

Bericht über den MyCoRe-Workshop auf dem 106. Bibliothekartag

Wiebke Oeltjen, Universität Hamburg, MyCoRe-Geschäftsstelle
Matthias Eichner, Thüringer Universitäts- und Landesbibliothek Jena
Steffen Hankiewicz, intranda GmbH, Göttingen
Frank Lützenkirchen, Universitätsbibliothek Duisburg-Essen
Kathleen Neumann, Verbundzentrale des GBV (VZG), Göttingen
Ulrike Stahl, Julius Kühn-Institut (JKI), Quedlinburg
Robert Stephan, Universitätsbibliothek Rostock

1. Einleitung

MyCoRe¹ ist eine Open-Source-Software, deren Name ein Akronym aus „My Content Repository“ ist.² Sie dient als Software-Framework zum Erstellen von Repository-Webanwendungen, wie z.B. digitalen Archiven, Onlinekatalogen, Bilddatenbanken, Dokumenten- oder Zeitschriftenservern. Die Software stellt Funktionen zum Erschließen, Präsentieren, Recherchieren, Veröffentlichen und Administrieren digitaler Objekte bereit. Erfasst werden neben den Metadaten optional Volltexte, Bilder, Digitalisate, Audio- und Videodateien verschiedenster Genres und in beliebigen Dateiformaten oder Forschungsdaten.

Im Rahmen des 106. Bibliothekartages veranstaltete die MyCoRe-Community am 1. Juni 2017 einen Workshop zum Thema MyCoRe. Die Referentinnen und Referenten des Workshops gingen auf verschiedene Themen im Zusammenhang mit MyCoRe ein

Der Workshop, der von Wiebke Oeltjen (MyCoRe-Geschäftsstelle) moderiert wurde, begann mit einem Einführungsvortrag von Kathleen Neumann. Sie ist bei der Verbundzentrale des GBV (VZG) für den Repository-Service zuständig.³ Der Dienst „VZG-Reposis“ ist ein Hosting-Service, in dem Installation, Anpassung und Betrieb der MyCoRe-Anwendung MIR⁴ angeboten wird. In dem Beitrag wurde ein Blick auf den aktuellen Stand der MyCoRe-Entwicklung gegeben. Im Vortrag von Frank Lützenkirchen (Universitätsbibliothek Duisburg-Essen) wurden der Metadatenimport und die Metadatenanreicherung in MyCoRe thematisiert. Anschließend berichtete Steffen Hankiewicz (intranda GmbH) über den automatischen Ingest aus der Software „Goobi“⁵. Er zeigte den Workflow mit Digitalisaten, die mit der Software Goobi erfasst werden und über die standardkonforme SWORD-Schnittstelle⁶ u.a.

1 Siehe „MyCoRe,“ zuletzt geprüft am 01.08.2017, <http://www.mycore.de>.

2 Frank Lützenkirchen, „MyCoRe – Ein Open-Source-System zum Aufbau digitaler Bibliotheken,“ *Datenbank Spektrum* 2, Nr. 4 (November 2002): 23–27, <urn:nbn:de:hbz:464-20120810-143808-4>.

3 „Reposis – Repository-Service der VZG,“ VZG, zuletzt geprüft am 01.08.2017, <https://www.gbv.de/Verbundzentrale/serviceangebote/reposis-repository-service>.

4 „MIR MyCoRe / MODS Institutional Repository,“ zuletzt geprüft am 31.07.2017, <http://www.mycore.de/mir/>.

5 „Goobi: Überblick über die Funktionen,“ intranda, zuletzt geprüft am 31.07.2017, <http://www.intranda.com/digiverso/goobi/>.

