

Kathleen Neumann, Wiebke Oeltjen, Ulrike Stahl, Robert Stephan

MyCoRe macht Forschungsdaten FAIR

Zusammenfassung: Forschungsdaten werden in Repositorien gespeichert und zur Nachnutzung erhalten. Dies setzt voraus, dass die Repositorien bestimmte Prinzipien einhalten. Die FAIR-Leitprinzipien besagen, dass wissenschaftliche Daten auffindbar, zugänglich, interoperabel und wiederverwendbar sein sollen. In diesem Artikel wird gezeigt, dass die Open-Source-Software MyCoRe geeignet ist, Repositorien bereitzustellen, die die FAIR-Prinzipien erfüllen. Dazu werden drei MyCoRe-Anwendungen, die Forschungsdaten verwalten und bereitstellen, untersucht und bewertet.

Schlüsselwörter: MyCoRe, Forschungsdaten, FAIR-Prinzipien, OpenAgris, Catalogus Professorum Rostochiensium, Corpus Musicae Ottomanicae

MyCoRe makes Research Data FAIR

Abstract: Research data are stored in repositories for long term use. This presupposes that the repositories comply with certain principles. The FAIR data guiding principles propose that research data should be Findable, Accessible, Interoperable, and Reusable (FAIR). This article shows, that the open source software MyCoRe is capable of providing repositories that meet the FAIR principles. Therefore three MyCoRe applications that manage and provide research data will be examined and evaluated.

Keywords: MyCoRe, research data, FAIR guiding principles, openagris, catalogus professorum rostochiensium, corpus musicae ottomanicae

1 MyCoRe trifft auf Forschungsdaten

Forschungsdaten¹ bilden einen Grundpfeiler wissenschaftlicher Erkenntnis und sind die Basis für weitere Forschung. Eine transparente Dokumentation der Forschungsdaten, ein verantwortungsvolles Forschungsdatenmanagement (FDM) einschließlich qualitätsgesicherter Archivierung und/oder Veröffentlichung sorgen für die Nachvollziehbarkeit und Reproduzierbarkeit von Forschungsprozessen und deren Ergebnissen und ermöglichen eine vielfältige Nachnutzung.

¹ Zur Definition von Forschungsdaten siehe z.B. Kindling und Schirmbacher (2013) 130.

MyCoRe² ['maiko:r] ist ein Open-Source-Framework zur Erfassung, Verwaltung und Präsentation digitaler Objekte³ und deren Metadaten. Die bis heute mehr als 70 realisierten Anwendungen (z.B. institutionelle Repositorien, Archive und Online-Lexika) enthalten auch verschiedene Forschungsdaten. Von der zitierfähigen Ablage einzelner Forschungsdaten auf Publikationsservern bis zu fachspezifischen Datenbanken und Portalen, zeigt sich dabei ein breites Spektrum.

MyCoRe wird seit 2001 von einer bundesweiten Gemeinschaft an Universitätsbibliotheken, universitären Rechenzentren und an der Verbundzentrale des GBV (VZG) kontinuierlich weiterentwickelt. Dabei standen schon immer Prinzipien im Mittelpunkt, die wir heute unter anderem als FAIR-Leitprinzipien⁴ kennen: Daten und Metadaten sollten in MyCoRe-Webanwendungen im Rahmen einer entsprechenden Infrastruktur auffindbar (**F**indable), zugänglich (**A**ccessible), interoperabel (**I**nteroperable) und wiederverwendbar (**R**eusable) sein. Dafür stellt das MyCoRe-Framework Schnittstellen und Funktionen bereit, die zum Verwalten, Speichern, Präsentieren und Austauschen von Metadaten und den digitalen Ressourcen benötigt werden. Zum besseren Verständnis sind die Prinzipien in der Tabelle 1 mit Stichworten aufgelistet.

Tab. 1: Die FAIR-Prinzipien und Rahmenrichtlinien für Repositorien

Bewertungskriterien

| Auffindbarkeit (Findable) | |
|--|---|
| F1 | (Meta-)Daten erhalten global eindeutige und dauerhafte PIDs |
| F2 | Beschreibung der Daten mit umfangreichen Metadaten |
| F3 | Klare Referenz von Metadaten zu Daten mittels ID |
| F4 | Metadaten sind in durchsuchbaren Verzeichnisdiensten erfasst |
| Zugänglichkeit (Accessible) | |
| A1 | Auffindbarkeit der (Meta-)Daten über ein standardisiertes Protokoll |
| A1.1 | Protokoll ist offen, frei und universell |
| A1.2 | Protokoll unterstützt Authentifizierung und Rechteverwaltung |
| A2 | Metadaten sind/bleiben verfügbar |
| Interoperabilität (Interoperable) | |

² Siehe <http://www.mycore.de>.

³ wie z.B. Dokumente, digitalisierte Urkunden, Akten, Handschriften, Bücher, Kataloge, Zeitschriften, Zeitungen, Fotos, Scans, Video- oder Audiodateien.

⁴ Siehe Box 2: „The FAIR Guiding Principles“ in Wilkinson et al. (2016) 4.

Bewertungskriterien

| | |
|---|--|
| I1 | Nutzung etablierter Formalismen zur Präsentation der (Meta-)Daten |
| I2 | Nutzung FAIRer Vokabulare in den (Meta-)Daten |
| I3 | Qualifizierte Referenz zwischen den (Meta-)Daten |
| Wiederverwendbarkeit (Reusable) | |
| R1 | Detailliert beschriebene (Meta-)Daten mit präzisen und relevanten Attributen |
| R1.1 | Klare Angabe der Nutzungslizenz |
| R1.2 | (Meta-)Daten enthalten Provenienz-Informationen |
| R1.3 | (Meta-)Daten entsprechen fachgebietsrelevanten Standards |
| Rahmenrichtlinie für Repositorien (Guidelines) | |
| G1 | Global eindeutige und dauerhafte PIDs (analog zu F1) |
| G2 | PIDs unterstützen mehrere Granularitäten |
| G3 | PIDs lösen auf die sogenannte „Landing Page“ (LP) auf |
| G4 | PID ist maschinenlesbar in LP eingebettet |
| G5 | Dokumentation und Unterstützung für Zitation der Daten |
| G6 | LP enthält Metadaten für Zitation und Auffindbarkeit |
| G7 | Metadaten sind entsprechend schema.org in JSON-LD eingebettet |
| G8 | Metadaten sind über HTML-Tags auslesbar |
| G9 | Download der Metadaten in Bibtex o.ä. möglich |
| G10 | „Content Negotiation“ (CN) für schema.org/JSON-LD |
| G11 | Unterstützung von sogenannten „HTTP Link Headern“ für CN |

