

12 (2009)
AUSGABE 2

Sonderdruck

B*online*.I.T.

Bibliothek Information Technologie

D 52614
ISSN 1435-7607

**Zeitschrift für Bibliothek, Information und Technologie
mit aktueller Internet-Präsenz: www.b-i-t-online.de**

FACHBEITRÄGE

Katastrophenplanung für
Informationseinrichtungen

Die Perfekte Bibliothek

Geschäftsmodelle für
elektronische Medien

GLOSSE

Erlesenes
von Georg Ruppelt

TRENDTHEMA

Das Ebook ist gekommen
– und es bleibt!

NACHRICHTEN

Was man über
WorldCat.org wissen sollte

Die Stuttgarter
Bibliothek 21

BAUTRENDS

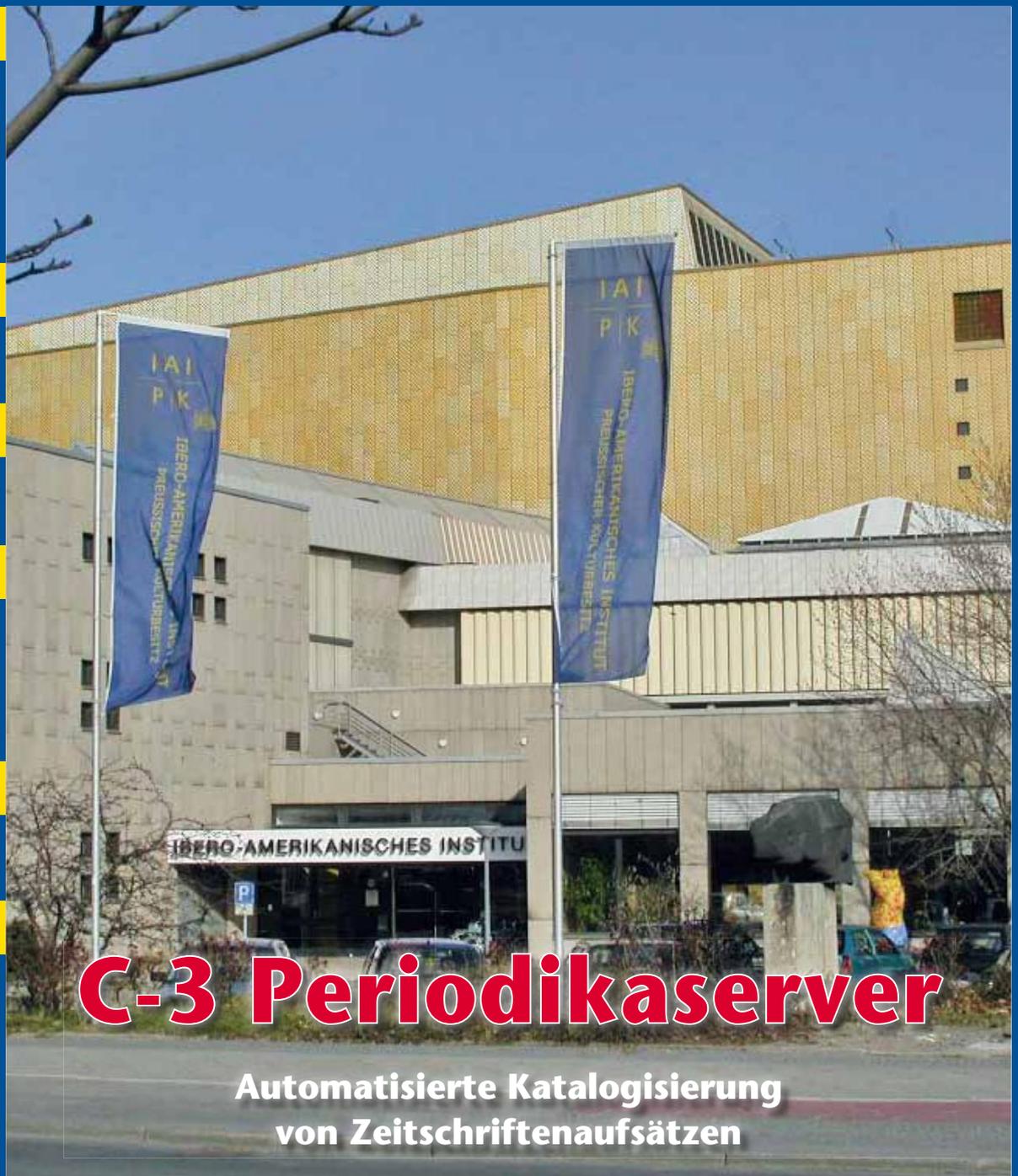
Bibliotheken planen
und bauen

REPORTAGEN

Welchen Wert hat eine
Bibliothek? – Bielefeld
Conference 2009

3. IFLA Presidential
Meeting

Die „British Architectural
Library“ in London



C-3 Periodikaserver

**Automatisierte Katalogisierung
von Zeitschriftenaufsätzen**

C-3 Periodikaserver

Automatisierte Katalogisierung von Zeitschriftenaufsätzen

Christoph Müller, Rüdiger Stratmann, Nicolai Sternitzke

Die Bibliothek des Ibero-Amerikanischen Instituts Stiftung Preussischer Kulturbesitz in Berlin (IAI) ist mit 830.000 Monografien, 33.000 Zeitschriften, 73.000 Landkarten, 38.500 Tonträgern, Videos und DVDs die größte Spezialbibliothek zu Lateinamerika, Spanien, Portugal und der Karibik in Europa