



Tafelmalerei Digital und FAIR

Mark Fichtner, Tobias Gradl und Canan Hastik

Das Germanische Nationalmuseum (GNM) vereint als größtes kulturgeschichtliches Museum des deutschen Sprachraums vielfältige Sammlungen und Archive, das Institut für Kunsttechnologie und Konservierung sowie die größte öffentlich zugängliche Spezialbibliothek für die deutsche Kulturgeschichte. Neben der Grundlagenforschung am Objektbestand widmet sich das Museum innovativer Tiefenforschung zusammen mit universitären und außeruniversitären Partnern im Rahmen von Drittmittelprojekten. Bei Projektbeginn wird auf dem bereits erfassten Datenbestand aufgebaut und dieser im anschließenden Forschungsprozess verfeinert. Am Ende eines Forschungsprojektes fließen die Ergebnisse zurück in den Objektkatalog und aktualisieren die Erkenntnisse aus der Grundlagenforschung oder reichern sie an. Doch nicht alle Ergebnisse aus den Forschungsprojekten können in dem Objektkatalog verzeichnet werden. Gleichzeitig fordern Drittmittelgeber eine Bereitstellung der Forschungsprimärdaten, so dass diese in einem dezierten Forschungsdatenkorpus gesammelt und zugänglich gemacht werden. Ein dezentraler und verteilter Zugang auf den Museumsgesamtbestand entspricht jedoch nicht den Anforderungen eines zeitgemäßen Forschungsdatenmanagements. Mit dem Ziel, die kooperative digitale Forschung in den Geistes- und Kulturwissenschaften voranzubringen, wurde mit der DARIAH-DE Datenföderationsarchitektur (DFA)¹ eine Infrastruktur für heterogene Forschungsdatensammlungen geschaffen, die mit Hilfe der Data Modelling Environment (DME)² und der Collection Registry (CR)³ die Zentralisierung verteilter Forschungsdatenkorpora unterstützt. Dieser Beitrag eruiert und validiert, ob mit der DFA unterschiedliche Forschungsdatensammlungen am Beispiel der Tafelmalerei gezielt zusammengeführt und somit die Anforderungen an ein zeitgemäßes, an den FAIR-Prinzipien orientiertes Forschungsdatenmanagement erfüllt werden können.

Forschungsdatenzyklus in DARIAH-DE

Eine Aufgabe von Gedächtnisinstitutionen ist es, Materialien und Objekte unserer kulturellen Aufzeichnung zu sammeln, zu sortieren und zu speichern. Es ist allerdings enorm aufwendig, diese Materialien für die Forschungsgemeinschaft und die breite Öffentlichkeit zugänglich zu machen. Der größte Anteil aller Kultursammlungen wird in Depots verwahrt und nur ein Bruchteil ist bisher digital zugänglich.⁴ Dies ist in erheblichem Maße auf eine extreme Selektivität

1 <https://de.dariah.eu/data-federation-architecture> (08.10.2019).

2 <http://genericsearch.gnm.de/dme/> (08.10.2019).

3 <http://genericsearch.gnm.de/colreg/> (08.10.2019).

4 <https://www.nytimes.com/2009/03/19/arts/artsspecial/19TROVE.html> (08.10.2019).

zurückzuführen und wird verstärkt von der Perspektive, dass Sammlungsvolumen und -vielfalt weiter zunehmen werden. Es ist ungewiss, ob diese Gesamtheit irgendwann online zugänglich gemacht werden kann.⁵ Umso wichtiger ist es, die bestehenden Repositorien offen und vernetzt dauerhaft zugänglich zu machen, um nicht nur die Nachnutzung für Kulturförderung und -entwicklung zu gewährleisten, sondern auch den wissenschaftlichen Diskurs zu sichern und zu befördern.

In den letzten Jahren sind hierfür bereits diverse technische Ansätze und Lösungen entwickelt worden, die museumsübergreifend⁶ oder fach- und institutionsübergreifend⁷ vernetzte generische Suchmöglichkeiten zur Verfügung stellen. Das Problem der sich jeweils stark in Breite und Granularität unterscheidenden Metadatenformate bleibt jedoch inhärent bestehen, skaliert sich bei zunehmender Vielfalt digitaler und digitalisierter Forschungsgegenstände und ist zudem technisch noch nicht vollständig gelöst. Mit *DARIAH-DE*⁸ wurden nicht nur disziplinübergreifende geistes- und kulturwissenschaftliche Infrastrukturlösungen entwickelt, sondern auch fachspezifische Bedarfe zur Entwicklung einer nationalen Forschungsdateninfrastruktur⁹ formuliert und somit sowohl technische als auch methodische Anforderungen für künftige Lösungsansätze umrissen. Dabei ist mit Blick auf das breite Feld der Geistes- und Kulturwissenschaften der Anspruch, neben sprach- und textbasierten Sammlungen auch nicht-textuelles, immaterielles, bildbasiertes Kulturmaterial und archäologische Artefakte nach den *FAIR*¹⁰-Prinzipien interoperabel, auffindbar, zugänglich und nachnutzbar zu machen.

Digitales Forschungsdatenmanagement am Germanischen Nationalmuseum

Das GNM präsentiert seine Objektbestände in seinem digitalen Objektkatalog,¹¹ der sich aus dem hauseigenen Inventarisierungssystem *Document Management System* (DMS) speist. Die selektive Tiefenforschung im Rahmen von Forschungsprojekten stellt sehr spezifische Anforderungen an die Objekterschließung und -verwaltung, die in dem engen Korsett des Inventarisierungssystems nur rudimentär erfüllt werden können. Aus diesem Grund kommt seit 2012 flächendeckend und konsequent die *Wissenschaftliche KommunikationsInfrastruktur* (WissKI)¹² für diesen Bereich zum Einsatz.

5 <https://blog.staedelmuseum.de/digitale-projekte-sind-nie-abgeschlossen-die-neue-digitale-sammlung/> (08.10.2019).

6 <https://blog.staedelmuseum.de/schlendern-suchen-teilen-die-staedel-digitale-sammlung/>, (08.10.2019).

7 <https://www.deutsche-digitale-bibliothek.de/> (08.10.2019).

8 <https://de.dariah.eu/> (08.10.2019).

9 <https://www.dfg.de/foerderung/programme/nfdi/absichtserklaerungen/index.html> (08.10.2019).

10 <https://www.force11.org/group/fairgroup/fairprinciples> (08.10.2019).

11 <http://objektkatalog.gnm.de> (08.10.2019).

12 http://risources.dfg.de/detail/RI_00397_de_druck.pdf (08.10.2019)

Zu Projektbeginn werden typischerweise die bestehenden Inventardaten zu den ausgewählten Objekten des Forschungsprojektes aus dem DMS exportiert und in das WissKI-basierte Projektsystem importiert. In der zumeist dreijährigen Projektlaufzeit werden diese Daten nun durch innovative Forschungsmethoden korrigiert, angereichert, erweitert und ergänzt. Die im Rahmen des Forschungsprozesses gewonnenen neuen Erkenntnisse zum ausgewählten Objektbestand werden in einer auf dem Projektsystem basierenden eigenen Forschungsplattform verzeichnet und für die Öffentlichkeit aufbereitet präsentiert. Das zentrale, langfristig gepflegte und von der Sammlungsleitung kuratierte System zur Datenhaltung bleibt jedoch nach wie vor das DMS. Entsprechend werden Erkenntnisse, für die es im digitalen Objektkatalog geeignete Beschreibungsmöglichkeiten gibt, manuell übernommen. Die WissKI-basierte Forschungsplattform hingegen repräsentiert zum Ende des Forschungsprojektes einen solitären, komplexen und detaillierten Informationsstand, der in der Regel nicht mehr verändert wird. Konsequenterweise stellt die Forschungsplattform somit weiterführende, neue und innovative Informationen zur Verfügung, die in Granularität und Breite über jene im Objektkatalog hinausgehen.

