

Index Theologicus – neue Produktionsverfahren bei der Bibliographieerstellung

Martin Faßnacht und Winfried Gebhard

Wissenschaftliche Bibliographien, die Forschende, Studierende und sonstige Nutzer zuverlässig und umfassend über neu erschienene Literatur informieren sollen, müssen möglichst alle relevanten Zeitschriften aus dem jeweiligen Wissenschaftsgebiet auswerten. Die Bibliographie Index Theologicus (IxTheo) widmet sich dieser Aufgabe seit vielen Jahren.¹ Allerdings war die Zahl der ausgewerteten Zeitschriften auf etwas mehr als 600 Zeitschriften begrenzt, da die traditionelle Produktionsweise – das Abtippen der Metadaten – personalintensiv und deswegen limitiert war. Eine Ausweitung der Auswertung auf möglichst alle relevanten Zeitschriften kann nur mit neuen Produktionsverfahren und auf die jeweilige Medienart der Zeitschrift zugeschnittenen Methoden erreicht werden.

Die Wahl des Produktionsverfahrens hängt davon ab, ob die Metadaten der Aufsätze aus Zeitschriften elektronisch oder gedruckt vorliegen. Für elektronisch vorliegende Metadaten wurde das Literaturverwaltungsprogramm Zotero so weiterentwickelt, dass die Metadaten in einem halbautomatischen Verfahren aus dem Internet übernommen und direkt in das im Südwestdeutschen Bibliotheksverbund (SWB) genutzte Pica-Format transformiert werden können.² Für Zeitschriften, die traditioneller Weise nur gedruckt verlegt werden, kann dieses Verfahren in der Regel nicht angewandt werden. Von den etwa 1.347 Zeitschriften, die zukünftig im IxTheo ausgewertet werden, ist knapp die Hälfte der Zeitschriften nur gedruckt verlegt. Um auch diese große Zahl an Zeitschriften auf Aufsatzebene effizient auswerten zu können, haben wir uns für ein halbautomatisches Verfahren entschieden, das die grundlegenden Informationen aus dem Inhaltsverzeichnis (IHV) eines Zeitschriftenheftes strukturiert auslesen kann.

Ein solches Verfahren bietet seit vielen Jahren die Firma ImageWare Components GmbH aus Bonn an. Auf einer Informationsreise zum Ibero-Amerikanischen Institut in Berlin (IAI), das die C-3 Software von

ImageWare seit 2007 einsetzt, konnten wir uns das Verfahren zeigen lassen, Vorteile und Nachteile kennenlernen und praktische Fragen klären.³

Die UB Tübingen hat sich für eine Einzelplatzlösung entschieden. Neben den finanziellen Vorteilen bot die Einzelplatzlösung für uns die Möglichkeit, dass wir das C-3 Verfahren schnell in einem speziellen Projekt zum Einsatz bringen konnten.

Weiterentwicklung der C-3 Software

Als notwendig erwies sich nach der Implementierung, dass das ImageWare inhärente Datenschema an die SWB Bedürfnisse angepasst werden musste. Im Workflow des IAI wird die Transformation der Metadaten aus dem ImageWare eigenen XML in entsprechende PICA-Felder bei der Einspielung in den GBV bewirkt. Die UB Tübingen wünschte aber eine davon unabhängige und direkt in der C-3-Software vorgenommene Datentransformation, bevor die Metadaten an den SWB-Verbund übergeben werden. Deswegen wurde von den Metadatenspezialisten der UB Tübingen eine Transformationsvorgabe erarbeitet und von der Firma ImageWare in der Software implementiert. Neben den Standardangaben für einen Aufsatz konnten so auch spezielle Abrufzeichen und Kennziffern bei der Ausgabe der Metadaten in Marc21XML berücksichtigt werden (*Abb. 1*). Dazu gehören beispielsweise das Bibliothekssigel und das SSG-Kennzeichen.