und nach der Library of Congress in Washington und der Nettie-Lee-Benson-Collection der University of Texas in Austin die drittgrößte Spezialbibliothek dieser Art in der Welt. Jährlich kommen durch Kauf, Tausch und Geschenke 30.000 Monografien dazu und es werden 5.000 Zeitschriften-Abonnements geführt.



IBERO-AMERIKANISCHES INSTITUT
PREUSSISCHER KULTURBESITZ

IAI
P | K



■ Dabei sind ca. 65 % der Bestände Alleinbesitz in Deutschland. Inhaltlich liegt der Schwerpunkt der Materialien des IAI im Bereich der Geistes- und Sozialwissenschaften, der Politikwissenschaft, Geographie, Ethnologie, Altamerikanistik sowie Archäologie jeweils bezogen auf Lateinamerika, Spanien, Portugal und die Karibik. Darüber hinaus betreut die Bibliothek des IAI die DFG-Sondersammelgebiete Recht, Parlamentaria und Tageszeitungen in bzw. aus Lateinamerika sowie Latino-Studies. Die Materialien des IAI sind sowohl vor Ort, aber auch über die nationale und internationale Fernleihe sowie den Dokumentenlieferdienst subito zugänglich. Pro Jahr werden dabei über 15.000 Medieneinheiten über die verschiedenen Verteilungswege verschickt.

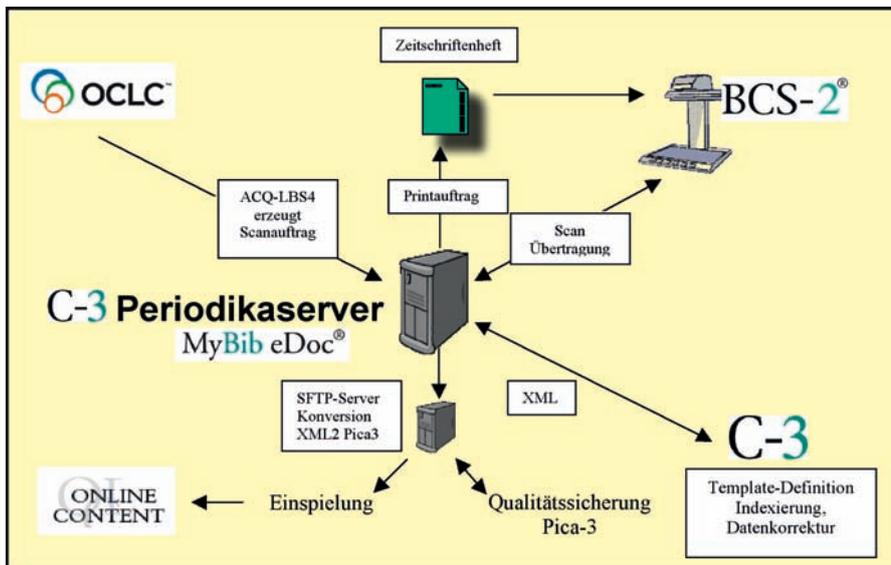
Als Mitglied des Gemeinsamen Bibliotheksverbundes (GBV) verfügt die Bibliothek des IAI über einen OCLC-Pica-Katalog, in dem