Für Nutzende, ob mit wissenschaftlichem, fachlichem oder öffentlichem Interesse, ist der divergierende Informationsstand nur bedingt sichtbar und nachvollziehbar. Einerseits fehlt es an Verweisen aus dem digitalen Objektkatalog auf die in den Forschungsprojekten generierte Kontextinformation, andererseits sind die entstandenen Forschungsplattformen nur zeitlich begrenzt im Rampenlicht und werden von aktuelleren Projektergebnissen fortlaufend überlagert.

Eine angebots- und systemübergreifende Such- und Recherchemöglichkeit gibt es nicht, mit dem Resultat, dass es für Nutzende sehr schwierig ist, den gesamten Forschungsstand eines Museumsobjektes samt Forschungsprimärdaten und -sekundärdaten zu überblicken und zu erschließen. Insbesondere ältere Projektergebnisse sind daher bedeutend schwerer auffindbar, obwohl sie oft nicht weniger relevant sind.

Während heutzutage kunsttechnologische Untersuchungen des Materials und der Bilddetails im Vordergrund der Forschung stehen, lag der Forschungsschwerpunkt früher auf kunsthistorischen Analysen von Motiven, Mustern, Hintergründen, Text usw. Dabei baut jedes Folgeprojekt auf dem Vorgängerprojekt auf, überprüft aber nur bedingt die bereits erhobene Datenlage und korrigiert und ergänzt somit nur partiell ältere Forschungsergebnisse. Zusätzlich zu der geschilderten heterogenen Datenhaltung in den digitalen Beständen stellt jedes System eine jeweils eigene Such- und Recherchemöglichkeit zur Verfügung. Allenfalls die Sammlungsleitung hat eine vollständige Kenntnis aller Forschungsprojekte und Kenntnis über die Historie der Forschungsobjekte und übernimmt an dieser Stelle eine zentrale Rolle im Sammlungs- und Forschungsdatenmanagement.

Forschung zum Bildträger Tafelmalerei

Im Rahmen des Leibniz-Forschungsprojektes *Deutsche Tafelmalerei des Spätmittelalters. Kunsthistorische und kunsttechnologische Erforschung der Gemälde im GNM* und anschließend im DFG-geförderten Projekt *Spätmittelalterliche Tafelmalerei im Spiegel von Geschichte, Material und Technik. Interdisziplinäre Objektforschung am GNM* entstand ein innovatives Forschungsportal *Deutsche Tafelmalerei des Spätmittelalters*.¹³ Dort finden sich aktuellste kunsttechnologische und objektspezifische Untersuchungen, beispielsweise zu dem Tafelgemälde *Gm1025 von Meister L.Cz.*,¹⁴ der *Kreuzaufladung*.¹⁵ Diese Forschungsprojekte waren jedoch nicht die ersten, die dieses Gemälde im Fokus hatten. Bereits 1995 erforschte das DFG-geförderte, interdisziplinäre Forschungsprojekt *Fränkische Tafelmalerei vor Dürer* dieses Gemälde und machte hierfür zahlreiche Detailaufnahmen, die heutzutage im Nachlass des Projektleiters am GNM zu finden sind. Das Museum digitalisierte die Materialien dieses Forschungsprojektes als Vorarbeit für das anstehende Folgeprojekt und stellte diese Materialien in einem weiteren *Medienarchiv zur fränkischen Tafelmalerei vor Dürer*¹⁶ zur freien Verfügung. Im digitalen *Objektkatalog* des Museums finden sich zu diesem Tafelgemälde *Gm1025 Kreuzaufladung* jedoch nur die Verwaltungsdaten wie Inventarnummer, Titel, Künstler und Herstellungsort sowie Material und Maße. Von den genannten drei Angeboten des Museums verweisen nur zwei gegenseitig aufeinander, nämlich der digitale *Objektkatalog* auf das Forschungsportal *Deutsche Tafelmalerei des Spätmittelalters* und vice versa.

Die folgende Abbildung 2.1 zeigt die gegenseitige Verknüpfung der drei Forschungsdatenportale zur Tafelmalerei am GNM, dabei handelt es sich für um inhaltlich verständliche, aber technisch nicht umsetzbare Verknüpfungen. Das *Medienarchiv zur fränkischen Tafelmalerei vor Dürer* steht solitär, obwohl es auf dieselbe technologische Basis aufbaut und das initiale Medienportal zur Tafelmalerei darstellt. Es enthält jedoch neben den am GNM beherbergten Objekten zum Großteil auch weitere Objekte aus Kirchen und Museen aus dem fränkischen Raum, weswegen eine strukturierte Anreicherung der Informationen zum Teilbestand des GNM bisher nicht erfolgte und die Forschungserkenntnisse lediglich als Forschungsprimärdaten zur Verfügung gestellt werden. Eine gezielte Aufarbeitung des gesamten Forschungsstandes zur Tafelmalerei müsste heute im Nachgang der bereits abgeschlossenen Projekte schließlich durch Hausmittel weitgehend manuell erfolgen. Eine Aufgabe, die in den gängigen Fördermodellen aus innovativer Projektarbeit und Grundlagenforschung bislang nicht erforschter Materialien so nicht vorgesehen ist.

Daher gibt es im GNM neben den miteinander rudimentär und uneinheitlich verknüpften Datenbeständen auch solitär stehende digitale Bestände, die unverknüpft sind. Eine nachträgliche Verknüpfung dieser Einzelbestände zur Gene-

13 <http://tafelmalerei.gnm.de> (08.10.2019).

14 <http://d-nb.info/gnd/138378967> (08.10.2019).

15 <http://objektkatalog.gnm.de/objekt/Gm1025> (08.10.2019).

16 <http://projektdb.gnm.de/medienarchiv-tafelmalerei> (08.10.2019).

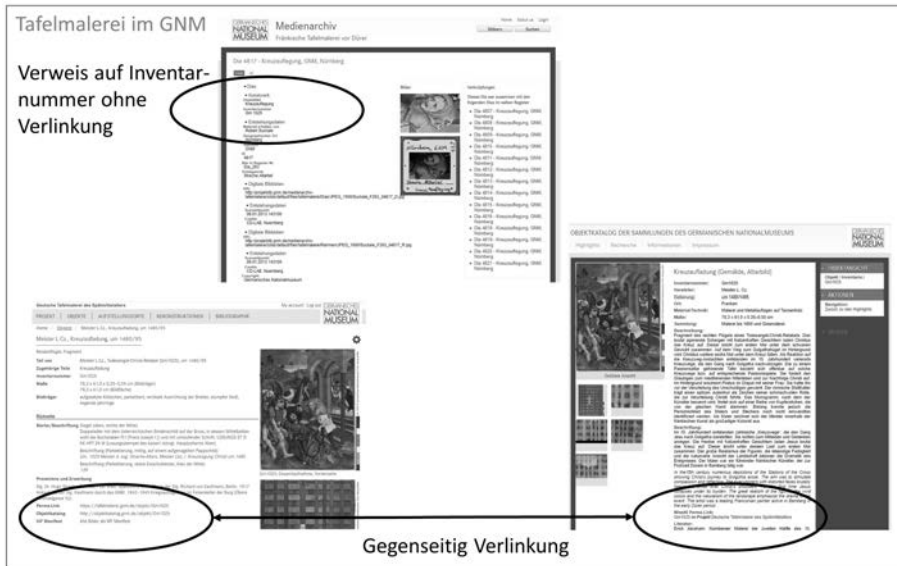


Abb. 2.1: Deutsche Tafelmalerei des Spätmittelalters (links), Medienarchiv Tafelmalerei (Mitte oben), Objektkatalog des GNM (rechts) mit gegenseitiger Verlinkung zwischen den Datenbeständen; Abbildung, Mark Fichtner (CC BY).