6 Simple Webservice Offering Repository Deposit.

auch an MyCoRe-Anwendungen weitergeleitet werden.⁷ Über die Arbeit mit und Erschließung von Digitalisaten ging es auch in dem folgenden Beitrag. Matthias Eichner (Thüringer Universitäts- und Landesbibliothek – ThULB) präsentierte seine Entwicklungsarbeit, die die Digitalisate mittels optischer Zeichenerkennung (OCR) durchsuchbar macht. Es folgte ein Erfahrungsbericht von Ulrike Stahl (Julius Kühn-Institut – JKI) über die MyCoRe-Webapplikation „OpenAgrar“⁸. Dieses Repositorium zeichnet sich dadurch aus, dass neben den Publikationen mit ihren Metadaten auch Forschungsdaten erfasst werden. In einem weiteren Vortrag ging Robert Stephan (Universitätsbibliothek Rostock) auf die Vernetzungsmöglichkeiten von Daten ein. Er zeigte am Beispiel der Professorenkataloge auf MyCoRe-Basis, wie sich Synergien nutzen, Informationen vernetzen und Daten auswerten lassen. Im Folgenden werden diese Beiträge jeweils kurz zusammengefasst.

2. MyCoRe 2017 – Aktuelle Entwicklungen und ein Blick nach vorn

Kathleen Neumann (VZG) gab in ihrem Vortrag einen kurzen MyCoRe-Einstieg und -Überblick. Weiterhin berichtete sie über aktuelle MyCoRe-Entwicklungen und wagte einen Blick nach vorn. Sie beschrieb die Möglichkeit der freien Konfiguration als einen Vorteil von MyCoRe, wodurch ein vielseitiger Einsatz des Software-Frameworks ermöglicht wird. Auch die kontinuierliche Weiterentwicklung von Schnittstellen zum Datenaustausch trägt dazu bei, dass MyCoRe in verschiedenen Kontexten eingesetzt werden kann. Die Orientierung an Standards wie z.B. METS⁹/MODS¹⁰ oder SWORD¹¹ trägt darüber hinaus zur Nachhaltigkeit bei.¹² Seit mehr als 15 Jahren wird MyCoRe von einer deutschen Community kontinuierlich erneuert und weiterentwickelt.

Das Framework bringt dabei alles mit, was funktional für das DINI-Zertifikat für Open-Access-Repositorien und -Publikationsdienste¹³ notwendig ist. Mit dem im Juni erschienenen Long-Term-Support-Release (MyCoRe LTS 2017) kann das aktuelle Zertifikat 2016 erlangt werden, was die DINIready-Evaluierung¹⁴ des VZG-Reposis-Dienstes belegt.

7 „MyCoRe-SWORD-Schnittstelle,“ zuletzt geprüft am 01.08.2017, <http://www.mycore.de/documentation/interfaces/sword.html>.

8 „OpenAgrar,“ zuletzt geprüft am 31.07.2017, <https://www.openagrar.de/content/index.xml>.

9 Metadata Encoding & Transmission Standard.

10 Metadata Object Description Schema.

11 Simple Webservice Offering Repository Deposit.

12 Siehe z.B. „Die METS-Unterstützung,“ zuletzt geprüft am 01.08.2017, http://www.mycore.de/documentation/basics/mcrderivate_ifs/mets.html.

13 „DINI-Zertifikat 2016 ‚Open-Access-Repositorien und -Publikationsdienste,“ Deutsche Initiative für Netzwerkinformation e. V. (DINI), zuletzt geprüft am 31.07.2017, <https://dini.de/dini-zertifikat/>.

14 „DINI-ready: Modularisierung des Zertifizierungsprozesses,“ DINI, zuletzt geprüft am 31.07.2017, <https://dini.de/dini-zertifikat/dini-ready/>.

Aus den umfangreichen Weiterentwicklungen wurden drei kurz vorgestellt:

- Die Möglichkeit im MODS-Datenmodell mehrere Personenidentifikatoren¹⁵ wie beispielsweise GND¹⁶, ORCID¹⁷ oder VIAF¹⁸ abzulegen und die Unterstützung dafür innerhalb von MIR.
- Die Möglichkeit in MyCoRe abgelegte Videos über Pseudo-Streaming direkt aus der Anwendung heraus abzuspielen oder alternativ Streaming-Server wie den Wowza-Media-Server¹⁹ anzubinden.
- Die Unterstützung der IIIF-Image-API²⁰ um abgelegte Bilder standardkonform für die Darstellung durch Dritte bereitzustellen.