der überwiegende Teil der Bestände der Bibliothek verzeichnet ist. Verwaltet werden die Bestände des IAI mit dem lokalen Bibliothekssystem LBS4 von OCLC-Pica. Seit der Gründung des IAI im Jahr 1930 werden auch systematisch Aufsätze aus Sammelbänden und Zeitschriften erschlossen. Seit 2000 werden dazu Bilddateien von Inhaltsverzeichnissen ausgewählter Zeitschriften in einer nach Zeitschriftentitel, Erscheinungsort, Erscheinungsjahr, Schlagwörtern, Heftnummer und Signatur durchsuchbaren Current-Contents-Datenbank angeboten.

Um die Nachweissituation von Zeitschriftenaufsätzen und gleichzeitig das bibliografische Informationsangebot im Sinne einer Spezialbibliothek zu verbessern, wurde im Jahr 2006 beschlossen, dem System der Online-Contents-Sondersammelgebietsauschnitt (OLC-SSG) des GBV beizutreten. Dazu wurde nach einer technischen Lösung

gesucht, die es gewährleisten sollte, die Inhaltsverzeichnisse von Zeitschriften automatisch formal zu erschließen.

Als einziges halbautomatisches Programm kam zum damaligen Zeitpunkt das in verschiedenen Sondersammelgebietsbibliotheken, seinerzeit an der SUB Göttingen entwickelte Current-Contents-Tool (CC-Tool) zum Einsatz. Die mit der OCR-Software Omnipage bearbeiteten Scans der Inhaltsverzeichnisse werden bei diesem Verfahren als ASCII-Daten in einem Texteditor ausgegeben und die Titel-, Autoren- und Seitenangaben mit erheblichem Zeitaufwand manuell markiert. Das Ergebnis wird dabei in einem nachgeschalteten Prozess vom CC-Tool ins PICA3-Format umgewandelt. Dieses Vorgehen erschien jedoch von Anfang an als ungeeignet, da die notwendige umfangreiche redaktionelle Arbeit nicht den Anforderungen an ein angestrebtes Massenverfahren entsprach. Außerdem



bildeten die Schwierigkeiten der Bearbeitung von nicht deutschsprachigen Inhaltsverzeichnissen mit CC-Tool – Sonderzeichen wie Tilden und Akzente, so wie sie im Spanischen und Portugiesischen verwendet werden, müssen jeweils ebenfalls per Hand nachgetragen werden – ein Ausschlusskriterium für den Einsatz der Software in der Bibliothek des IAI.

Eine mögliche Lösung für diese erweiterten Anforderungen stellte die Software C-3 dar, die sich in 2006 noch im Entwicklungsstadium befand. Die von der Bonner Firma ImageWare Components GmbH entwickelte modulare Software sollte die automatische Erkennung der Titel-, Autoren- und Seitenzahlinformationen in den gescannten Inhaltsverzeichnissen sowie deren automatische Indexierung und Konversion in Katalogisate ermöglichen. Als Pilotbibliotheken beteiligten sich die Niedersächsische Staats- und Universitätsbibliothek Göttingen und die Deutsche Zentralbibliothek für Medizin Köln an der Weiterentwicklung der C-3 Programmsuite und lieferten die bibliotheksspezifischen Anforderungskriterien zur Anpassung der Software an die Erschließungsfunktionalitäten und an den bibliothekarischen Geschäftsgang. Da zum damaligen Zeitpunkt die OCR-Erkennung fremdsprachiger Texte und die Erkennung der Verzeichnisstrukturen systemseitig noch nicht vollständig ausgereift waren, beschloss man im IAI die weitere Entwicklung des Systems abzuwarten und solange den bestehenden Current-Content-Dienst weiter zu führen.