rierung eines Überblicks über den Gesamtbestand zur Tafelmalerei im GNM ist ohne umfangreiche händische Nacharbeit nur bedingt umsetzbar. Weiter erschwert wird die Nachnutzbarkeit der Forschungsergebnisse und -materialien dadurch, dass das GNM auch über keine zentrale Such- und Recherchemöglichkeit über die verteilten Forschungsportale hinweg verfügt. Die gesamte Informationslage am Museum ist daher weder angemessen recherchierbar, noch ist eine Qualitäts- und Aktualitätsbeurteilung zu den einzelnen Objekten selbst möglich. Eine Kontaktaufnahme zur Sammlungsleitung bleibt schließlich der einzige Weg, um einen Gesamtüberblick zum Forschungsstand zu erhalten.

Mit Blick auf die Notwendigkeit und Gewährleistung der Nachnutzung von Sammlungs- und Forschungsdaten im Sinne der FAIR-Prinzipien lässt sich für das digitale Datenmanagement am GNM zusammenfassen, dass insbesondere die Forschungsrepositorien auf Basis von WissKI technisch für sich die Prinzipien erfüllen und auch die Schnittstellen für eine disziplinübergreifende Vernetzung und Nachnutzung zur Verfügung stellen. Museumsübergreifend hingegen sind die FAIR-Prinzipien jedoch nicht vollständig auf allen Ebenen abbildbar, da die gesamte Breite und Tiefe an Daten aus rechtlich-politischen, aber auch kuratorischen Gründen nicht uneingeschränkt in andere Kontexte übertragbar ist. So können beispielsweise Provenienzenzen wie Kaufpreis, Käufer und Versicherungswert aus datenschutzrechtlichen Gründen oft nicht vollständig öffentlich

gemacht werden. Zudem muss die Sammlungsleitung für jedes einzelne dargestellte Objekt in der Lage sein, aufgrund mehrerer sich überschneidender Quellsysteme bzw. für einzelne Teilsammlungen spezifischer Objektaspekte oder feingranular auf Ebene der Bildmetadaten eine andere Quelle als den Objektkatalog als normgebend zu bestimmen.

Technische Hintergrundinformation

Das für das GNM entwickelte *Document Management System* (DMS) verwaltet die Inventardaten des Hauses auf Basis gängiger Metadatenstandards wie *LIDO*¹⁷ und *SPECTRUM*.¹⁸ Derzeit umfasst dieses System ca. 80.000 Objekte aus den verschiedenen Sammlungsbereichen. Sie sind formal und inhaltlich mit deskriptiven Metadaten wie Inventarnummer, Material/Technik, Sammlungszugehörigkeit, Beschreibung, Bild/Abbildung, Datierung, Orts- und Künstlerreferenz digital erschlossen und über einen angebundenen digitalen Objektkatalog¹⁹ recherchierbar. Zusätzlich erfolgt ein Literaturnachweis, der die wesentlichen Publikationen zum jeweiligen Objekt auflistet.

Mit dem DFG-finanzierten Projekt *Wissenschaftliche Kommunikationsinfrastruktur* (WissKI) ist von 2009 bis 2011 eine virtuelle Forschungsumgebung für die Anwendung im Bereich der digitalen Geisteswissenschaften entwickelt worden. In einer zweiten Förderphase wurde die Forschungsumgebung in den Jahren 2014 bis 2016 von den Projektpartnern, der Friedrich-Alexander-Universität Erlangen-Nürnberg, dem GNM sowie dem Zoologischen Forschungsmuseum Alexander Koenig, konsequent erweitert. Hauptaspekt der Datenerfassung und -haltung in WissKI sind die Erzeugung und Darstellung semantischer Zusammenhänge durch umfassende Unterstützung aktueller Semantic Web Technologien unter Berücksichtigung der Standards zur Veröffentlichung und Nutzung maschinenlesbarer Daten.²⁰ Die Klassifizierung und Speicherung der erhobenen Daten erfolgt auf Grundlage des *ISO-Standards 21127* (CIDOC CRM).²¹ Somit ist das System ideal für nachhaltige Datenverwaltung ausgelegt und kommt seither aus diesem Grund im GNM in nahezu allen Forschungs- und Ausstellungsprojekten zum Einsatz.

Fast zeitgleich wurde in insgesamt drei Förderphasen in *DARIAH-DE* die DFA entwickelt und aufgebaut. Als eine modulare Infrastruktur unterstützt die DFA die Publikation von Forschungsdaten im *DARIAH-DE Repository*, die Beschreibung wissenschaftlicher Sammlungen durch die *Collection Registry* und die Modellierung von Forschungsdaten mit Unterstützung der *Data Modelling Environment* (DME). Zudem bietet die DFA umfassende Suchfunktionalitäten über

17 <http://www.lido-schema.org/> (08.10.2019).

18 <http://collectionstrust.org.uk/spectrum/> (08.10.2019).

19 <http://objektkatalog.gnm.de/> (08.10.2019).

20 <https://www.w3.org/standards/semanticweb/> (08.10.2019).

21 <http://cidoc-crm.org/> (08.10.2019).

die *Generic Search* (GS).²² Mit der DFA ist es also möglich, Forschungsdaten und Sammlungsbeschreibungen aus verteilten Quellen mit unterschiedlicher Provenienz, inhaltlicher Ausrichtung und Beschreibungstiefe zu aggregieren und auf Basis gemeinsamer Referenzen für Orte, Namen, Daten oder anderer logischer Einheiten, ebenfalls standardisiert auf mittels *Dublin Core* (DC),²³ zu korrelieren und für weiterführende Analysen zugänglich zu machen.

Anforderungen an eine Integration von Bilddatenrepositorien

Unumstritten stellen Forschungsdaten ein wertvolles Gut dar, das nachhaltig gesichert, aufbereitet, zugänglich und nutzbar verwaltet werden muss.²⁴ Insbesondere wissenschaftliche Daten sollten leicht auffindbar, langzeitverfügbar und -zugänglich, mit anderen Datenbeständen interoperabel und für künftige Forschung ordnungsgemäß wiederverwendbar sein. Mit den Prinzipien *Findable, Accessible, Interoperable und Reusable* (FAIR)²⁵ werden die Basisstrategien für eine datengetriebene Forschung und Lehre sowie eine optimale Aufbereitung von Forschungsdaten definiert. Diese Prinzipien werden auf das hier geschilderte Beispielszenario übertragen und daraus die Anforderungen an eine zu etablierende, technische Infrastruktur zur Integration von Bilddatenrepositorien abgeleitet. Die Beschreibung zur technischen Ausgangssituation zeigt, dass die Prinzipien umsetzbar sind, wenn sie nicht bereits in Teilen umgesetzt wurden. Als Grundvoraussetzung für eine Anbindung externer Ressourcen an die DFA gilt, dass die in den Repositorien verzeichneten Daten wenigstens über standardisierte webbasierte Protokolle zugänglich sein müssen.