Eine Eigenart der C-3-Software war die Gleichsetzung der Schlussseitenzahl eines Aufsatzes mit der Anfangsseitenzahl des darauffolgenden Aufsatzes. In nahezu allen Zeitschriften, die im IxTheo ausgewertet werden, beginnt jedoch ein neuer Aufsatz auf einer neuen Seite, sodass die Schlussseitenzahl des vorgehenden Aufsatzes um die Zahl 1 reduziert ist.⁴ Dies wurde auf Wunsch der UB Tübingen nun als Default-Einstellung übernommen. Wünschenswert wäre, wenn pro Zeitschrift im Template der jeweiligen

1 <https://ixtheo-test.de>

2 Timotheus Chang-Whae Kim, Philipp Zumstein, Semiautomatische Katalogisierung und Normdatenverknüpfung mit Zotero im Index Theologicus. URN: urn:nbn:de:kobv:11-100238157 URL: <http://libreas.eu/ausgabe29/05kim/>

3 Erfahrungsbericht des IAI: http://www.imageware.de/wp-content/uploads/2015/03/IAL_Automatisierte-Katalogisierung-von-ZeitschriftenaufsätzenDT.pdf

4 Die Schlussseitenzahl eines Aufsatzes wird aus der Anfangsseitenzahl des nächsten Aufsatzes errechnet. Deswegen muss die Schlussseitenzahl des allerletzten Aufsatzes eines Heftes / Bandes immer manuell eingetragen werden, da es keine weitere Anfangsseitenzahl mehr gibt.

```
<?xml version="1.0" encoding="UTF-8"?>
<collection xmlns="http://www.loc.gov/MARC21/slim">
  <record id="1" type="Bibliographic">
    <leader> naa a22 4500</leader>
    <controlfield tag="003">DE-21</controlfield>
    <controlfield tag="007">tu</controlfield>
    <datafield tag="041" ind1="0" ind2="">
      <subfield code="a">fre</subfield>
    </datafield>
    <datafield tag="084" ind1="" ind2="">
      <subfield code="a">1</subfield>
      <subfield code="2">ssgn</subfield>
    </datafield>
    <datafield tag="100" ind1="1" ind2="">
      <subfield code="a">Siegwalt, Gérard</subfield>
    </datafield>
    <datafield tag="245" ind1="1" ind2="0">
      <subfield code="a">La crise écologique et l'actualité
    </datafield>
    <datafield tag="260" ind1="" ind2="">
      <subfield code="c">2016</subfield>
    </datafield>
    <datafield tag="773" ind1="0" ind2="">
      <subfield code="i">In</subfield>
      <subfield code="w">(DE-576)015182495</subfield>
    </datafield>
    <datafield tag="852" ind1="" ind2="">
      <subfield code="a">DE-Tue135</subfield>
    </datafield>
    <datafield tag="935" ind1="" ind2="">
      <subfield code="a">mteo</subfield>
    </datafield>
    <datafield tag="935" ind1="" ind2="">
      <subfield code="a">imwa</subfield>
    </datafield>
    <datafield tag="935" ind1="" ind2="">
      <subfield code="a">ixzs</subfield>
    </datafield>
    <datafield tag="936" ind1="u" ind2="w">
      <subfield code="d">64</subfield>
      <subfield code="j">2016</subfield>
      <subfield code="e">3</subfield>
      <subfield code="h">215-228</subfield>
    </datafield>
  </record>
</collection>
```

Sommaire

Gérard SIEGWALT

La crise écologique et l'actualité du témoignage biblique
concernant la création 215

Marc LIENHARD

Réformes et Réformation au XVI^e siècle 229

Matthieu ARNOLD

Considérations actuelles sur les Quatre-vingt-quinze
thèses (1517) Prélude à l'« année Luther » 249

Anne-Cathy GRABER

Marie. Une lecture comparée de Redemptoris Mater
(Jean-Paul II) et du Commentaire du Magnificat (Luther)
à la lumière des dialogues œcuméniques 263

Matthieu ARNOLD

Quelques ouvrages récents relatifs à Martin Luther et à la
Réformation (XXIV) 279

Matthieu ARNOLD

Chronique Sur les traces de Martin Luther : Orte der
Reformation (VIII) 289

Abbildung 1 und Abbildung 2

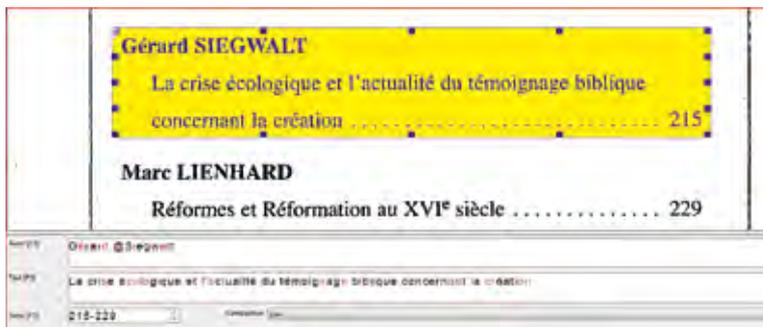


Abbildung 3

Zeitschrift eine Wahlmöglichkeit für die Erstellung der Schlussseitenzahl möglich wäre.