Im Herbst 2007, als sich der C-3 Produktionsbetrieb an der SUB Göttingen etabliert hatte und das C-3 Modul um wesentliche Zusatzfunktionen optimiert worden war, beschloss das IAI, die Software für die Erschließung der eigenen Zeitschriftenbestände zu verwenden und das Hostingangebot „C-3 Periodikaserver“ der Verbund-

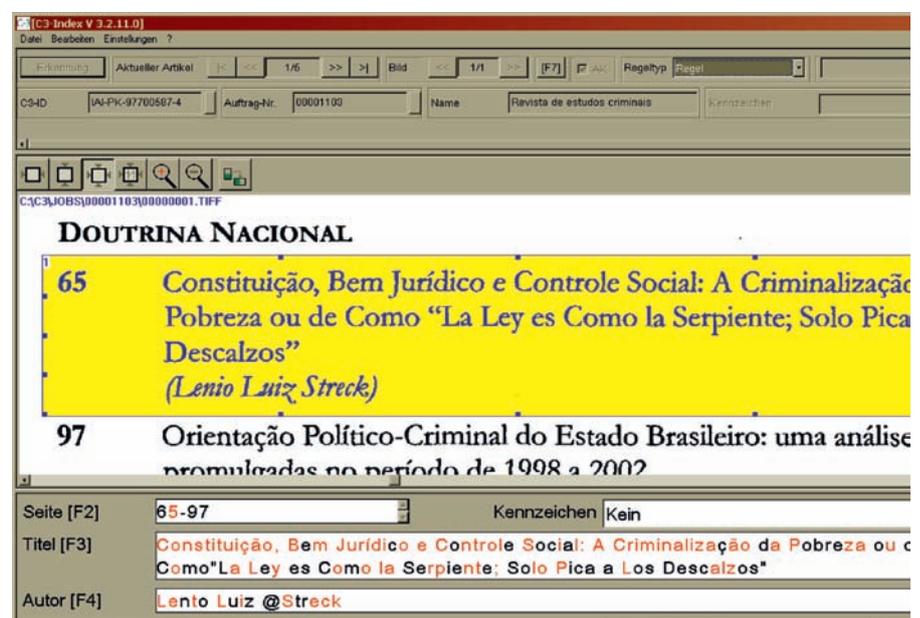
zentrale des GBV in Anspruch zu nehmen. Zu den Komponenten des Gesamtworkflows gehören neben der C-3 Erschließungssoftware der Scan-Client BCS-2®, über den die zu indexierenden Zeitschrifteninhaltsverzeichnisse gescannt werden und ein von der Verbundzentrale des GBV gehostetes MyBib eDoc® System, das mit seinen Schnittstellen zu allen Workflowkomponenten eine webbasierte mehrplatzfähige Auftragsverwaltung bietet und die Arbeitsprozesse im Batchverfahren steuert. In der Zwischenzeit konnten diakritische Sonderzeichen, mehrsprachige Inhaltsverzeichnisse und auch komplex strukturierte Inhaltsverzeichnisse problemlos verarbeitet werden, sodass ein vollautomatisches Indizieren möglich war.

Bei der ImageWare Components GmbH und der Göttinger Verbundzentrale des GBV (VZG) wurde die Zusammenstellung der Hard- und Softwarekomponenten für zwei Auftragserzeugungsarbeitsplätze, eine Scanstation, zwei C-3 Indexierungsarbeits-

plätze und drei SFTP-Zugriffsmöglichkeiten für die Qualitätssicherung der Artikeldaten in Auftrag gegeben.

Hardwareseitig wurde ein WideTekA3-Scanner angeschafft, der mit der für das System speziell konfigurierten BCS2-Scan-Software in kürzester Zeit die Inhaltsverzeichnisse scannt. Ansonsten kommen normale Standard-PCs zum Einsatz, über die auf die Auftragsverwaltung des zentralen C-3 Periodikaservers per Browser zugegriffen wird.