Anforderung: Übergreifende Suche

Alle Informationsquellen des Museums müssen für die interessierten Nutzer leicht auffindbar und recherchierbar sein, daher ist jedem Datensatz ein global eindeutiger und persistenter Identifier zuzuweisen. Die Informationen sind mit standardisierten maschinenlesbaren Metadaten erschlossen, die den notwendigen Kontext für die Interpretation des Forschungsstandes liefern. Da unterschiedliche Informationsquellen zu einem Objekt einen variierenden Forschungsstand repräsentieren können und der Objektkatalog in der Regel die normgebende Informationsinstanz darstellt, ist es eine zentrale Herausforderung beim Zusammenführen der verteilten Repositorien, die Herkunft der Information nachvollziehbar und transparent zu dokumentieren.

²² <http://genericsearch.gnm.de/colreg/> (08.10.2019).

²³ <https://www.dublincore.org/specifications/dublin-core/dces/> (08.10.2019).

²⁴ <https://www.forschung-und-lehre.de/datengetriebene-forschung-1005/> (08.10.2019).

²⁵ <https://www.go-fair.org/fair-principles/> (08.10.2019).

Anforderung: Digitale Qualitätssicherung

Die verschiedenen Informationsquellen des Museums enthalten teilweise redundante, sich überlagernde, aber auch sich ergänzende Daten. Sie müssen im Rahmen einer digitalen Qualitätssicherung gewichtet und/oder bereinigt werden. Das hier beschriebene Problem manifestiert sich anschaulich an den Bilddaten. So sind einerseits digitale Bilddaten aus einem Forschungsprojekt von 1990 nicht so hochauflösend wie Bilddaten aus einem heutigen Forschungsprojekt, gleichzeitig können sie jedoch damalige Zustände dokumentieren, die heute nicht mehr nachvollziehbar sind. Aus diesem Grund kann nicht davon ausgegangen werden, dass der aktuelle Stand auch der inhaltlich validere ist und nicht notwendigerweise sind alte Bilddaten durch moderne Bilddaten zu ersetzen. Die nachhaltige Verfügbarkeit der Daten muss in jedem Fall langfristig gewährleistet sein.

Anforderung: Maschinell unterstützter Abgleich mit Rückführung

Die Forschungsdaten sollen austauschbar, interpretierbar und (semi-) automatisch mit anderen Datensätzen kombinierbar sein. Da der erste Schritt in der Projektarbeit die Zurverfügungstellung aller bisher erfassten Informationen umfasst, müsste zum Projektende diese Information idealerweise automatisiert zurückgeführt werden. Die Forschungsdatenrepositorien nutzen in der Regel auf CIDOC CRM-basierte Domänen-Vokabulare. Eine weitere Herausforderung besteht darin, dass am selben Objektbestand gleichzeitig mehrere Forschungsprojekte mit unterschiedlichen Forschungsschwerpunkten und unterschiedlicher semantischer Tiefe und Breite arbeiten, was die Bearbeitungsstände nur schwer vergleichbar und derzeit gar nicht automatisch interoperabel mit dem Objektkatalog oder gar mit diesem referenzierbar macht.

Anforderung: Informationsfusion

Forschungsdaten sollen nachnutzbar und mit anderen kompatiblen Datenquellen vergleichbar sein. Ideal wäre eine (semi-) automatische Zusammenführung und Homogenisierung²⁶ der Forschungsdaten anstelle einer Bereitstellung der sortierbaren Aggregation der Suchtreffer aus den verteilten Repositorien. Zum einen hat das Museum den Anspruch, eine autoritative aufgearbeitete Darstellung des jeweiligen Objektes zu präsentieren und nicht mehrere, ungewichtete und verteilte Darstellungen desselben Objektes, die zueinander nicht oder in einer unklaren Relation stehen. Zum anderen wäre es so möglich, den aktuellen Stand eines spezifischen Forschungsschwerpunkts als aggregierte Information zum jeweiligen Objekt zum aktuellen Zeitpunkt ohne zusätzliche Recherchearbeit zur Verfügung zu stellen und zu präsentieren. Eine Fusion der bestehenden Bilddatenrepositorien sollte zu einem integrierten Gesamtsystem dann unter Vermeidung von Duplikationen der Meta- und Bilddaten umgesetzt werden.

26 <http://www.dcc.ac.uk/resources/curation-lifecycle-model> und <https://univerlag.uni-goettingen.de/handle/3/isbn-978-3-86488-032-2/> (08.10.2019).

Das Integrationskonzept von DARIAH-DE

Gemeinsam von dem GNM und DARIAH-DE werden seit 2017 Machbarkeitsstudien zur Anbindung von WissKI-basierten Forschungsdatensammlungen an DARIAH-DE am Beispiel der Tafelmalerei²⁷ durchgeführt. Primäres Ziel dieser Evaluierung ist es, zu ermitteln, ob und in welchem Umfang die Konzepte und Werkzeuge der DARIAH-DE DFA die oben definierten, an die FAIR-Prinzipien angelehnten Anforderungen unterstützen können. Im Rahmen der Machbarkeitsstudie wurde eine prototypische Instanz der DFA installiert. Sie besteht aus den für die Umsetzung einer übergreifenden Suchmöglichkeit erforderlichen *Collection Registry*, DME und *Generic Search* (GS). Zudem wurden die drei oben beschriebenen Forschungsdatensammlungen des GNM verzeichnet, modelliert, verknüpft und somit durchsuchbar gemacht.

Registrierung von Forschungsdatenrepositorien

Das Konzept der DFA definiert einen mehrstufigen Prozess für die Anbindung von Forschungsprimärdaten, der im Ergebnis zu einem verbesserten Zugang zu verteilt gelagerten Datenrepositorien führt und sie so im Rahmen übergreifender Anwendungsfälle zugänglich und nutzbar macht. Hierfür werden in einem ersten Schritt relevante Sammlungen in der *Collection Registry* verzeichnet und auf diese Weise für weitere DARIAH-DE-Dienste sichtbar gemacht. Enthält die Sammlungsbeschreibung Informationen über unterstützte technische Schnittstellen, wie *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH)²⁸ oder *Search/Retrieve via URL* (SRU),²⁹ sind die von der Sammlungen angebotenen Forschungsdaten unmittelbar für die Dienste der DFA zugänglich. Wenngleich die Heterogenität der Daten durch diesen Schritt noch nicht adressiert wird, sind die Inhalte der angebundenen Sammlungen im Rahmen der einfachen Volltextsuche der *Generic Search* bereits auffindbar. Eine Dokumentation und Einführung in die Nutzung der *Collection Registry* mit anwendungsorientierten Empfehlungen ist auf der DARIAH-DE Portalseite zu finden.³⁰

Zusammenführung und Anreicherung der Daten

Die Herstellung integrativer Sichten über heterogene Bestände von Forschungsdaten erfolgt mit Hilfe der DME und umfasst eine Anreicherung der Daten, um deren Interpretierbarkeit außerhalb ihres Erstellungskontextes zu erreichen. Die

27 Mark Fichtner u.a., „Vom Wandel hin zur objektbasierten Forschung im Kontext von DARIAH-DE“, in: *Forschungsinfrastrukturen in den digitalen Geisteswissenschaften : wie verändern digitale Infrastrukturen die Praxis der Geisteswissenschaften?*, hg. von Martin Huber u.a., Frankfurt a.M. 2019, URL: <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:hebis:30:3-519453> (18.12.2019).