2 Offenen Zugang zu Daten FAIR gestalten

Die FAIR-Prinzipien, die für Menschen und Maschinen gleichermaßen gelten sollen,⁵ haben die Wiederverwendung von Forschungsdaten als übergeordnetes Ziel. Sie werden heute oft im Zusammenhang mit der Forderung nach offenem Zugang zu Forschungsdaten proklamiert, wobei „offen“ in erster Linie verlangt, dass klar beschrieben ist, wo die Daten liegen. Mit dem expliziten Bezug auf die FAIR-Prinzipien beim geforderten Datenmanagementplan im aktuellen Horizon 2020

⁵ Siehe Mons et al. (2017) 51.

Förderprogramm⁶ werden sie zu einem Grundstein des Forschungsdatenmanagements.⁷ Eine wesentliche Rolle in den Prinzipien spielt die Zitierung von Forschungsdaten und Metadaten. Daher verfeinern Empfehlungen zur Zitierung von Forschungsdaten einzelne FAIR-Kriterien.⁸ Sie sind in Tab. 1 in den Rahmenrichtlinien für Repositorien zusammengefasst.

Die Umsetzung der sich daraus ergebenden Anforderungen stellen datenerzeugende Forschungseinrichtungen vor enorme Herausforderungen. Es ist eine Infrastruktur zu schaffen, die die anfallenden spezifischen Forschungsdaten und Metadaten bestmöglich erfasst, dokumentiert, analysiert, publiziert, archiviert und gleichzeitig den FAIR-Prinzipien genügt. Dadurch wächst die Notwendigkeit bestehende Software zu evaluieren, ob und wie sie den FAIR-Prinzipien genügt⁹ und deren FAIRness¹⁰ transparent zu messen.

Wir zeigen hier, inwieweit das MyCoRe-Framework selbst, sowie die folgenden drei MyCoRe-Anwendungen, die alle frei zugängliche Forschungsdaten enthalten, die FAIR-Prinzipien und die Empfehlungen zur Zitierung für Datenrepositorien vollständig und optimal (++), vollständig mit Potential (+), teilweise (?) oder nicht (–) erfüllen. Dabei vergeben wir vollständig und optimal (++) nur dann, wenn eine technische Möglichkeit vollständig implementiert und auch konsequent in der Anwendung umgesetzt ist. MyCoRe kann somit bestenfalls nur vollständig mit Potential (+) bewertet werden.

- OpenAgrar,¹¹ der Publikationsserver der Bibliotheken der Ressortforschungseinrichtungen des Bundesministerium für Ernährung und Landwirtschaft (BMEL-Forschung),
- der Catalogus Professorum Rostochiensium,¹² ein Personenlexikon über alle an der Universität Rostock tätigen Professorinnen und Professoren seit ihrer Gründung 1419, kurz Rostocker Professorenkatalog und
- der Corpus Musicae Ottomanicae¹³ (CMO) Musik-Editionen und -Quellenkatalog des gleichnamigen Gemeinschaftsprojektes des Institutes für Musikwissenschaft der Westfälischen Wilhelms-Universität Münster, der Max-Weber-Stiftung und des Orient-Institutes Istanbul.

⁶ Siehe EU (2016).

⁷ Siehe Boeckhout et al. (2018).

⁸ Siehe „Table 1. Guidelines for Repositories“ in Fenner et al. (2016) 2.

⁹ Siehe Dunning et al. (2017).

¹⁰ Siehe Wilkinson et al. (2018).

¹¹ Siehe <https://www.openagrar.de/>.

¹² <http://cpr.uni-rostock.de/>.

¹³ <https://corpus-musicae-ottomanicae.de/>.

3 MyCoRe und die FAIR-Prinzipien

3.1 Auffindbarkeit (Findable)

Sowohl Daten als auch Metadaten sollen von Mensch und Maschine leicht zu finden sein. Zu diesem Zweck sollen die Forschungsdaten global eindeutige und dauerhafte Persistent Identifier (PIDs) erhalten (F1) und mit umfangreichen Metadaten beschrieben werden (F2). Weiterhin ist eine klare Referenz der Metadaten zu den entsprechenden Daten mittels einer ID zu realisieren (F3). Eine hochwertige Indizierung ermöglicht schlussendlich das Harvesting und die Meldung der Metadaten an einschlägige Verzeichnisdienste (F4).

Einer der in MyCoRe bereitgestellten Softwarebausteine, ermöglicht es beliebige PIDs zu generieren. Eine Implementierung zur Erzeugung von DNB-URNs, Datacite-DOIs und PURLs wird bereits mitgeliefert. Weitere Realisierungen z.B. für Crossref-DOIs oder Handle-Dienste können leicht ergänzt werden (F1: +). Das Datenmodell ist in einer MyCoRe-Anwendung frei definierbar. Einzige Einschränkung ist die Festlegung auf XML. MyCoRe ermöglicht einerseits eine flexible Datenmodellierung, so dass Daten detailliert zu erfassen sind, andererseits können aber auch XML-Metadatenstandards wie MODS oder MEI direkt als Datenmodell eingebunden werden (F2: +). Ist das Datenmodell entsprechend definiert, können die Metadaten eindeutig und explizit Informationen über die Daten enthalten, die sie beschreiben (F3: +). MyCoRe nutzt die Open-Source-Suchplattform Apache Solr und stellt damit umfangreiche Suchfunktionen zur Verfügung. Sowohl Metadaten als auch Volltexte können somit direkt über Solr durchsucht werden. Um von der konkreten Suchplattform zu abstrahieren, stellt MyCoRe weiterhin eine eigene Anfragesprache bereit, die „MyCoRe Query Language“ (MCR-QL)(F4: +).