Außerdem zeigte Kathleen Neumann kurz, wie sich einzelne Entwicklungen im MIR präsentieren. Durch die Nutzung der DataCite-REST-API ist es möglich DataCite-DOIs²¹ aus der Anwendung heraus zu vergeben, wobei auch die spätere Metadatenaktualisierung bei DataCite²² direkt aus der MyCoRe-Anwendung heraus ermöglicht wird. Zukünftig soll MyCoRe stärker hin zu einem Backend-Framework mit entsprechenden Schnittstellen für die flexible Nutzung der abgelegten Daten entwickelt werden, so dass auch die Wahl des Frontends flexibler und unabhängig von XSL²³ wird.

3. Import und Anreicherung von Metadaten

Frank Lützenkirchen (UB Duisburg-Essen) stellte in seinem Vortrag den „Enrichment Resolver“ vor. Mit dem aktuellen LTS-Release wurde der MyCoRe-Kern um Funktionen erweitert, die den Import und die Anreicherung von Publikationsdaten aus externen Quellen vereinfachen. Der Enrichment Resolver ermöglicht es, Publikationsdaten auf MODS Basis aus konfigurierbaren Quellen zu importieren. Für eine erste Anwendung, die Universitätsbibliographie der Universität Duisburg-Essen²⁴, wurden bereits Adapter für zwölf Datenquellen implementiert, darunter Scopus²⁵, PubMed²⁶, IEEE²⁷, DataCite, Crossref²⁸ und mehrere Verbundkataloge. Der Resolver verwendet dazu eindeutige Identifikatoren wie DOI, IEEE Article Number, PubMed ID, ISBN, ISSN usw., um Publikationsdaten aus der externen Quelle zu laden, in das MODS-Format zu konvertieren und vorhandene Daten damit anzureichern. Die Datenquellen können dabei unterschiedlich priorisiert werden. Das Anreichern der Daten erfolgt iterativ und hierarchisch für verschiedene Ebenen der

15 Siehe Dokumentation zu „Persistent Identifier,“ zuletzt geprüft am 01.08.2017, <http://www.mycore.de/documentation/interfaces/pid.html>.

16 Gemeinsame Normdatei.

17 Open Researcher and Contributor ID.

18 Virtual International Authority File.

19 „Wowza Streaming Engine,“ WOWZA media systems, zuletzt geprüft am 01.08.2017, <https://www.wowza.com/products/streaming-engine>.

20 International Image Interoperability Framework.

21 Digital Object Identifier.

22 „DataCite,“ zuletzt geprüft am 01.08.2017, <https://www.datacite.org/>.

23 Extensible Stylesheet Language.

24 „Universitätsbibliographie,“ Universität Duisburg-Essen, zuletzt geprüft am 01.08.2017, <https://bibliographie.ub.uni-due.de/>.

25 „Scopus,“ Elsevier, zuletzt geprüft am 01.08.2017, <https://www.elsevier.com/solutions/scopus>.

26 „PubMed,“ zuletzt geprüft am 01.08.2017, <http://www.pubmed.gov>.

27 „IEEE Xplore,“ Institute of Electrical and Electronics Engineers, zuletzt geprüft am 03.11.2017, <http://ieeexplore.ieee.org/Xplore/home.jsp>.

28 „Crossref,“ zuletzt geprüft am 01.08.2017, <https://www.crossref.org/>.

Publikation (z. B. Artikel, Zeitschrift bzw. Aufsatz, Sammelband, Serie). Wissenschaftler/innen können so z. B. durch Eingabe des DOI ihre Publikationen in der Universitätsbibliographie melden. Die Funktionalität eignet sich auch zum automatisierten Zusammenführen identifizierter Dubletten. Eine weitere Komponente ermöglicht den automatisierten Import neuer Publikationen über einen RSS Feed, den viele externe Quellen wie Scopus oder PubMed anbieten.