28 <https://www.openarchives.org/pmh/> (08.10.2019).

29 <http://www.loc.gov/standards/sru/> (08.10.2019).

30 <https://dfa.de.dariah.eu/doc/colreg/#wie-benutze-ich-die-collection-registry/> (08.10.2019).

Interpretierbarkeit der Forschungsdaten ist eine Datenqualität und abhängig von formalen, semantischen und standardisierten Metadaten-Beschreibungen.

Abbildung 2.2 zeigt eine erste Anreicherung von Quelldaten am Beispiel von Bildmaterial. Hierfür wird zunächst die in den originären Daten vorhandene Bildadresse in ihre Bestandteile zerlegt (*BildUrlsParsen*), um Dateinamen zu ermitteln und Bilder zu beziehen, und um diese entsprechend weiterer Regeln in *BilderLadenMetadatenLesen* zu verarbeiten.³¹ Darüber hinaus werden die ermittelten *Exchangeable Image File Format* (Exif)-Metadaten dem Datensatz zugeführt und erweitern ihn. Die hier stark verkürzte Darstellung der Anreicherung dient der Überprüfung und Konvertierung von Datentypen, der Erkennung von zeitlichen Abfolgen, der Erkennung und Löschung von Dubletten oder der Vervollständigung von Datenlücken. Reichhaltigere Forschungsdaten bieten einen enormen Mehrwert für die Verarbeitung und Nachnutzung. Die Darstellung der auf Beispieldatensätzen ausgeführten Transformationsregel (s. Abb. 2.2, Ergebnisdarstellung links) deutet beispielsweise die unterschiedlichen Datumsansetzungen als eine unmittelbar anschließende Modellierungsaufgabe an. Die Erfahrung mit bereits durchgeführten Integrationsvorhaben zeigt, dass eine Zusammenführung und Anreicherung der Daten durch ihre Modellierung mit dem Ziel einer Explizierung des Erstellungskontextes und Vorbereitung einer Überführbarkeit in unterschiedliche Verwendungskontexte typischerweise iterativ erfolgt. Dies kann zudem die Extraktion weiterer Exif-Metadaten oder auch die Anwendung bildverarbeitender Verfahren für die Verbesserung der Auffindbarkeit bzw. des Vergleichs von Bildern sein.

Die anhand des erarbeiteten Datenmodells verfeinerten und angereicherten Daten können mit Hilfe von Mappings in beliebige Zielmodelle überführt werden. Auswahl und konkrete Gestaltung solcher Zielmodelle sind dabei stark von den betrachteten Verwendungskontexten abhängig und können von einfachen Standards, wie DC über komplexere Metadatenmodelle hin zu spezifischen, tief erschlossenen Modellen von Forschungsdaten reichen. Ein Beispiel für letzteres wäre ein Mapping in unterschiedliche Profile nach den *Text Encoding Initiative (TEI): P5 Guidelines*³². Einfache Metadatenschemata hingegen haben häufig einen sehr breit und heterogen zusammengesetzten Anwendungskontext.

Für die Verarbeitung und Transformation von Daten steht im Rahmen von Mappings derselbe Funktionsumfang zur Verfügung, wie sie auch für die Modellierung des Datenmodells (vgl. Abb. 2.2 oben) eingesetzt werden können. Abhängig vom Verwendungskontext kann beispielsweise die Homogenisierung von Zeitangaben global am Datenmodell oder jeweils für einzelne Zielformate modelliert werden. Beispiele für Mappings finden sich unter anderem in der DME-Instanz des GNM.³³

31 Auf eine Detaillierung der Verarbeitung wird an dieser Stelle verzichtet. Diese kann in der DME-Instanz des GNM im Modell ProvenienzObjekte eingesehen und nachvollzogen werden.

32 <https://tei-c.org/guidelines/p5/> (08.10.2019).

33 <https://dme.mww-forschung.de/> (08.10.2019).

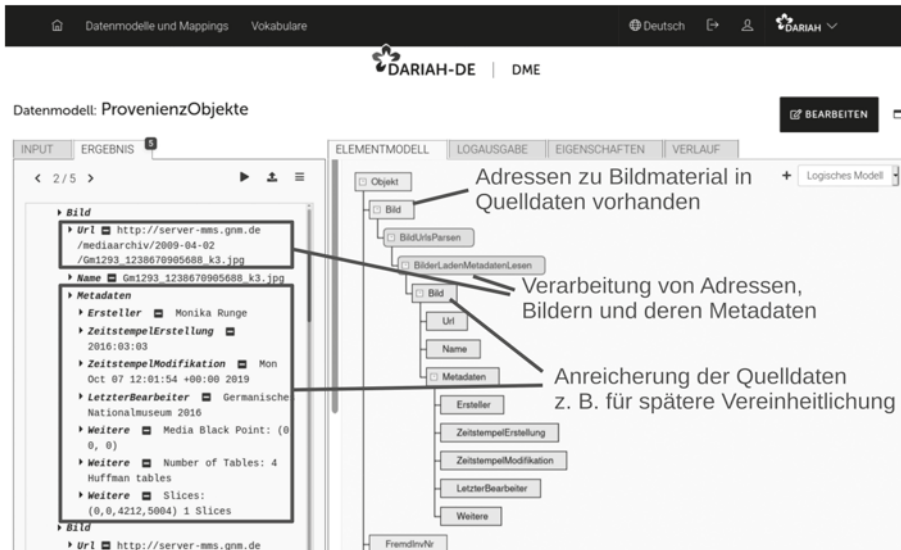


Abb. 2.2: Modellierung von Struktur und Logik zur Extraktion, Verarbeitung und Anreicherung deskriptiver Metadaten; Abbildung, Tobias Gradl (CC BY).

Integrierte Suche der Datenförderationsarchitektur

Ein zur Sammlungslandschaft des GNM ähnliches Anwendungsbeispiel für die DFA findet sich in der gemeinsamen Suche des *Forschungsverbunds Marbach Weimar Wolfenbüttel* (MWW).³⁴ Der seit 2013 durch das BMBF geförderte Verbund umfasst das *Deutsche Literaturarchiv Marbach*, die *Klassik Stiftung Weimar* und die *Herzog August Bibliothek Wolfenbüttel*. Die drei Einrichtungen widmen sich im Rahmen von MWW gemeinsamen der Sammlung, Erschließung, Bewahrung und dem Erhalt deutschsprachiger Ideengeschichte, Literatur und Kultur. Die im Rahmen einer übergreifenden Suche miteinander verbundenen Sammlungen zeichnen sich durch die Anwendung unterschiedlicher Beschreibungsstandards aus. Der Schwerpunkt der Metadatenbeschreibungsformate liegt auf bibliographischen Daten, wie dem *Metadata Object Description Schema* (MODS).³⁵ Für Editionen ist vorwiegend TEI in Verwendung. Insbesondere in Fällen bereits abgeschlossener Sammlungen werden verwendete Standards um eine Vielzahl eigens entwickelter, ergänzender Datenstrukturen angepasst und erweitert.

³⁴ <https://www.mww-forschung.de/> (08.10.2019).

³⁵ <http://www.loc.gov/standards/mods/> (08.10.2019).