Workflow

Das Verfahren basiert auf den eingescannten Inhaltsverzeichnissen (IHV) der auszuwertenden Zeitschriftenhefte (Abb. 2). Diese werden nicht nur OCR erkannt, sondern in Strukturen übertragen. Auf diese Weise können die OCR erkannten Strings als Name des Autors/der Autoren, Titel und Seitenzahl richtig zugeordnet werden. Das ist die Voraussetzung, um diese Daten in die Verbunddatenbank und von dort aus in den IxTheo übernehmen zu können.

Die IHV werden nach Eingang der Hefte direkt in der Zeitschriftenabteilung der UB gescannt und auf einen für alle Mitarbeiter zugänglichen Server abgelegt. Das Scannen der IHV ist von den späteren Verarbeitungsschritten modular getrennt und ermöglicht auf diese Weise, die in einem zweischichtigen Bibliothekssystem

an unterschiedlichen Stellen eingehenden Hefte einbeziehen zu können. Es spielt also keine Rolle, ob eine Zeitschrift in der UB oder in einer Institutsbibliothek abonniert wird. Vor Ort bedarf es nur der Absprache der beteiligten Bibliotheken und eines einfach zu bedienenden Flachbettscanners.

Der nächste Schritt ist die Verarbeitung der gescannten IHV und die Umsetzung in eine strukturierte Form. Auf einem mit der C-3-Software ausgestatteten Computer werden die IHV als Bilddatei geöffnet, wobei in einem zweiten Fenster die OCR erkannten Strings in die entsprechenden Kategorien Titel, Autor und Seitenzahl umgesetzt werden. Farbliche Hervorhebungen möglicher OCR Fehler erleichtern die Überprüfung und Korrektur (Abb.3). Ebenso können hier direkte Manipulationen an den Strukturen vorgenommen werden:

- das Eintragen eines Semikolons innerhalb des Titel-Strings trennt im Ausgabeformat Haupttitel vom Untertitel
- das automatisch gesetzte Sortierzeichen @ innerhalb des Autor-Strings kann manuell verschoben werden
- die automatisch errechnete Schlussseitenzahl eines Aufsatzes kann manuell angepasst werden.

Templates

Da Inhaltsverzeichnisse sehr unterschiedlich gestaltet sein können, muss für jede Zeitschrift ein eigenes

Template erarbeitet werden, in dem die Auswertungseigenschaften des jeweiligen IHV einer Zeitschrift hinterlegt werden. Hier kann nicht nur das grundlegende Design eines IHV bestimmt werden, sondern in User-Fields gleichbleibende Informationen, die in die Metadaten aller Aufsätze einer Zeitschrift übernommen werden sollen, festgelegt werden. Die UB Tübingen legt hier die Sprachcodes einer Zeitschrift, die Zuordnung zum Sondersammelgebiet (SSG) und lokale Abrufzeichen fest.

Über die ebenfalls im Template hinterlegte Pica-Production-Number (PPN) der Gesamtaufnahme einer Zeitschrift im SWB werden die einzelnen Aufsätze bei der Metadatentransformation der entsprechenden Zeitschrift richtig zugeordnet.

Vorteile / Nachteile

Der größte Vorteil der C-3 Software gegenüber der traditionellen Produktionsweise, in der die Metadaten abgetippt wurden, liegt in der Schnelligkeit und damit der zeitlichen Einsparung. Da sowohl die OCR-Erkennung als auch die Umsetzung in die entsprechenden Strukturen zuverlässig funktionieren, müssen nur wenige Korrekturen manuell ausgeführt werden.