4. Automatischer Ingest aus Goobi – Ein Praxiseinblick

Steffen Hankiewicz (intranda GmbH) ging in seinem Vortrag über das Zusammenspiel zwischen Goobi und MyCoRe zunächst kurz darauf ein, dass sich hinter Goobi neben der Workflowsteuerung ebenso ein Framework für digitale Bibliotheken verbirgt. Weiter führte er aus, für welche Einsatzzwecke Goobi eingesetzt wird. Im Verweis auf konkrete Anwender, die ihre Digitalisierungsergebnisse in MyCoRe einspielen, zeigte er die individuellen Wege auf, die hierfür beschritten wurden. So ließ sich am Beispiel der UB Braunschweig nachvollziehen, wie die Datenübergabe größtenteils manuell erfolgen kann. Die Herangehensweise der UB Rostock hingegen zeigt, dass eine vollautomatisierte Übergabe der Daten aus Goobi nach MyCoRe mit etwas Implementierungsaufwand ebenso gelingt und dass sogar eine Datenübergabe aus MyCoRe in Richtung eines extern betriebenen Goobi Viewers dank des Standards IIIF möglich ist, wie dies im Landesportal Mecklenburg-Vorpommern der Fall ist. Aus dem Zusammenspiel mit verschiedenen MyCoRe Instanzen und aufbauend auf den parallel gewonnenen Erfahrungen der jeweiligen Ingest-Implementierungen in andere Repositorien wie Fedora²⁹ und Langzeitarchivierungssysteme wie Rosetta³⁰ und Preservica³¹ entstand die Idee der Zusammenarbeit zwischen der ThULB Jena und intranda, einen standardisierten Ingest zu implementieren. Steffen Hankiewicz schilderte, welche Stärken und Schwächen die bisherigen Implementierungen von automatisierten Ingests in andere Systeme hatten und welche Konsequenzen sich für die Implementierung auf der Basis der SWORD-Schnittstelle ergaben. Er zeigte auf, dass der Ingest auf der Basis von SWORD bereits vollständig funktioniert, dass allerdings auch noch Optimierungspotenzial bezüglich der Datenpakete besteht, wenn diese einen Umfang von 10 GB überschreiten. Abschließend schilderte er die in der Zusammenarbeit mit der ThULB Jena gewonnenen Erfahrungen, die belegen, dass ein Ingest auf der Basis von SWORD sinnvoll ist: Technisch, weil es sich hierbei um eine solide Implementierung handelt, die stark von den Erfahrungen anderer Implementierungen profitieren konnte und entsprechend verlässlich ist, und strategisch, weil die beiden Communities rund um Goobi und MyCoRe mit der Wahl eines akzeptierten Standards einerseits ihre gute Zusammenarbeit verdeutlichen und andererseits zugleich bereitwillig als Referenzen und Ansprechpartner für andere Communities und Implementierungen im Rahmen von digitalen Bibliotheken und Repositorien bereitstehen.

29 „Fedora.“ zuletzt geprüft am 01.08.2017, <http://www.fedora.info/>.

30 „Rosetta.“ ExLibris, zuletzt geprüft am 01.08.2017, <http://www.exlibrisgroup.com/de/category/Rosetta>.

31 „Preservica.“ zuletzt geprüft am 01.08.2017, <http://preservica.com/>.

5. Von der Digitalisierung zur Präsentation

Matthias Eichner (ThULB Jena) zeigte in seinem Vortrag den aktuellen Stand der OCR-Integration in MyCoRe. Wesentliche Verbesserungen wurden im MyCoRe-Bildbetrachter implementiert.³² Der Bildbetrachter unterstützt nun ein Texthighlighting³³, die reine Ausgabe des OCR-Textes bei Bewahrung der Struktur und eine verbesserte Suchfunktion. Weiterhin wurden die notwendigen XML-Formate vorgestellt. Zum einen ALTO³⁴, ein von der *Library of Congress* entwickeltes Format zur Speicherung von OCR-Informationen und dem Layout von Seiten. Die Seite wird dabei in Paragraphen zerlegt, welche Zeilen enthalten und diese wiederum Wörter. Und zum anderen METS, welches als Container-Format dient und alle relevanten Informationen bündelt. Das betrifft die Verknüpfung zwischen der physikalischen Struktur (Digitalisat, ALTO) und der logischen Struktur (vereinfacht das Inhaltsverzeichnis). Am Beispiel des Jenaer Volksblattes wurde gezeigt, wie man ein OCR-Projekt mit MyCoRe realisiert. Von der Vorbereitung, also Digitalisierung, OCR-Erkennung und Nachstrukturierung bis zum automatisierten Hochladeprozess.