3.2 Zugänglichkeit (Accessible)

Eine langfristige Verfügbarkeit der (Meta-)Daten über standardisierte (A1), offene und frei zugängliche Protokolle (A1.1) macht diese sowohl für Mensch als auch für Maschine leicht zugänglich. Auch der Zugriff stark geschützter Daten muss entsprechend realisiert werden (A1.2). Für den Fall, dass die eigentlichen Daten nicht mehr verfügbar sind, wird zu Dokumentationszwecken der Zugriff auf die Metadaten weiterhin sichergestellt (A2).

MyCoRe-Anwendungen nutzen das offene Protokoll HTTP/HTTPS, so dass die (Meta-)Daten entsprechend ihrer Zugriffsrechte frei zugänglich sind (A1: +). Es ist ein Rollen- und Rechtesystem implementiert, das den Zugriff auf die (Meta-)Daten steuerbar macht. Jedem Benutzer ist ein Authentifizierungsbereich zugeordnet, der konfiguriert werden kann. Privilegien steuern Berechtigungen und Access Control Lists (ACLs) regeln den Zugriff auf Metadaten, auch wenn die Ressourcen nicht mehr verfügbar sein sollten (A2: +).

3.3 Interoperabilität (Interoperable)

Die Daten müssen derart bereitgestellt werden, dass ein Austausch und Interpretation innerhalb anderer Systeme möglich ist. Daher sollen etablierte Formalismen (I1) und Vokabulare (I2) zur Erfassung und Präsentation genutzt werden. Bei entsprechender Abhängigkeit verschiedener Datensätze untereinander, sollen diese qualifiziert referenziert werden (I3).

Die Auszeichnungssprache XML spielt in MyCoRe eine zentrale Rolle. Sie gewährleistet die Interoperabilität, da Metadaten-Objekte intern als XML-Dokumente persistent abgelegt werden. Auch können XSL-Stylesheets für beliebige Ausgabeformate wie beispielsweise XML, CSV, JSON oder BibTeX implementiert werden. So lassen sich Metadaten und Klassifikationen vielfältig im- und exportieren. Dritte Systeme können diese Schnittstelle für den Batch-Import von Metadaten verwenden (I1: +). Um das Prinzip I2 zu erfüllen, müssen die Anwendungsentwickler von Repositorien das verwendete Vokabular - den FAIR-Prinzipien entsprechend - leicht auffindbar dokumentieren und falls möglich Standards verwenden. Eine Vernetzung von Datenquellen ist in MyCoRe über qualifizierte Verweise möglich (I2: +). Auch können externe Daten via Linked Open Data in MyCoRe-Webanwendungen integriert werden (I3: +). Das Framework bringt also die technischen Grundvoraussetzungen mit um Interoperabilität in der jeweiligen Anwendung zu realisieren.

3.4 Wiederverwendbarkeit (Reusable)

Durch detaillierte und präzise Beschreibung der Daten (R1), die Angabe einer eindeutigen Nutzungslizenz (R1.1) und Provenienz-Informationen (R1.2) können die Forschungsdaten auch für zukünftige Forschung wiederverwendet werden. Werden fachgebietsrelevante Standards verwendet können die Daten noch leichter verstanden und nachgenutzt werden (R1.3).

Dies liegt im Verantwortungsbereich der Daten erfassenden Institutionen und Personen. Auf der technischen Seite können mit MyCoRe-Anwendungen (Meta-)Daten menschen- und maschinenlesbar bereitgestellt werden (R1: +). Begünstigt wird die Wiederverwendbarkeit der Daten, wenn sie nativ und in einem aussagekräftigen bzw. standardisierten Datenmodell vorliegen. Die Software bietet lediglich die Werkzeuge zur Wiederverwendung der Daten. So ist beispielsweise in der Anwendung MIR¹⁴ die Lizenzangabe ein Pflichtfeld. Weiterhin kann der Rechteinhaber im Sinne des Urheberrechts benannt (R1.1: +) und eine Institutionszugehörigkeit angegeben werden. Auch können Provenienz-Informationen je nach Metadatenschema in den Metadaten erfasst (R1.2: +) und je nach Fachgebiet relevante Standards implementiert werden (R1.3: +).

¹⁴ MIR steht für MyCoRe MODS Institutional Repository, siehe auch mycore.de/mir.

3.5 Zitierbarkeit der (Meta-)Daten

Mit MyCoRe-Anwendungen können aus technischer Sicht die erforderlichen und empfohlenen Richtlinien zur Zitierung¹⁵ von Forschungsdaten umgesetzt werden (G1-G9: +). Dazu gehört, dass Persistent Identifier – auch granular – vergeben werden, die auf eine Landing Page verweisen, die wiederum umfangreiche Metadaten in verschiedenen Formaten menschen- und maschinenlesbar bereitstellen. An der Umsetzung der beiden optionalen Richtlinien „Content Negotiation“ (CN) für schema.org/JSON-LD und die entsprechende Unterstützung von sogenannten „HTTP Link Headern“ für CN (G10, G11: -) wird zurzeit gearbeitet.

4 OpenAgrar – ein Repository auch für Agrardaten

OpenAgrar ist das gemeinsame Repository von Einrichtungen im Geschäftsbereich des Bundesministeriums für Ernährung und Landwirtschaft (BMEL). Es dient einerseits dazu die Veröffentlichungen der Forschenden aus diesen Einrichtungen zu erfassen, zu archivieren und der Öffentlichkeit zugänglich zu machen und andererseits, um erzeugte Dokumente aber auch Forschungsdaten dieser Einrichtungen zu veröffentlichen und entsprechend nachnutzbar zu machen. Es werden alle Datenformate unterstützt – Bilder, Tabellen oder ganze Datenbanken können über eine Weboberfläche hochgeladen und entsprechend mit Metadaten angereichert werden.

