it to its SAP Exchange Infrastructure (SAP XI), a component of the SAP NetWeaver platform; it is branded “SAP Conversion Agent by Informatica.” This enabled the team in charge of implementation to use the resources they already had available.

## The Results

Implementation of SAP Conversion Agent by Informatica was rapid. The travel management project began in May 2007 and went live mid-September of the same year. The implementation of automated data recording itself took slightly more than one day using the SAP Conversion Agent by Informatica. Rainer Kreft is very pleased with the ease of use offered by the tool: “All we need to do is select the information we need in the PDF invoice. In principle, access is already there.”

Automated extraction of relevant data from PDF documents was an important constituent of the overall automation concept. Today, the travel agency sends ticket invoices as PDF documents to an anonymous mailbox at SAP via SMTP, which is queried at regular intervals using a Java service that connects periodically. Through this process, the PDF documents are retrieved from the server and transferred to a pipeline server, which acts as a type of work preparation process and retains a connection to the visual archive. The pipeline server collects all documents, sorts them per the respective trips, and processes them so that the archive can integrate them and assign them to the trips appropriately. In a subsequent step, the SAP Conversion Agent analyzes the PDF documents located on the pipeline server. Content relevant to the additional steps of the process is then extracted and recreated in a structured format in an XML file. A second SAP Conversion Agent service has been set up to ensure that the archive system is also updated. This service serializes an attribute file from the XML file, which the archive server uses to assign the PDF document for the trip in the SAP system.

“The greatest innovation is the fact that a manual process based on e-mail was successfully converted to an automated, fully

structured, and organized process,” comments Stefan Kraus. “We save about three person-days a week in the Shared Service Center through the new process. It used to take two people working three days a week to find the right PDF and attach it to the appropriate record in the SAP system. Furthermore, there are now fewer errors (less than 5 percent of cases) and therefore we deliver better customer service. The previous manual process resulted in many queries since quite often PDF documents were attached to the work records in SAP. Now people monitor the process instead of laboriously addressing issues after the fact.”

“In the future we will certainly consider SAP XI with Conversion Agent a viable option for enhancements to other invoice receipt processes,” concluded Rainer Kreft.

### LEARN MORE

Learn more about B2B Data Transformation and the entire Informatica product platform. Visit us at [www.informatica.com](http://www.informatica.com) or call 800.653.3871.

### ABOUT INFORMATICA

Informatica enables organizations to gain a competitive advantage in today's global information economy by empowering them to access, integrate, and trust all their information assets. As the independent data integration leader, Informatica has a proven track record of success helping the world's leading companies leverage all their information assets to grow revenues, improve profitability, and increase customer loyalty.



Worldwide Headquarters, 100 Cardinal Way, Redwood City, CA 94063, USA  
phone: 650.385.5000 fax: 650.385.5500 toll-free in the US: 1.800.653.3871 [www.informatica.com](http://www.informatica.com)

Informatica Offices Around The Globe: Australia · Belgium · Canada · China · France · Germany · Ireland · Japan · Korea · the Netherlands · Singapore · Switzerland · United Kingdom · USA

© 2008 Informatica Corporation. All rights reserved. Printed in the U.S.A. Informatica, the Informatica logo, and Informatica B2B Data Transformation are trademarks or registered trademarks of Informatica Corporation in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks of their respective owners.

6870 (07/28/2008)