Der größte Nachteil dieses Verfahrens liegt im beschränkten Informationsgehalt eines IHV. Auch die beste Software kann nur Informationen entnehmen, die im IHV abgedruckt sind. Deswegen fehlen bei diesem Verfahren Abstracts, sacherschließende Elemente wie Keywords und weitergehende Autoreninformationen, da diese Informationen in Zeitschriften an anderen Stellen abgedruckt sind. IHV können darüber hinaus jedoch auch noch Kurztitel oder abgekürzte Vornamen enthalten. Die Metadaten sind deswegen so gut wie ein IHV vollständig ist.

Die in den Verbund eingespielten IxTheo-Metadaten werden deswegen in der Regel im SWB nachbearbeitet. Geprüft wird, ob die im Template hinterlegte Sprache, die für alle Titel einer Zeitschrift übernommen wird, für den einzelnen Aufsatz richtig ist. Für Kernzeitschriften wird nachträglich noch die Autorenverknüpfung zur Gemeinsamen Normdatei (GND) hergestellt.⁵

Problemanzeige

Problematische Inhaltsverzeichnisse

Die Einstellungsmöglichkeiten beim Anlegen eines

Templates berücksichtigen sehr viele unterschiedliche Anordnungsmöglichkeiten in einem IHV. Bei den meisten der in das C-3 Verfahren zu überführenden IxTheo-Zeitschriften führt das zu korrekt zugeordneten Daten, in den restlichen Fällen ist weiterhin eine manuelle Erfassung der Aufsatzdaten erforderlich. Die Gründe dafür sind sehr unterschiedlich:

- graphisch gestaltete IHV

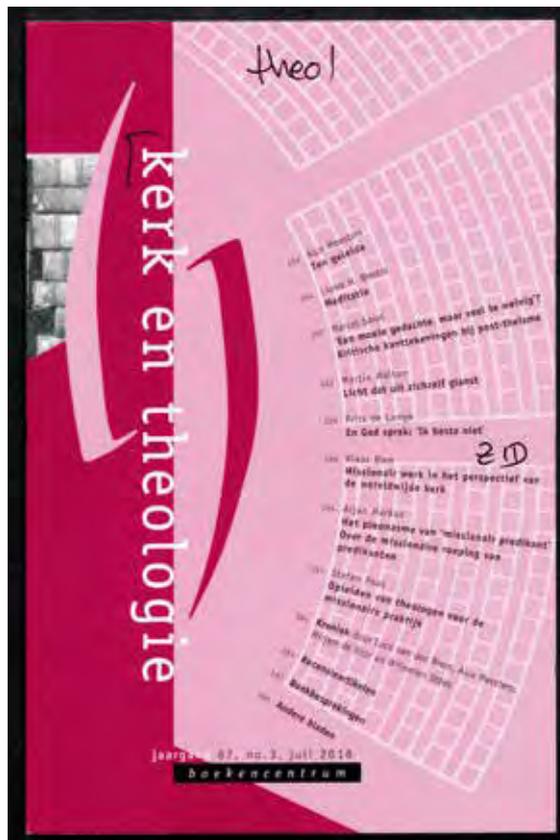


Abbildung 4

- farbliche IHV, die dazu führen, dass der Text kontrastarm ist (z.B. gelber Text auf braunem Hintergrund)
- IHV, in denen die Titel nur gekürzt sind oder Seitenzahlen fehlen
- Titel in Großbuchstaben⁶
- IHV hat fremdsprachige Paralleltitel

Bibliothekarische Zuordnungsvorgaben

In der Verbunddatenbank des SWB werden die Aufsätze aus Zeitschriften – anders als man das vermuten könnte – nicht hierarchisch zugeordnet:

Gesamtaufnahme der Zeitschrift

↳ Heft / Band der Zeitschrift

↳ Aufsatz innerhalb eines Heftes / Bandes

⁵ Obwohl einzelne Zeitschriften – wie die Theologische Quartalschrift – die GND der Autoren mittlerweile im Heft abdrucken, sind diese Informationen nicht Bestandteil des IHV und müssen manuell nachgetragen werden. Zur Problematik fehlender Autoren-Identifikationsnummern in Zeitschriften siehe den Aufsatz von Martin Faßnacht, „Druck die GND Nummer in der Publikation ab! Vereinfachung der Normdatenverwendung in Bibliotheken und Datenbanken“ (<http://dx.doi.org/10.5282/o-bib/2014H1S192-196>)

⁶ Es kommt oft vor, dass Autorennamen in Großbuchstaben abgedruckt werden. In diesen Fällen wandelt die C-3 Software die Großbuchstaben mit Ausnahme des ersten Buchstabens in Kleinbuchstaben um. Es wäre wünschenswert, diese Funktionalität in der C-3 Software auch auf Wörter des Titels anzuwenden.