Für alle erstpublizierten Datenobjekte im OpenAgrar können DOIs¹⁶ aus dem System heraus vergeben und über DataCite registriert werden. Sie sind über einen HTTP-Resolver auflösbar (G1: ++) und ihre Langlebigkeit ist durch eine entsprechende Policy der DOI-Registrierungsagentur¹⁷ garantiert (F1: ++). Außerdem kann ein Embargo auf das Datenobjekt selbst vergeben werden, sodass nur die Metadaten selbst sichtbar sind. Veröffentlichte Datenobjekte können gesperrt aber nicht undokumentiert gelöscht werden, wobei die Metadaten in allen Fällen erhalten bleiben. Das verwendete MODS-Metadaten-Schema im OpenAgrar wurde dafür auf das DataCite-Schema 4.1¹⁸ abgebildet.

Eine strukturierte granulare Vergabe von DOIs für einzelne Datenobjekte, die zu einer Datensammlung gehören oder verschiedene Teile, Versionen darstellen bzw. in einem anderen Format abgelegt sind, ist nur teilweise möglich (G2: ?). Die Datenobjekte können derzeit nur über die Beziehungstypen „Version von“, „anderes Format zu“, „Original zu“, „verbunden mit“ entsprechend

¹⁵ Siehe Tabelle 1 in Fenner et. al (2016).

¹⁶ Siehe z.B. DOI: [10.5073/20180816-152451](https://doi.org/10.5073/20180816-152451).

¹⁷ Siehe https://www.doi.org/doi_handbook/6_Policies.html.

¹⁸ DataCite Metadata Working Group (2017).

verlinkt werden (I3: +). Registrierte DOIs lösen immer auf eine Landing Page (LP) auf, wo neben dem DOI selbst u.a. Kontaktinformationen, bibliographische und datenbeschreibende Metadaten, Verlinkungen, Lizenzen und Zitierweisen stehen (G3, G5, G6: ++). All diese Metadaten sind von Menschen als auch Maschinen lesbar, wobei Letzteres über schema.org-Auszeichnung im JSON-LD Format, HTML Meta Tags, HighWirePress Tags, sowie Dublin Core HTML Meta Tags realisiert (F2, F3, G4, G7, G8: ++) ist.

Für jedes Datenobjekt ist es möglich, sowohl die für die DOI-Registrierung nötigen fünf obligatorischen als auch alle weiteren sechs empfohlenen Metadaten des DataCite-Schemas 4.1 zu erfassen, womit die Zitierung garantiert ist und die Auffindbarkeit gewährleistet wird. OpenAgrar ermöglicht die beidseitige Verlinkung zwischen erstpubliziertem Datenobjekt und dem erfassten dazugehörigem wissenschaftlichen Artikel über den Bezug „ist referenziert in“ bzw. „referenziert“ und erhöht somit die Auffindbarkeit im OpenAgrar selbst als auch über externe Data Literature Interlinking Services wie ScholeXplorer.¹⁹ Zusätzlich können Metadaten wie „Rechteinhaber“, „Beschreibung der Datenstruktur“ wie z.B. Parameter oder Messgrößen aber auch „Umfang/Größe“ des Datenobjektes, „Version“ und „Sprache“ erfasst werden (R1: +). Um Informationen über die Entstehung und Herkunft des Datenobjektes zu bekommen, erlaubt OpenAgrar das Erfassen von verschiedenen Beitragenden wie z.B. „Laborleitung“ oder „Projektleitung“ ebenso wie von verschiedenen zeitlichen Angaben wie „erstellt am“, „überprüft am“, „geändert am“, „aufgenommen am“ (R1.2: +). Das Metadatum „Rechte/Lizenz“ ist obligatorisch und erlaubt die Auswahl aus verschiedenen erfassten Lizenzen über eine Liste (R1.1: ++). Die zur Zitierung nötigen Metadaten sind in einem extra Feld „Zitierform“ ersichtlich und in verschiedenen Formaten wie z.B. BibTeX, RIS und ISI exportierbar (G9: ++). OpenAgrar erlaubt die Eingabe von IDs für Personen und Schlagwörtern (ORCID, GND, etc.) und ermöglicht somit eine eindeutige Zuordnung von Autoren bzw. zu Themengebieten. Das Datenobjekt selber muss obligatorisch einer Sachgruppe der Deutschen Nationalbibliothek zugeordnet werden. Alle diese nach extern referenzierten Metadaten lösen über eine entsprechende URL auf und sind auffindbar und dokumentiert (I2: +). Außerdem bietet OpenAgrar den Export im MODS und Dublin Core Schema an (I1: ++).

Die Daten und Metadaten sind über verschiedene Sucheinstiege im OpenAgrar recherchierbar. Da Volltexte ebenfalls indiziert sind, kann sowohl das Datenobjekt selbst besser gefunden werden als auch mitveröffentlichte Methodenbeschreibungen oder datenbeschreibende Dateien. Diese zusätzlichen Informationen geben dem Datennachnutzenden Auskunft über die Herkunft, den Prozessierungsgrad, oder über die Erzeugungsmethode des Datenobjektes. Durch beschränkte

¹⁹ <https://dliservice.research-infrastructures.eu>.

Zugriffsrechte kann der Zugang zum Datenobjekt selbst eingeschränkt bzw. gesteuert werden, dem Datennutzenden wird dies auf der LP angezeigt, dabei bleiben die Metadaten immer zugänglich. Im OpenAgrar können über ein Rollen- und Rechtemanagement entsprechend verschiedene Lese-/Schreibrechte für Autoren, Administratoren oder Bearbeiter zugewiesen werden (A1.2: ++).

Datennutzende haben freien Zugang, wohingegen datenpublizierende Autoren sich z.B. via SHIBBOLETH oder LDAP authentifizieren und registrieren müssen.

Über die OAI-PMH Schnittstelle werden Metadaten ausgegeben und können von Suchmaschinen gesammelt und indiziert (z.B. von Bielefeld Academic Research Engine – BASE²⁰), über die Rest-API exploriert und exportiert werden bzw. über die OpenAIRE Schnittstelle entsprechend Projekten zugeordnet werden (F4: ++). Ebenso ist OpenAgrar in re3data.org²¹, einem Katalog für Datenrepositorien, registriert. Die Datenobjekte und ihre Metadaten sind über HTTPS zugänglich (A1: ++) – einem offenen und universellen Format (A1.1: ++).