Die als *MWW Verbundsuche* eingerichtete Instanz der GS³⁶ bietet eine integrative Sicht auf die unterschiedlichen Sammlungen des Verbunds.³⁷ Die vorwiegend bibliothekarischen Quellen und Editionen, beispielsweise die in TEI vorliegende *Zeitschrift für digitale Geisteswissenschaften*,³⁸ ist in tief integrierter Form im Rahmen der MWW Verbundsuche zugänglich. Viele Sammlungen der Verbundpartner verzeichnen zudem digitalisierte mittelalterliche Handschriften,³⁹ Gemälde und Portraits, vorwiegend aus Nachlässen. Einen Überblick über die derzeit 26 in der *Collection Registry* zugänglich gemachten Sammlungen des Verbunds ist jederzeit online frei abrufbar.⁴⁰

Die drei aufeinander aufbauenden Module der DFA ermöglichen somit grundsätzlich die Föderation verteilter und heterogener Forschungsdatensammlungen. Insbesondere der hier eingeführte Anwendungsfall der *MWW-Verbundsuche* ähnelt dem GNM-Szenario insofern, als die Sammlungslandschaft einen hohen Grad an Heterogenität aufweist, deren Überwindung durch die Autonomie der Organisationen und Sammlungen erschwert wird. Ein grundsätzlicher Unterschied der Szenarien besteht in den unterschiedlichen Herangehensweisen an die Integration der Sammlungen. Während im *MWW-Verbund* zunächst eine möglichst große Zahl relevanter Sammlungen in eher geringer semantischer Tiefe verbunden und in einer gemeinsamen Suche angeboten wird, besteht das wesentliche Ziel des hier skizzierten Szenarios zur Tafelmalerei in einer weitreichenden Fusion der relevanten Bestände. Diese Fusion muss nicht nur Informationsverluste ausschließen, sondern möglichst eine Optimierung der Datenqualität, Informationsqualität und einen Mehrwert gegenüber den heterogenen, verteilten Repositorien erzielen. Der Integrationsprozess des *MWW-Verbunds* ist ein fortlaufender, iterativer Prozess und profitiert gleichzeitig von den Entwicklungen und dem Erkenntnisgewinn aus dem hier beschriebenen Anwendungsbeispiel zur Tafelmalerei.

Zusammenfassend ermöglicht die GS als Suchmaschine die Suche in digital verfügbaren und in der *Collection Registry* registrierten Forschungsdatensammlungen. Die einfache Volltextsuche kann hierbei sämtliche Sammlungen durchsuchen, für eine Facettierung im Rahmen der erweiterten Suche oder die Anwendung von Suchfiltern wird auf relevante Datenmodelle in der DME zurückgegriffen. Für übergreifende Suchen sind zudem Mappings zwischen den Modellen erforderlich, da mit deren Hilfe erst eine Transformation von Suchanfragen möglich wird.

36 <https://search.mww-forschung.de> (08.10.2019).

37 Timo Steyer und Tobias Gradl. *A Research-oriented and Case-based Data Federation for the Humanities*, 2019, URL: <https://hcommons.org/deposits/item/hc:23443> (08.10.2019).

38 <http://www.zfdg.de> (08.10.2019).

39 Torsten Schaßan und Timo Steyer, „Vom lokalen Bestand zur weltweiten Vernetzung. Mittelalterliche Handschriften im Netz“, in: *Das Mittelalter* 24/1 (2019), S. 188–204: <https://doi.org/10.1515/mial-2019-0013> (21.10.2019).

40 <https://colreg.mww-forschung.de/colreg/collections/> (08.10.2019)

Die bisherige Schwerpunktsetzung von *DARIAH-DE*, auch im Rahmen der Kollaboration mit MWW, liegt primär auf textuellen Ressourcen. Die digitale Erschließung von Objekten in der DME erfolgt häufig in einer für Bibliotheken typischen Fokussierung auf Dokumente und Textformen. Metadaten werden dabei ebenso erschlossen wie freie textuelle Beschreibungen und Inhalte zum und über das Forschungsobjekt selbst. Das Modellierungskonzept der DME sieht daher die Möglichkeit vor, neben Forschungs- und Metadaten insbesondere auch Hintergründe des Erstellungskontexts der Daten in den Datenmodellen zu explizieren, um hierdurch die spätere Nachnutzung dieser Daten in unterschiedlichen Verwendungskontexten zu erleichtern. Kontextinformationen wie Bezüge zwischen Objekten, Orten oder Personen und Zeiten werden in der DME als deskriptive Metadaten des betrachteten Forschungsobjektes verwaltet. Dieses Vorgehen bindet somit sämtliche verfügbare Kontextinformationen an das Dokument und ermöglicht die Anwendung textanalytischer Verfahren, die insbesondere auf großen Datenmengen effizient eingesetzt werden können. Die *Generic Search* nutzt dies zur Transformation, Anreicherung und Bereinigung von Daten bei deren Verarbeitung und Indexierung. Mit Hilfe von klassischen Werkzeugen des Text-Retrievals kann die *Generic Search* eine einfache Volltextsuche über die heterogenen und verteilten Sammlungen bereitstellen.

Anwendungsbeispiel Tafelmalerei

Auf Basis des *DARIAH-DE* Integrationskonzeptes können die verschiedenen Forschungsrepositorien des GNM, hier mit dem Fokus auf der Tafelmalerei, übergreifend durchsuchbar gemacht werden.

Abbildung 2.3 zeigt Ergebnisse zu einer Suchanfrage in der für das GNM implementierten Instanz der GS. Die Anfrage wurde hierbei als einfache Volltextsuche ausgeführt und führt zu Resultaten in drei aus derzeit vier prototypisch registrierten und modellierten Sammlungen. Ergebnisse aus den WissKI-Systemen *Medienarchiv – Fränkische Tafelmalerei vor Dürer* und *Die deutsche Tafelmalerei des Spätmittelalters* werden ergänzt durch Treffer aus dem übergreifenden digitalen Objektkatalog des GNM. Die Suchanfrage *gm1025 OR "gm 1025"* adressiert die verteilten Datensätze des Beispiel-Tafelgemäldes *Gm1025*. Der Suchausdruck entspricht der Notation der *Generic Search* und der Identifizierung unterschiedlicher Schreibweisen von Inventarnummern, wobei die zweite Variante als sogenannte Suchphrase aus zwei Termen zusammengesetzt wird. Werden auf diese Weise unterschiedliche Ansetzungen für Inventarnummern ermittelt, können mit dieser Erkenntnis iterativ die Modellierung weiter optimiert und die Informations- und Datenqualität weiter verbessert werden. Diese Suchanfrage liefert sowohl für homogenisierte Daten als auch bei noch bestehender Heterogenität, beispielsweise Dopplung, jeweils sämtliche Datensätze.