Die Hefte / Bände sowie die Aufsätze hängen vielmehr direkt an der Gesamtaufnahme der Zeitschrift:
Gesamtaufnahme der Zeitschrift

↳ Heft / Band der Zeitschrift

↳ Aufsatz innerhalb eines Heftes / Bandes

Dieses Verfahren hat seine Gründe: die meisten Bibliotheken katalogisieren keine Hefte von Zeitschriften, da der Informationsgehalt eines solchen Katalogisates begrenzt ist. Dazu kommt, dass die einzelnen Hefte später in der Regel zu Bänden gebunden werden und gar nicht als physische Einheit weiter existieren. Im C-3 Verfahren hat das zum Vorteil, dass im Template einmalig die PPN einer Zeitschrift hinterlegt werden muss, die Aufsätze bei der Bearbeitung in C-3 aber immer richtig zugeordnet werden.

Anders verhält es sich jedoch bei zeitschriftenartigen Reihen wie z.B. Jahrbüchern, bei denen die Aufsätze in der Verbunddatenbank nicht an die Gesamtaufnahme der Reihe, sondern an den jeweiligen Band gehängt werden. Hier könnte die richtige Zuordnung nur durch eine ständige Änderung der PPN-Nummer im Template erfolgen, was nicht praktikabel ist. Deswegen wäre es wünschenswert, wenn bei der Auftragserstellung innerhalb der C-3 Software, bei der auch jetzt schon die jeweiligen Heft-, Band- und Jahresangaben erfasst werden müssen, auch die Eingabe der PPN des jeweiligen Bandsatzes möglich wäre, um die bibliothekarischen Zuordnungsvorgaben umsetzen zu können.

Fazit

Die Einführung des C-3 Scanverfahrens für die Erstellung der IxTheo-Bibliographie führt zu einer schnelleren Erfassung der grundlegenden Aufsatzmetadaten: Vor- und Nachname des Autors / der Autoren, Titel / Untertitel des Aufsatzes, Seitenzahlen sowie die Zuordnung zur Gesamtaufnahme der Zeitschrift. Da bei jeder einzelnen Zeitschrift geprüft wird, ob das Zotero- oder das C-3 Verfahren zur Bearbeitung geeig-

net ist, dauert die Überführung der 1.349 Zeitschriften in eines der beiden Verfahren noch an. Momentan sind von den etwa 600 in Frage kommenden Print-Zeitschriften 160 in das C-3 Verfahren übernommen worden. Eine abschließende Aussage zum Effizienzgewinn ist zum jetzigen Zeitpunkt noch nicht möglich. Die Erfassung der Metadaten durch das C-3 Scanverfahren und das damit ermöglichte automatisierte Anlegen von Lokaldaten in der Verbunddatenbank beschleunigen aber signifikant den Arbeitsablauf im Vergleich zur bisherigen manuellen Eingabe. Die Effizienzgewinne können jedoch erst im Routinebetrieb erzielt werden. Zuvor müssen die hausinternen Geschäftsgänge etabliert werden, die Mitarbeiterinnen und Mitarbeiter die nötige Routine ausbilden sowie die Templates neu angelegt werden.

Ohne die neuen Produktionsverfahren wäre aber eine Verdoppelung der im IxTheo ausgewerteten Zeitschriften nicht möglich. |



Dr. Martin Faßnacht

Leiter des FID Theologie / Index Theologicus
martin.fassnacht@uni-tuebingen.de
GND: 122054482



Winfried Gebhard

Dipl.-Bibl. (FH)
w.gebhard@uni-tuebingen.de
GND: 13773039X

Eberhard Karls Universität Tübingen
Universitätsbibliothek
Wilhelmstr. 32
72074 Tübingen

Täglich News auf www.b-i-t-online.de