Im Falle der Beendigung des Betriebs von OpenAgrar sind die Betreiber für den (Meta-)Datenerhalt verantwortlich und müssen dies entsprechend in ihrer Policy verankern (A2: ?), vertraglich ist momentan eine Datensicherung für 10 Jahre garantiert.

OpenAgrar ist ein Datenrepositorium für den Agrarbereich, aber nicht spezialisiert auf einen bestimmten Datentyp wie z.B. Gensequenzdaten, Geodaten oder Charakterisierungsdaten pflanzengenetischer Ressourcen, die jeweils eigene Metadatenstandards haben, es kann somit nur generelle Metadaten verpflichtend erheben und indizieren (R1.3: +).

5 Rostocker Professorenkatalog

Der Rostocker Professorenkatalog ist ein biographisches Online-Lexikon, in dem alle Rostocker ProfessorInnen seit Gründung der Universität 1419 bis heute erfasst und mit ihren biographischen Informationen darstellt werden. Die Einträge werden mit Bildern und historischen Quelldokumenten angereichert und mit weiteren digitalen Ressourcen verknüpft.

In dieser Anwendung findet man die biografischen Informationen (Daten) nicht in Dateien, die in ein Repository hochgeladen und mit Metadaten beschrieben werden, sondern sie werden über Webformulare erfasst und direkt in den MyCoRe-Metdatenobjekten als XML gespeichert. Die Repository-Funktionalität von MyCoRe wird genutzt, um zusätzliches Bild- und Quellenmaterial für einen Datensatz zu veröffentlichen. Im Folgenden untersuchen wir, ob auch für diesen Spezialfall die FAIR-Prinzipien und die Kriterien zur Zitierung von Datensätzen erfüllt werden können.

²⁰ <https://www.base-search.net>.

²¹ <https://www.re3data.org/repository/r3d100012467>.

Für jeden Datensatz wird auf dem PURL-Server der Universitätsbibliothek Rostock ein PID erzeugt. Sollten in Zukunft eine Migration oder Integration in eine andere Anwendung unter einer neuen Web-Adresse notwendig werden, bleibt die zitierfähige PURL konstant. In diesem Fall wird lediglich eine neue Resolving-URL auf dem PURL-Server registriert (F1, G1: ++). In einer Empfehlung, fassen wir die für das Zitieren eines Datensatzes notwendigen Informationen (u.a. Titel, PID und Abrufdatum) zusammen. (F3: ++). Durch die Registrierung der Datensätze in Suchmaschinen (u.a. durch Implementierung des Google-Sitemap-Protokolls) werden diese weltweit sichtbar (F4: +).

Die Daten werden über das HTTP-Protokoll in Form von Webseiten für den menschlichen Leser und über die MyCoRe-REST-API maschinenlesbar im XML-Format bereitgestellt (A1: ++, A1.1: ++). Sämtliche Daten stehen ohne Zugriffsbeschränkung zur Verfügung. Somit besteht keine Notwendigkeit, die in MyCoRe enthaltenen Funktionen des Rollen- und Rechte-Managements für den Zugriff auf die Webseiten und die REST-API zu verwenden (A1.2: +). Da sich Daten und Metadaten nicht trennen lassen, bzw. die Metadaten aus den Datensätzen abgeleitet werden, ist eine Bereitstellung der Metadaten unabhängig von den Primärdaten nicht vorgesehen (A2: –). Dieses Kriterium könnte erfüllt werden, wenn Metadaten im Rahmen der Registrierung eines PIDs erzeugt werden, wie es z.B. von MyCoRe bei der Registrierung von DOIs unterstützt wird. Die Speicherung der Daten im XML Format und die Bereitstellung über die REST-API ermöglichen den Datenaustausch. Durch die Verwendung der GND-Nummer²² der beschriebenen Personen, wird eine eindeutige Zuordnung der Person und Verknüpfung mit anderen Datenquellen, beispielsweise über den GND-Beacon-Service²³ möglich. Die Verknüpfung weiterer Aspekte (z.B. Ortsnamen²⁴) ist im Datenmodell vorgesehen, wurde allerdings bislang nicht explizit erfasst (I3: +).

Die Beschreibung der Datenerhebung und des Datenumfangs lässt sich aus der Projektbeschreibung der Webseite entnehmen. Diese Daten sind aber nicht formalisiert oder maschinenlesbar (R1: ?). Die Notwendigkeit der Angabe einer Lizenz für die Nachnutzung der Daten wurde erkannt und zur Zeit diskutiert (R1.1: –). Durch die Anzeige des Zitierhinweises sowie Angaben zum Erfasser und letzten Bearbeiter eines Datensatzes sind grundlegende Angaben zur Provenienz möglich (R1.2: ?). Die Identifizierung einzelner Datensatzbestandteile oder verschiedener Versionsstände eines Objektes sind derzeit nicht vorgesehen (G2: ?). MyCoRe bietet jedoch die Möglichkeit mehrere Versionsstände für einen Datensatz zu speichern. Die persistenten URLs verweisen auf eine individuelle Webseite für jeden Datensatz, mit Hinweisen zur gewünschten Zitierform (G3: ++, G5: +). Weder eine

²² Gemeinsame Normdatei http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html.

²³ See-Also-Dienst auf <https://beacon.findbuch.de>.

²⁴ Siehe GeoNames <https://www.geonames.org/>.

maschinenlesbare Codierung (mittels JSON-LD bzw. HTML-Metatags) noch eine Bereitstellung in einem Zitierformat (BibTEX, RIS) wurde bislang umgesetzt (G4: -, G6: ?, G7: -, G8: -, G9: -), ließe sich aber durch die von MyCoRe bereitgestellten XML-Funktionen (z.B. XSLT) einfach implementieren.