Dieses einfache Szenario einer übergreifenden Suche verdeutlicht bereits die Synergiepotenziale der *DARIAH-DE* Föderationsarchitektur. Die Verschränkung von dem inhaltlichen Fokus textbasierter Dokumente in *DARIAH-DE* und der

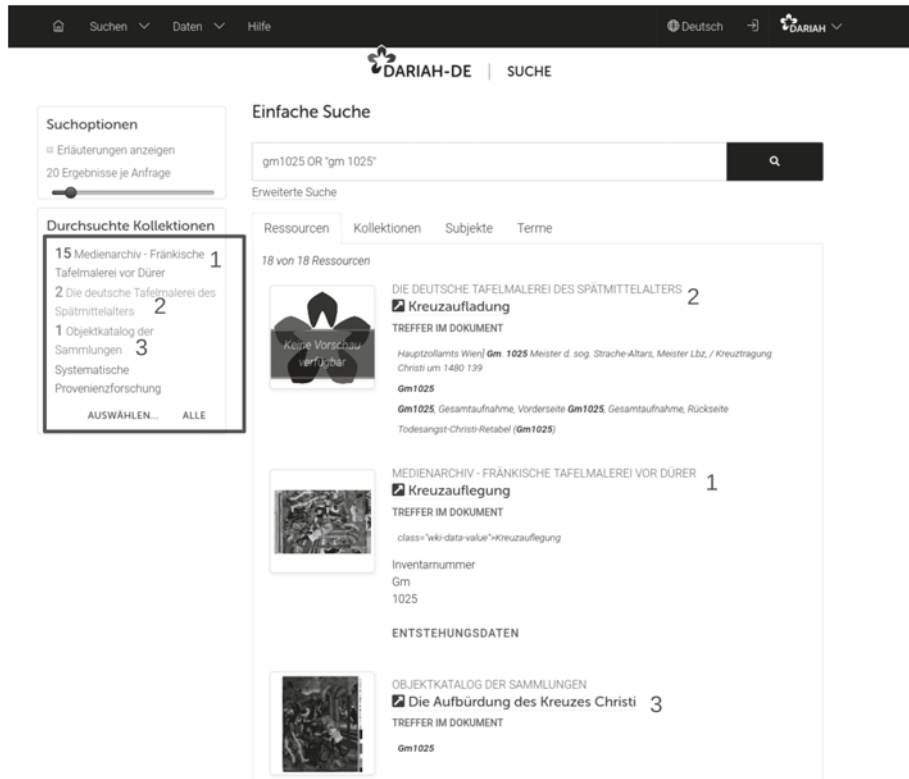


Abb. 2.3: Suche anhand einer bekannten Inventarnummer mit Ergebnissen aus verschiedenen Datenquellen; Abbildung, Tobias Gradl (CC BY).

objektzentrierten Information in WissKI wird anhand folgender drei Anwendungsszenarien erörtert.

Im Rahmen des ersten Szenarios wurde eine übergreifende Suche ermöglicht. Die Forschungsrepositorien des GNM können bislang nicht übergreifend durchsucht werden. Zwar sind die Datenbestände der jeweiligen Repositorien miteinander vergleichbar und basieren auch auf einem gemeinsamen Schema, so dass prinzipiell Heterogenität zwischen WissKI-basierten Repositorien überwunden werden kann. Der *Objektkatalog* und weitere, nicht-semantische Datenquellen bleiben jedoch von dieser Art der Integration zunächst ausgeschlossen. An dieser Stelle bietet die DFA ein einfaches Lösungsmodell, das exemplarisch für die Tafelmalerie bereits umgesetzt ist.

Für das zweite Szenario soll neben den jeweiligen Forschungsportalen ein weiterer Sammlungszugang geschaffen werden. Eine, wenngleich deutlich flachere, übergreifende Suche, bietet neben der Traversierung einen zusätzlichen Einstieg in die integrierten Sammlungen. Beide Varianten ergänzen sich, indem von

Suchtreffern auf Detailansichten weiternavigiert werden kann. Ebenso könnte aus einer solchen Ansicht direkt eine Suche, beispielsweise nach Objekten mit ähnlichen Titeln oder der gleichen Klassifikation, angeboten werden.

Abschließend geht es im dritten Szenario um die weitere Anreicherung der Ergebnisse auf semantischer Ebene. Die Formulierung der Suchanfrage lässt erkennen, dass Daten zum derzeitigen Entwicklungsstand der DFA nicht als semantische Konzepte aufbereitet werden. Durch eine entsprechende Anpassung von Datenmodellen in der DME könnte das Format der Inventarnummern zwar problemlos homogenisiert werden, sie stünden aber auch dann weiterhin nur in Form deskriptiver Metadaten zur Verfügung. Durch die Integration semantischer Konzepte, beispielsweise CIDOC CRM, könnte der Bezug zwischen Datensätzen mit gleicher Inventarnummer auch explizit gemacht werden.

Grenzen bisheriger Anwendungsszenarien

Bis zu der in diesem Beitrag beschriebenen und diskutierten Machbarkeitsstudie wurden in den bisherigen von *DARIAH-DE* umgesetzten Anwendungsszenarien alle Forschungsobjekte, Annotationen und assoziierten Daten ausschließlich in disjunkten Repositorien vorgehalten. Es ist üblich, im Sinne der Abbildung der Realität, äquivalente Objekte und Datensätze in mehreren Repositorien und Sammlungen gleichzeitig zu dokumentieren und zu erschließen. Duplikate entstehen aber erst dann, wenn mehrere Repräsentationsformen desselben Objektes (z.B. in TEI, MODS und DC) zur Verfügung stehen und unabhängig voneinander indiziert werden.

Eine für *DARIAH-DE* neue Interpretation von Datensätzen als verteilte Teilbeschreibungen eines Gesamtobjektes eröffnet neue Anforderungen an die DFA, die zunächst konzeptuell auf der Ebene von Datenmodellen durch die Berücksichtigung semantischer Bezüge unterstützt wird. In einem ersten Schritt werden Orte, Personen und Zeitangaben als eigenständige Objekte gehandhabt, was eine Darstellung von möglichen Bezügen oder auch die Suche nach diesen ermöglicht. In der GS selbst wird diese Ergänzung auf technischer Ebene dadurch unterstützt, dass die existierende und bewährte Dokumentindex-Struktur durch eine graphorientierte Speicherstruktur flankiert wird. Technisch bleibt für die Suche die Auffindbarkeit anhand deskriptiver Metadaten bestehen, wodurch mit den Dokumenten auch die Eigenschaften verknüpfter Objekte redundant zu indizieren sind. Würden im Sinne normalisierter Daten lediglich Identifikatoren zu verknüpften Objekten an den Dokumenten indiziert, so stünden auch nur diese in den Volltextindizes zur Verfügung.

Durch die Verfügbarkeit von Text- und Bildmaterial entsteht durch das Sammeln und Indexieren der verfügbaren Daten nach und nach ein umfassender Datenpool. Für die Anwendung bildverarbeitender Verfahren, beispielsweise zur Erkennung unterschiedlicher Varianten desselben Bildes (Fotografie, Graustufen, Röntgenaufnahme, Ausschnitte etc.), liegen zudem die Aufnahmen bereits an ei-

ner zentralen Stelle, wodurch die Erprobung und Evaluation entsprechender Verfahren erheblich erleichtert wird.

Konstruktive Kritik und offene Aufgaben

Die DFA erfüllt teilweise per se die definierten Anforderungen zur Integration von Bilddatenrepositorien. Die DFA ermöglicht nicht nur eine zentrale Durchsuchbarkeit und Nachnutzbarkeit des Datenbestandes von verteilten Repositorien, sondern macht auch die Herkunft und Aktualität der Daten nachvollziehbar. Die Aktualität der Daten könnte durch eine Zeit- oder Epochenangabe, die Projektlaufzeit oder ein konkretes Erstellungsdatum explizit gemacht werden. Dies wäre einfach umsetzbar und Letzteres wird von den WissKI-basierten Repositorien bereits unterstützt.