An den Indexierungsarbeitsplätzen wird die C-3 Software, bestehend aus den Programmmodulen C-3 Template und C-3 Index, verwendet. In C-3 Template werden einmalig für jeden Zeitschrifttitel die Struktur der Inhaltsverzeichnisse (Interpretationstyp Regel, Tabelle oder Freiform), die Abfolge von Aufsatztitel, Autoren und Seitenzahlen sowie die Schriftattribute (fett, kursiv etc.) der bibliografischen Daten festgelegt. Auf Basis dieser Templatedefinition werden die Scans der jeweiligen Inhaltsverzeichnisse im C-3 Index Modul mit der OCR-Software Abbyy Fine Reader so erkannt, dass die bibliografischen Informationen der einzelnen Aufsätze bereits kategorisiert und aufsatzweise separiert ausgegeben werden. Die dabei erzeugten Daten können, sofern erforderlich, mit komfortablen C-3 Nachbearbeitungstools korrigiert und im Anschluss im XML-Format an den Periodikaserver exportiert werden. Automatische Konversionsroutinen erzeugen aus den xml-Dateien serverseitig Artikeldaten im Pica3-Format, die nach einer abschließenden Qualitätskontrolle in die Online Contents Datenbanken des GBV eingespielt werden. Um die Aufsätze der jeweiligen Zeitschrift eindeutig zuordnen zu können, werden die Titelaufnahmen der Zeitschriften, die Templates und später auch die jeweiligen Aufsatztiteldaten über eine eindeutige Ident-Num-



mer, die C-3 ID, verknüpft. Dazu bietet sich neben der Pica-Produktionsnummer auch die so genannte Swets-Nummer an. Das OLC-SSG-System des GBV beruht auf der Swets-Datenbank Online Contents, in der alle von Swets gelieferten Aufsatzdaten verzeichnet und recherchierbar sind.

Die Steuerung des Gesamt-Workflows erfolgt über den MyBib eDoc®-basierten C-3 Periodikaserver, über den die Mitarbeiter jederzeit den Status des Bearbeitungsprozesses verfolgen können. Darüber hinaus wird auf dem Periodikasytem der Indexierungsauftrag generiert und mit den Metadaten des Auftrages, dem gescannten Inhaltsverzeichnis und der xml-Exportdatei verknüpft. Damit werden alle zum Auftrag gehörenden Daten übersichtlich und nachvollziehbar verwaltet und können bei Bedarf von den Mitarbeitern editiert werden.

Eine besondere Schwierigkeit stellte bei der Implementierung des Geschäftsgangs in

bearbeiteten Zeitschriften einer inhaltlichen und formalen Revision unterzogen. Dabei wurden die inhaltliche Relevanz der Zeitschriften, der Status des jeweiligen Abonnements und die formale Eignung der Inhaltsverzeichnisse für die Erschließung mit C-3 geprüft. Die Liste der mit dem neuen System zu bearbeitenden Zeitschriften umfasste am Ende dieses Prozesses ca. 750 Titel.

Zusätzlich zu diesen 750 Zeitschriften, die sukzessive durch neue bzw. erneuerte Abonnements ergänzt werden, wurden noch ca. 150 Titel identifiziert, die von anderen Bibliotheken bearbeitet werden, aber inhaltlich zu den Zeitschriften der Bibliothek des IAI passen. Auf Basis dieser 900 Zeitschriftentitel wurde dann der OLC-Sondersammelgebietsausschnitt *Ibero-Amerika* generiert, der über den GVK+ des GBV zugänglich ist.

Nach Aufbau und Konsolidierung der technischen Infrastruktur und nach Auswahl der zu bearbeitenden Zeitschriften wurden

den Mitarbeiterinnen und Mitarbeitern des IAI die erforderlichen Kenntnisse zur Bedienung der Softwarekomponenten im Rahmen einer zweitägigen vor Ort Schulung vermittelt. Im direkten Anschluss ging das IAI im Mai 2008, ca. 6 Monate nach der Entscheidung für das C-3 Periodikasytem, in den Produktionsbetrieb. Monatlich entstehen seitdem ca. 1.000 Aufsatztiteldatensätze, die in den GVK+ eingespielt werden.