6 CMO-Editionen und Quellen-Katalog

Die abschließend vorgestellte Anwendung Corpus Musicae Ottomanicae (CMO) ist eine Mischung aus Publikationsserver und Quellen-Katalog und Teil des auf 12 Jahre angelegten, gleichnamigen Projektes CMO. Dieses beschäftigt sich mit der Aufarbeitung von Musikhandschriften aus dem Nahen Osten, die sich auf Manuskripte osmanischer Musik aus dem 19. Jahrhundert konzentrieren. Es wird eine umfassende Onlinedatenbank aufgebaut, die Drucke, Manuskripte, zugehörige Online-Quellen und Beschreibungen der beteiligten Personen enthält. Ein weiterer wesentlicher Bestandteil des Projektes ist die Transkription und Edition der Quellen. Die so entstehenden kritischen Berichte, Musik- und Texteditionen werden ebenfalls auf der Online-Plattform veröffentlicht und mit dem zugehörigen Quellmaterial verknüpft.²⁵

Früh stand fest, dass für die Quellenbeschreibung der Metadatenstandard MEI (Music Encoding Initiative),²⁶ ein XML-basiertes Dokumentenformat für die Kodierung, den Austausch und die Archivierung von Musik verwendet werden soll²⁷ (I1: ++, R1.3: ++). Eine weitere Anforderung war es, keine reine Webpräsentation der Daten zu schaffen, sondern gleichzeitig ein Werkzeug mit dem die Informationen direkt erfasst, analysiert und erweitert werden können. Nach der Analyse verschiedener Lösungsmöglichkeiten fiel die Wahl unter anderem aus den folgenden drei Gründen auf MyCoRe:

1. Das Framework bietet die Möglichkeit XML-basierte Metadatenmodelle frei zu modellieren – also auch das hier geforderte MEI und dessen Relationen nach dem FRBR-Modell.²⁸
2. Der Einsatz eines Open-Source-Frameworks mit aktiver Community und Entwicklung sowie Betrieb bei der Verbundzentrale des GBV (VZG) gewährleisten eine nachhaltige Softwarelösung (A2: +).
3. Die ebenfalls in diesem Kontext angesiedelte MyCoRe-Anwendung Bach digital²⁹ ist seit 2010 online und belegt nochmals Umsetzbarkeit und Nachhaltigkeit (A2: +, R1.3: ++).

²⁵ Siehe Corpus Musicae Ottomanicae (2018) und Projekt CMO, Projektbeschreibung: <https://www.uni-muenster.de/CMO-Edition/cmo/beschreibung.html>.

²⁶ MEI (2018)

²⁷ Riley (2009)

²⁸ Functional Requirements for Bibliographic Records (FRBR).

²⁹ Siehe Bach Digital <https://www.bach-digital.de/>.

Seit Mai 2018 ist die Online-Plattform von CMO über HTTPS erreichbar (A1, A1.1, A1.2: ++) und der Bestand wächst seither stetig. Im Vordergrund des Projektes steht nachhaltiges Forschungsdatenmanagement. So sind neben den aktuellen Forschungsarbeiten auch für die internationale akademische Welt und Künstlergemeinschaft neue Einblicke und weitere Forschungen möglich, da diesen die Quellen zugänglich gemacht werden. Ein weiteres langfristiges Ziel ist es, die Datenbank für weitere Quellen zu öffnen um zukünftig eine breitere Datenbasis für die Musikforschung bereitzustellen.

Durch die Verwendung der Metadatenstandards MODS und MEI sind sowohl Quell- als auch Publikations-(Meta-)Daten umfangreich (F2:++), nachhaltig und maschinenlesbar (F4: +) beschrieben. Innerhalb der Metadaten wird auf standardisiertes Vokabular gesetzt: Standard List of Musical Terms (Makâm, Usûl, Tempo, Genre)³⁰ (I1: ++, R1.3: ++). In den Bereichen, in denen die derzeitigen Standards nicht ausreichen, wird im CMO-Projekt an einer entsprechenden Erweiterung des Standards gearbeitet.

Die Umsetzung der Relationen nach dem FRBR-Modell ist wie in der MEI-Spezifikation beschrieben implementiert. Dadurch ist es möglich die z.T. komplexen Zusammenhänge der Daten auch anwendungsunabhängig, also nativ auf der Platte nachzuvollziehen. Einen leicht verständlichen Zugang zu den Daten wie er in Cremer et al. (2018) gefordert wird, der auch ohne Webanwendung mit deren Suche, sowie Filtern funktioniert, kann jedoch derzeit nicht geboten werden.³¹

Eine langfristige Verfügbarkeit der Quelldaten um u.a. auch spätere Forschung unter neuen Aspekten zu ermöglichen ist dadurch realisiert, dass die Anwendung nach Projektlaufzeit als Standarddienst bei der VZG weiter gewartet und gehostet wird. Ein entsprechender Vertrag, der auch die Bereitstellung der Daten nach Vertragskündigung gewährleistet, ist bereits abgeschlossen.

7 Zusammenfassung

Wir haben gezeigt, dass die untersuchten FAIR-Prinzipien und die Zitierempfehlungen von MyCoRe bis auf die beiden optionalen Richtlinien G10 und G11 technisch sehr gut realisiert sind und somit entsprechende Anforderungen an das Forschungsdatenmanagement umsetzbar sind. Trotzdem verdeutlicht die Übersicht in Tabelle 2, dass die vorgestellten MyCoRe-Anwendungen einige Prinzipien noch nicht vollständig und optimal umsetzen. Folgende Gründe können genannt werden:

³⁰ Liste der standardisierten Musiktermini https://www.uni-muenster.de/imperia/md/content/cmo-edition/publikationen/standard_list_of_musical_terms.pdf.

³¹ Siehe Cremer et al. (2018).

Im OpenAgrar sind einige Prinzipien noch nicht vollständig und optimal umgesetzt (siehe + in Tab. 2), da sie bisher nicht für alle Datensätzen konsequent angewendet werden. Dies liegt zum überwiegenden Anteil in der Verantwortung der Betreiber, z.B. fehlende Verträge mit DOI Registrierungsagenturen (F1) oder fehlende Policies, die obligatorische und freiwillige Metadatenfelder festlegen können (R1, R.1.2, I3).