6. Forschungsdaten im OpenAgrar-Repository – Möglichkeiten und Herausforderungen

Ulrike Stahl (JKI) stellte in ihrem Vortrag das Repository OpenAgrar³⁵ vor und beschrieb die Möglichkeiten und Herausforderungen bei der Speicherung von Forschungsdaten. Das Repository OpenAgrar ist der gemeinsame Publikationsserver der Bundesforschungseinrichtungen im Geschäftsbereich des Bundesministeriums für Ernährung und Landwirtschaft sowie Einrichtungen der Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz. OpenAgrar weist einerseits die gesamte Bandbreite der Veröffentlichungen der Beschäftigten wie z.B. Hochschul- und Institutsschriften, Reports, Bücher, Konferenz- und Zeitschriftenbeiträge oder AV-Medien nach, dient aber andererseits auch als Publikationsserver für Forschungsdaten. Aktuell umfasst das Repository ca. 90.000 Dokumente, davon 6 erstpublizierte Forschungsdatensätze. Um Forschungsdaten öffentlich zugänglich und zitierbar zu machen, wird OpenAgrar entsprechend etablierter Standards sowie anhand der Bedarfe der Institute gemeinschaftlich und in Zusammenarbeit mit der Verbundzentrale des GBV (VZG) als Hosting Service angepasst und entwickelt. Zum derzeitigen Stand können für Forschungsdaten eine DOI via DataCite vergeben werden und folgende Metadaten erfasst und angezeigt werden: Titel, Abstract, Autor, Publikations-/Produktionsjahr, Publisher, Typ, Fileformat/-größe, Schlagwörter, Sachgruppe DNB, Lizenz, Version, verbundene Quellen wie dazugehörige Textpublikation. Es erlaubt damit das Erfassen der zwingend erforderlichen Metadaten für Forschungsdatenrepositorien entsprechend aktueller Empfehlungen. Zusätzlich werden Exportformate für Literaturverwaltungsprogramme sowie Zitierlink und Social Media Metriken angeboten. Die Qualität von Forschungsdatenpublikationen wird eigenverantwortlich von den Einrichtungen z.B. durch Prüfung der Vollständigkeit der Metadaten und

32 „Bildbetrachter – MyCoRe-Viewer,“ zuletzt geprüft am 01.08.2017, <http://www.mycore.de/documentation/frontend/imageviewer.html>.

33 Die Abbildung der Zeitung „Jenaer Volksblatt“ veranschaulicht dies: siehe „Jenaer Volksblatt“ Nr. 1, Dienstag, den 15. April 1890, zuletzt geprüft am 31.07.2017, http://zs.thulb.uni-jena.de/rsc/viewer/jportal_derivate_00250870/JVB_18900415_001_167758667_B1_001.tif.

34 „ALTO,“ zuletzt geprüft am 01.08.2017, <https://www.loc.gov/standards/alto/>.

35 „OpenAgrar,“ zuletzt geprüft am 31.07.2017, <https://www.openagrar.de/content/index.xml>.

der Ausführlichkeit der Beschreibung der Daten hinsichtlich Nachnutzbarkeit garantiert. Um Daten der Öffentlichkeit besser und schneller zugänglich zu machen, sind Entwicklungen zu erweiterten Such-/Filterfunktionen, Betrachtungsmodi für Tabellen, Bilder und Videos sowie Schnittstellen zu einrichtungsspezifischer Datenmanagementsoftware herausfordernde Entwicklungen. Zeitnah hingegen können zusätzliche Metadaten zu Förderern wie DFG oder BMBF implementiert oder aus Forschungsprojektdatenbanken eingebunden und mit den entsprechenden Dokumenten verknüpft werden. Das Julius Kühn-Institut (JKI) strebt als eine betreibende Einrichtung die Entwicklung eines Workflows sowie Implementierung für den Datenupload und Metadateneingabe via Webformular für JKI-Autoren mit anschließendem Review durch die Bibliothek und/oder durch externe Reviewer an, wenn Daten zum Artikel in einem Journal im Eigenverlag des JKI über das Open Journal System (OJS) publiziert werden sollen. Ebenso sollen Listen und Kennzahlen gemäß den Anforderungen von Evaluationen auf Autoren-/ Instituts-/ oder Gesamteinrichtungsniveau erstellt werden können.