Durch die vielen Darstellungen desselben Objektes in den unterschiedlichen Forschungssystemen ergibt sich dabei das zentrale Problem, dass nicht jedes System auf die anderen verweist. So wird oft von den Forschungsprojektsystemen die Verlinkung hin zum Objektkatalog gewährleistet, weil dieser die Bezugsquelle und die nachhaltige Infrastruktur im Museum ist. Die verschiedenen Forschungssysteme hingegen besitzen untereinander nur Verweise, wenn die Forschungsinhalte den gleichen Kontext repräsentieren. Eine leistbare Erweiterung wäre es, den Gesamtforschungsstand in Form aller Suchtreffer basierend auf der DFA zu aggregieren und zusätzlich weiterführende Information anzubieten, um so ein vollständiges Bild zur Datenlage am Museum zu erhalten. Darüber hinaus ist der Anspruch des Museums, die autoritative Norm für die Bereitstellung von Informationen zu den hauseigenen Objekten zu sein. Daher stellt sich die Frage, wie der Kurationsbedarf zur Bereitstellung der Forschungsdaten in der DFA abgebildet werden kann. Jedem Datensatz ist ein global eindeutiger und persistenter Identifier zuzuweisen und die Sammlungsleitung muss in der Lage sein, bei mehreren, sich überschneidenden Quellsystemen, Teilsammlungen, Objektinformation oder sogar einzelnen Teildaten eine andere Quelle als den Objektkatalog als normgebend zu bestimmen. Sämtliche Informationen zum Objekt, zu dem aktuellen Zustand der Informationslage sowie frühere Zustände und Erkenntnisse werden somit recherchierbar. Die Pluralität von Treffern eines konkreten Objektes in einem museumsübergreifenden Portal erfüllt diese Anforderung nicht.

Eine automatisierte Zusammenführung der Informationen zu einem Objekt bleibt derzeit noch ungelöst. Zwar ist durch die Inventarnummer eine eindeutige Identifikation des Objektes oft gewährleistet. Dadurch kann im Prinzip für jedes Objekt auch eine jeweils eigene Website angeboten werden, auf der sämtliche Informationen zu diesem Objekt gebündelt veröffentlicht werden. Schwieriger automatisiert nachvollziehbar bleibt jedoch, welche Quelle den aktuellsten Informationsstand für das Objekt als Ganzes oder in Ausschnitten bereitstellt und ihn über mehrere Forschungssysteme hinweg zu einem homogenen, widerspruchs-

freien Gesamtdatensatz zu vereinen. Ohne menschliche Unterstützung ist beides derzeit nicht zu leisten.

An diesem Punkt sind innovative Ansätze gefragt, die idealerweise eine dynamische Berechnung des Informationsstandes auf Grundlage der Aktualität der Information in Verbindung mit dem Fokus der diversen Forschungsprojekte ermöglichen. So ist ein Forschungssystem für kunsttechnologische Untersuchungen in aller Regel eine bessere Quelle für Informationen über kunsttechnische Daten, als der *Objektkatalog* – allerdings nur bis die Information aus dem Forschungssystem durch ein neueres abgelöst wird. Jedoch ist auch dann die ‚alte‘ Information als Vergleichsmaterial noch wertvoll und sollte transparent, sichtbar und unter Kennzeichnung einer Art Versionierung zugänglich bleiben.

Die Ergebnisse im globalen Kontext

Eingangs wurden bereits die FAIR-Prinzipien eingeführt. Bislang unbetrachtet blieb jedoch, ob die resultierende Infrastruktur diesen Prinzipien genügt und inwiefern die in diesem Beitrag erörterten Anforderungen diesen Prinzipien unterliegen. Die DFA ist als gemeinsame Infrastruktur dabei von den einzelnen, angebundenen Repositorien abhängig. Nur wenn diese die FAIR-Prinzipien erfüllen, kann auch die DFA diese Eigenschaft ‚erben‘. Liegen Forschungsdaten in einem nach den FAIR-Prinzipien angelegten Repository vor, bleiben die Eigenschaften auch im Kontext der DFA erhalten. Die einzelnen Forschungsdaten werden jedoch aus dem Kontext des Forschungssystems somit in einen globaleren und generalisierten Kontext mit anderen Forschungsdaten übernommen. Entsprechend sind manche der FAIR-Prinzipien nicht mehr so einfach zu erfüllen, z.B. können sich die gemeinsam anerkannten Metadatenstandards in einem einzelnen Fachsystem von denen einer übergeordneten Infrastruktur unterscheiden. Hierfür stellt die DFA teilweise bereits notwendige Werkzeuge zur Verfügung, andere, wie z.B. die Nutzung von global eindeutigen Identifikatoren für Suchtreffer, gilt es hingegen noch zu optimieren und zu erweitern.

Die Beurteilung, ob die FAIR-Prinzipien von Forschungsdaten erfüllt werden oder nicht, hängt dabei aber immer von dem Kontext ab, in den die Forschungsdaten eingebunden sind. Während sie im Forschungssystem vollständig erfüllt sein können, können sie im Zusammenspiel mit anderen Forschungsdaten schon nicht mehr hinreichend beschrieben sein, um den gemeinsamen Kontext verständlich zu machen. Insbesondere im Kontext des World Wide Web braucht es weitaus mehr Zusatzinformationen.

Als Fazit zeigt sich also, dass je größer der Kontext der jeweiligen Forschungsinformation, desto schwieriger ist es, die FAIR-Prinzipien vollständig zu erfüllen. Die Lösung des hier beschriebenen Problems ist aktuell eine große Herausforderung für existierende Initiativen auf europäischer, nationaler und regionaler Ebene. Aus unserer Perspektive kann dieses Problem jedoch nur schrittweise pro Ebene und aus dem Kontext der jeweiligen Sammlung heraus gelöst werden, denn nur dort existiert die Kenntnis darüber, wie die Forschungsdaten zu veror-

ten sind. Das Problem muss also zuerst in den jeweiligen ‚Häusern‘ gelöst sein, um übergeordnet, z.B. auf Ebene der *Nationalen Forschungsdateninfrastruktur* (NFDI), der *Deutschen Digitalen Bibliothek* oder *Europeana* gelöst werden zu können.

In diesem Beitrag wurden am Beispiel der Tafelmalerei aktuelle Bedarfe des Forschungsdatenmanagements formuliert und ein Ansatz zur Informationsfusion mittels *DARIAH-DE* DFA erörtert und diskutiert. Während des Lösungsprozesses wurden jedoch zusätzliche Fragen aufgeworfen, die im Hinblick auf die künftige Forschung in den digitalen Geistes- und Kulturwissenschaften von besonderem Interesse sind. Das in diesem Beitrag vorgestellte Anwendungsbeispiel zur Integration von Bilddatenrepositorien des Germanischen Nationalmuseums liefert nur eine partielle Lösung. Das langfristige Ziel, heterogene Quellen übergreifend und vollständig nach den *FAIR*-Prinzipien zu integrieren und zu homogenisieren, bleibt erhalten.

Literaturverzeichnis

- Fichtner, Mark, Tobias Gradl und Canan Hastik, „Vom Wandel hin zur objektbasierten Forschung im Kontext von DARIAH-DE“, in: *Forschungsinfrastrukturen in den digitalen Geisteswissenschaften: wie verändern digitale Infrastrukturen die Praxis der Geisteswissenschaften?*, hg. von Martin Huber, Sybille Krämer und Claus Pias, Frankfurt a.M. 2019, S. 99–110, URL: urn:nbn:de:hebis:30:3-519453 (18.12.2019).
- Gradl, Tobias und Timo Steyer, *A Research-oriented and Case-based Data Federation for the Humanities*, 2019, URL: <http://dx.doi.org/10.17613/d9j7-9q83> (08.10.2019).
- Schaßan, Torsten und Timo Steyer, „Vom lokalen Bestand zur weltweiten Vernetzung. Mittelalterliche Handschriften im Netz“, in: *Das Mittelalter* 24/1 (2019), S. 188–204, URL: <https://doi.org/10.1515/mial-2019-0013> (08.10.2019).