Im Professorenkatalog werden die intern als XML vorliegenden Daten in Form eines Webportals publiziert. Dadurch wird eine klare Abgrenzung von Daten und Metadaten erschwert und es besteht die Gefahr, dass die genannten Prinzipien nicht ausreichend Beachtung finden. Viele der zur Erfüllung der Kriterien notwendigen Informationen sind in den Daten enthalten und müssen lediglich in Form von zusätzlichen Angaben auf der Webseite oder die Aufbereitung in maschinenlesbarer Form explizit gemacht werden. Für die Umsetzung stellt MyCoRe die notwendige Funktionalität bereit. Aus unserer Sicht ist auch eine stärkere Sensibilisierung der Fachwissenschaftler für dieses doch eher technische Thema notwendig.

Der CMO-Editionen und Quellen-Katalog ist die jüngste der hier vorgestellten Anwendungen und sowohl inhaltlich als auch funktional noch im Aufbau befindlich. Auch wenn MyCoRe aus technischer Sicht alles mitbringt, was für die Umsetzung der hier evaluierten Kriterien notwendig ist, sind einige Funktionen nur teilweise oder gar nicht umgesetzt.

Abschließend können wir sagen, dass Forschungsdaten und Metadaten in MyCoRe-Repositoryen potentiell FAIR sind. Es kommt aber darauf an, dass in den Projekten, die den Aufbau eines Repositoriums planen und realisieren, kontinuierlich an der Einhaltung der Prinzipien mitgewirkt wird. Dies setzt eine stärkere Sensibilisierung der Fachwissenschaftler bzw. Repository-Betreiber für die FAIR-Prinzipien voraus. Notwendig ist auch eine kontinuierliche Weiterentwicklung der Software, wie dies in der MyCoRe-Entwicklergemeinschaft gepflegt wird. Da die grundlegenden Prinzipien der Auffindbarkeit, Zugänglichkeit, Interoperabilität und Wiederverwendbarkeit von (Meta-)Daten in Repositoryen schon bei der Entwicklung von MyCoRe eine wichtige Rolle gespielt haben, sind viele technische Möglichkeiten schon lange verfügbar und bei entsprechenden institutionellen Publikationsservern mit diversen digitalen Objekten umgesetzt. Mit MyCoRe-Anwendungen kann so ein Beitrag zum Aufbau einer langfristig stabilen, vertrauenswürdigen und verlässlichen Forschungsdateninfrastruktur geleistet werden.

Tab. 2: Bewertungen anhand der FAIR-Prinzipien und Richtlinien für Repositorien

| Bewertungskriterium | | MyCoRe | OpenAgrar | Professoren-katalog | CMO |
|---------------------|--|--------|-----------|---------------------|-----|
| F1 | (Meta-)Daten erhalten global eindeutige und dauerhafte PIDs | + | ++ | ++ | ++ |
| F2 | Beschreibung der Daten mit umfangreichen Metadaten | + | ++ | ? | ++ |
| F3 | Klare Referenz von Metadaten zu Daten mittels ID | + | ++ | ++ | ++ |
| F4 | Metadaten sind in durchsuchbaren Verzeichnisdiensten erfasst | + | ++ | + | + |
| A1 | Auffindbarkeit der (Meta-)Daten über ein standardisiertes Protokoll | + | ++ | ++ | ++ |
| A1.1 | Protokoll ist offen, frei und universell | + | ++ | ++ | ++ |
| A1.2 | Protokoll unterstützt Authentifizierung und Rechteverwaltung | + | ++ | + | ++ |
| A2 | Metadaten sind/bleiben verfügbar | + | ? | – | + |
| I1 | Nutzung etablierter Formalismen zur Präsentation der (Meta-)Daten | + | ++ | ? | ++ |
| I2 | Nutzung FAIRer Vokabulare in den (Meta-)Daten | + | + | ? | ? |
| I3 | Qualifizierte Referenz zwischen den (Meta-)Daten | + | + | + | + |
| R1 | Detailliert beschriebene (Meta-)Daten mit präzisen und relevanten Attributen | + | + | ? | ? |
| R1.1 | Klare Angabe der Nutzungslizenz | + | ++ | – | ++ |
| R1.2 | (Meta-)Daten enthalten Provenienz-Informationen | + | + | ? | ? |
| R1.3 | (Meta-)Daten entsprechen fachgebietsrelevanten Standards | + | + | ? | ++ |
| G1 | Global eindeutige und dauerhafte PIDs (analog zu F1) | + | ++ | ++ | ++ |
| G2 | PIDs unterstützen mehrere Granularitäten | + | ? | ? | ? |
| G3 | PIDs lösen auf die sogenannte „Landing Page“ (LP) auf | + | ++ | ++ | ++ |
| G4 | PID ist maschinenlesbar in LP eingebettet | + | ++ | – | ++ |
| G5 | Dokumentation und Unterstützung für Zitation der Daten | + | ++ | + | – |
| G6 | LP enthält Metadaten für Zitation und Auffindbarkeit | + | + | ? | + |
| G7 | Metadaten sind entsprechend schema.org in JSON-LD eingebettet | + | ++ | – | – |
| G8 | Metadaten sind über HTML-Tags auslesbar | + | ++ | – | + |

| Bewertungskriterium | | MyCoRe | OpenAgrar | Professoren-katalog | CMO |
|---------------------|--|--------|-----------|---------------------|-----|
| G9 | Download der Metadaten in Bibtex o.ä. möglich | + | ++ | - | - |
| G10 | „Content Negotiation“ (CN) für schema.org/JSON-LD | - | - | - | - |
| G11 | Unterstützung von sogenannten „HTTP Link Headern“ für CN | - | - | - | - |

Legende:
 Kriterium ist vollständig und optimal erfüllt: ++
 Kriterium ist vollständig mit Potential erfüllt: +
 Kriterium ist teilweise erfüllt: ?
 Kriterium ist nicht erfüllt: -

Literaturverzeichnis

Arbeitsgruppe Forschungsdaten (2018): Forschungsdatenmanagement. Eine Handreichung [Arbeitsgruppe Forschungsdaten der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen]. Potsdam: Deutsches GeoForschungsZentrum GFZ. DOI: [10.2312/allianzoa.029](https://doi.org/10.2312/allianzoa.029).

Allianz der deutschen Wissenschaftsorganisationen (2018): Schwerpunktinitiative „Digitale Information. Forschungsdaten“. Verfügbar unter <https://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten>.