7. Professorenkataloge auf MyCoRe-Basis: Synergien nutzen, Informationen vernetzen, Daten auswerten

Robert Stephan (UB Rostock) stellte in seinem Vortrag den Rostocker Professorenkatalog „*Catalogus Professorum Rostochiensium*“³⁶ vor. Der Katalog wird in Vorbereitung auf das 600-jährige Jubiläum der Universität Rostock 2019 aufgebaut. Erfasst sind mehr als 2.300 Einträge (Stand Juli 2017) mit biografischen Angaben zu allen an der Universität Rostock tätigen Professorinnen und Professoren von 1563 bis heute.

Eingesetzt wird die Software MyCoRe, die es ermöglicht ein angepasstes XML-Datenmodell und das MyCoRe-Editorframework für die Datenerfassung zu nutzen. Dokumente in Form von Anhängen, wie z.B. Auszügen aus Personalakten oder persönliche Dokumente und Bilder bereichern die Einträge. Mittels Klassifikationen lassen sich die erfassten Personen unter verschiedenen Aspekten, wie Fakultätszugehörigkeit, Epochen oder Fachgebieten gruppieren. Die Möglichkeit Suchmasken und Trefferlisten frei zu definieren findet ebenso ihre Anwendung.

Weiter wurden Vernetzungsmöglichkeiten mit anderen Portalen über den GND-Beacon-Dienst³⁷ vorgestellt. Durch Bereitstellung und Auswertung von Listen mit GND-Nummern in einem einfachen Format³⁸ lassen sich verschiedene Portale, wie Biografien, Lexika oder Bibliothekskataloge untereinander vernetzen. Auf der Grundlage von Beacon-Dateien weiterer deutscher Professorenkataloge wurden Statistiken über die gemeinsame Schnittmenge von Personen gezeigt. Diese Daten könnten die Basis für weitere biografische Forschung zu Personennetzwerken oder Karrierewegen sein.

Die Software wird an mehreren Standorten nachgenutzt: Die Universität Hamburg hat anlässlich ihres 100. Jubiläums einen Professorenkatalog erstellt.³⁹ Dort sind annähernd 5.000 Professorinnen

36 „Catalogus Professorum Rostochiensium,“ zuletzt geprüft am 31.07.2017, <http://cpr.uni-rostock.de/>.

37 „Beacon Findbuch,“ zuletzt geprüft am 31.07.2017, <http://beacon.findbuch.de/>.

38 „Wikipedia: Beacon,“ zuletzt geprüft am 31.07.2017, <https://de.wikipedia.org/wiki/Wikipedia:BEACON>.

39 „Hamburger Professorinnen- und Professorenkatalog,“ zuletzt geprüft am 31.07.2017, <https://www.hpk.uni-hamburg.de/>.

und Professoren der Universität Hamburg seit 1919 mit biografischen Einträgen erfasst. Auch die Universität Braunschweig bereitet die Einführung eines Professorenkatalogs vor.

8. Ausblick

Die Beiträge des Workshops haben gezeigt, dass MyCoRe als Software-Framework vielfältig eingesetzt wird. Das Spektrum reicht vom Hosting-Dienst mit standardisierten Anwendungen auf der Basis der Repository-Anwendung MIR über Repositorien zur Forschungsdatenspeicherung bis hin zu individuell angepassten Applikationen wie z.B. den Professorenkatalogen, die darüber hinaus mit Datenquellen vernetzt sind. Schnittstellenprogrammierungen sind dabei das entscheidende Bindeglied, wie die Goobi-Anbindung gezeigt hat. Wie jedes Jahr ist ein Long Term Release (MyCoRe LTS 2017) veröffentlicht worden. Zur Förderung des Austauschs zwischen den MyCoRe-Entwicklern, MyCoRe-Usern und -Interessenten wird im November 2017 an der Universität Hamburg ein MyCoRe-Workshop stattfinden.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/2017H4S242-248>