Die Erkennungsqualität – auch bei komplexen, mehrsprachigen Inhaltsverzeichnissen – ist dabei so hoch, dass Qualitätskontrollen und Datenkorrekturen nur in kleinem Rahmen vorgenommen werden müssen. Insgesamt ist der seitens des IAI investierte Arbeitsaufwand für die Bearbeitung der aktuellen Zeitschriftenhefte durch die von C-3 gebotenen automatischen Indexierungsfunktionen geringer als erwartet. So erfolgen die Auftragserzeugung im LBS4 und die Qualitätskontrolle im laufenden Betrieb (Akzessionierung und Katalogbereinigung). Für das Scannen kommen ca. 0,2 VZÄ und bei der Indexierung ein VZÄ zum Einsatz. Die Sicherung der Datenqualität wird am IAI von einem Diplom-Bibliothekar übernommen, alle anderen Arbeitsschritte werden von Assistenten bzw. Fachangestellten für Medien- und Informationsdienste ausgeführt.

Aufgrund des störungsfreien Ablaufs der Datenproduktion und des vergleichsweise geringen personellen sowie zeitlichen Aufwandes, wurde im Januar 2009 begonnen,

die vom IAI für den SSG-OLC Ibero-Amerika bearbeiteten Zeitschriften auch retrospektiv zu bearbeiten. In einem ersten Schritt werden dabei die Bestände ab dem Jahr 2000 erschlossen, dem Jahr, in dem im IAI die systematische Katalogisierung von Zeitschriftenaufsätze eingestellt und durch den Current-Contents-Dienst abgelöst wurde. In einem zweiten Schritt sollen dann auch alle weiter zurückliegenden Jahrgänge der für Online Contents ausgewählten Zeitschriften bearbeitet werden. Bis Ende März 2009 konnten dabei schon zusätzlich zum normalen Betrieb 850 Inhaltsverzeichnisse älterer Zeitschriftenhefte bearbeitet werden.

Zum einen wäre angesichts der vergleichsweise hohen Anzahl unterschiedlicher Softwarekomponenten aus Nutzersicht eine Zusammenführung unter einer einheitlichen Oberfläche und in ein integriertes Programm wünschenswert, zum anderen wird durch den modularen Aufbau die Möglichkeit geboten, die Systemkomponenten in die lokalen Geschäftsgänge der Bibliothek des Ibero-Amerikanischen Instituts zu integrieren. Insgesamt hat sich das System als sehr effizient erwiesen. Der Produktionsbetrieb verläuft störungsfrei und die Bedienung der Software ist komfortabel. Durch den hohen Automatisierungsgrad wird die Katalogisierung von Zeitschriftenaufsätze für den SSG Online Contents-Dienst sehr erleichtert und beschleunigt.

■ AUTOREN

DR. CHRISTOPH MÜLLER, RÜDIGER STRATMANN,

Ibero-American Institute
Prussian Cultural Heritage
Foundation
Potsdamer Straße 37
D 10785 Berlin
Tel.: +49 (0) 30 / 266 0
Fax: +49 (0) 30 / 266-2503
info@iai.spk-berlin.de
www.iai.spk-berlin.de



NICOLAI STERNITZKE

ImageWare
Components GmbH
Am Hofgarten 20
53113 Bonn
info@imageware.de



ImageWare

ImageWare Components GmbH
Am Hofgarten 20
D 53113 Bonn
Tel.: +49 (0) 228 / 969 85-0
Fax: +49 (0) 228 / 969 85-84
info@imageware.de
www.imageware.de



der Bibliothek des IAI die automatische Auftragserzeugung beim Einchecken der Zeitschriftenhefte im Erwerbungsmodul von LBS4 dar. Das Programm, das seitens der Verbundzentrale des GBV für die Auftragserzeugung entwickelt wurde, war zunächst nur mit LBS3 – der Vorgängerversion des im IAI verwendeten Lokalen Bibliothekssystems – kompatibel. Somit waren in diesem Zusammenhang noch weitere Entwicklungen seitens der VZG und Anpassungen der Schnittstelle durch die ImageWare Components GmbH erforderlich, die aber bereits im Frühjahr 2008 erfolgreich abgeschlossen waren. Mit dem neuen Zusatzprogramm werden die Daten eines durch LBS4 erzeugten Belegzettels ausgelesen und in eine ILL-subito konforme Bestellmail umgewandelt, die anschließend auf dem C-3 Periodikaserver einen Indexierungsauftrag erzeugt.

Neben der technischen Implementierung, die von Ende Dezember 2007 bis Ende Februar 2008 dauerte, waren auch inhaltliche Vorbereitungen nötig. So wurde die Liste der im Current-Contents-Dienst des IAI