Boeckhout, Martin; Zielhuis, Gerhard A.; Bredenoord, Annelien L. (2018): The FAIR guiding principles for data stewardship: fair enough? In: *European Journal of Human Genetics*, 26 (2018), 931–36, DOI: [10.1038/s41431-018-0160-0](https://doi.org/10.1038/s41431-018-0160-0).

Corpus Musicae Ottomanicae (2018): Projekt CMO. Verfügbar unter <https://www.uni-muenster.de/CMO-Edition/cmo/cmo.html>.

Cremer, Fabian; Klaffki, Lisa; Steyer, Timo (2018): Der Chimäre auf der Spur: Forschungsdaten in den Geisteswissenschaften. In: *O-Bib. Das Offene Bibliotheksjournal*, 5 (2), 142–62. DOI: [10.5282/o-bib/2018H2S142-162](https://doi.org/10.5282/o-bib/2018H2S142-162).

DataCite Metadata Working Group (2017): DataCite Metadata Schema for the Publication and Citation of Research Data, Version 4.1. DataCite e.V. DOI: [10.5438/0014](https://doi.org/10.5438/0014).

DFG (2018): Umgang mit Forschungsdaten. DFG-Leitlinien zum Umgang mit Forschungsdaten. Letzte Aktualisierung am 16.05.2018. Verfügbar unter http://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/nachnutzung_forschungsdaten.

Dunning, Alastair; de Smaele, Madeleine; Böhmer, Jasmin (2017): Are the FAIR Data Principles fair? In: *International Journal of Digital Curation*, 12 (2), 177–95. DOI: [10.2218/ijdc.v12i2.567](https://doi.org/10.2218/ijdc.v12i2.567).

- Fenner, Martin; Crosas, Mercè; Grethe, Jeffrey; Kennedy, David; Hermjakob, Henning; Rocca-Serra, Philippe et al. (2016): A Data Citation Roadmap for Scholarly Data Repositories. In: *bioRxiv* 097196. DOI: [10.1101/097196](https://doi.org/10.1101/097196).
- EU (2016): H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020. Verfügbar unter http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.
- Kindling, Maxi; Schirmbacher, Peter (2013): „Die digitale Forschungswelt“ als Gegenstand der Forschung. In: *Information – Wissenschaft & Praxis*, 64 (2-3), 127–36. DOI:[10.1515/iwp-2013-0017](https://doi.org/10.1515/iwp-2013-0017).
- Mons, Barend; Neylon, Cameron; Velterop, Jan; Dumontier, Michel; da Silva Santos, Luiz Olavo Bonino; Wilkinson, Mark (2017): Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. In: *Information Services & Use*, 37, 49–56, DOI: [10.3233/ISU-170824](https://doi.org/10.3233/ISU-170824).
- MEI (2018): The Music Encoding Initiative. Verfügbar unter <https://github.com/music-encoding>.
- Nature (2016): Announcement: Where are the data? In: *Nature*, 537, 138. Verfügbar unter <https://www.nature.com/news/announcement-where-are-the-data-1.20541>.
- Rat der Europäischen Union (2016): Der Übergang zu einem System der offenen Wissenschaft – Schlussfolgerungen des Rates (angenommen am 27.5.2016). Dokumentennummer 9526/16. Verfügbar unter <http://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/de/pdf>.
- Riley, Jenn (2009): Seeing Standards: A Visualization of the Metadata Universe. Verfügbar unter <http://jennriley.com/metadatamap>.
- Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie, Blomberg, Niklas; Boiten, Jan-Willem; da Silva Santos, Luiz Bonino; Bourne, Philip E.; Bouwman, Jildau; Brookes, Anthony J.; Clark, Tim; Crosas, Mercè; Dillo, Ingrid; Dumon, Olivier; Edmunds, Scott; Evelo, Chris T.; Finkers, Richard; Gonzalez-Beltran, Alejandra; Gray, Alasdair J.G.; Groth, Paul; Goble, Carole; Grethe, Jeffrey S.; Heringa, Jaap; 't Hoen, Peter A.C.; Hooft, Rob; Kuhn, Tobias; Kok, Ruben; Kok, Joost; Lusher, Scott J.; Martone, Maryann E.; Mons, Albert; Packer, Abel L.; Persson, Bengt; Rocca-Serra, Philippe; Roos, Marco; van Schaik, Rene; Sansone, Susanna-Assunta; Schultes, Erik; Sengstag, Thierry; Slater, Ted; Strawn, George; Swertz, Morris A.; Thompson, Mark; van der Lei, Johan; van Mulligen, Erik; Velterop, Jan; Waagmeester, Andra; Wittenburg, Peter; Wolstencroft, Katherine; Zhao, Jun; Mons, Barend (2016): The FAIR Guiding Principles for scientific datamanagement and stewardship. In: *Scientific Data*, 3, 160018. DOI: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).
- Wilkinson, Mark D; Sansone, Susanna-Assunta; Schultes, Erik; Doorn, Peter; Bonino da Silva Santos, Luiz Olavo; Dumontier, Michel (2018): A design framework and exemplar metrics for FAIRness. In: *Scientific Data*, 5, 180118, DOI: [10.1038/sdata.2018.118](https://doi.org/10.1038/sdata.2018.118).



Kathleen Neumann

Verbundzentrale des GBV (VZG)
Digitale Bibliothek, DMS
Platz der Göttinger Sieben 1
D-37073 Göttingen
Kathleen.Neumann@gbv.de



Dr. Wiebke Oeltjen

Universität Hamburg
Regionales Rechenzentrum
Service und Anwendungen, MyCoRe
Schlüterstr. 70
D-20146 Hamburg
wiebke.oeltjen@uni-hamburg.de



Dr. Ulrike Stahl

Julius Kühn-Institut (JKI)

Bundesforschungsinstitut für Kulturpflanzen
Informationszentrum und Bibliothek
Erwin-Baur-Str. 27
D-06484 Quedlinburg
ulrike.stahl@julius-kuehn.de



Robert Stephan
Universität Rostock
Universitätsbibliothek
Digitale Bibliothek
D-18051 Rostock
robert.stephan@uni-rostock.de