

Jan Horstmann

FORSCHUNGSVERBUND MARBACH WEIMAR WOLFENBÜTTEL

OFFENE DATEN UND TOOLS FÜR DIE DIGITALE
GEISTESWISSENSCHAFTLICHE FORSCHUNG UND LEHRE

1. Drei Standorte digital zusammenbringen: Herausforderungen und Chancen

Der Wissenschaftsrat sprach 2011 die Empfehlung aus, die drei renommierten sammlungsführenden Häuser in Marbach, Weimar und Wolfenbüttel – das Deutsche Literaturarchiv, die Klassik Stiftung und die Herzog August Bibliothek – sollten »unter Beibehaltung ihrer rechtlichen Selbständigkeit in einem von den Einrichtungen selbst mitzugestaltenden Verbund zusammengeführt werden.«¹ »Dabei sollte eine Verbesserung der internationalen Sichtbarkeit und Vernetzung als eines der wichtigsten Ziele gelten. Die drei Einrichtungen sollten ein gemeinsames Stipendienprogramm entwickeln und gemeinsame Lösungen für die Langzeitarchivierung und Speicherung großer Datenmengen suchen.«² Vor diesem Hintergrund fördert das Bundesministerium für Bildung und Forschung (BMBF) seit 2013 den »Forschungsverbund Marbach Weimar Wolfenbüttel (MWW)« als Drittmittelprojekt mit dem zentralen Auftrag, ein Verstetigungskonzept zu erarbeiten.³ Der Anspruch, neue Standards in der digitalen Sammlungsforschung zu etablieren, wurde 2019 mit Einleitung der zweiten Förderphase und der damit verbundenen Einrichtung eines Digitalen Labors geisteswissenschaftlicher Forschung zum konzeptionellen Kern dieser Verstetigung erklärt. Die verschiedenen Forschungsprojekte der drei Standorte sollen methodisch-digital zusammengeführt und damit gleichzeitig paradigmengestaltend für eine ortsübergreifende kollaborative Forschungsarbeit werden.

1 Wissenschaftsrat: Stellungnahme zur Klassik Stiftung Weimar, Jena 2011, S. 17.

2 Wissenschaftsrat: Stellungnahme zur Herzog August Bibliothek (HAB) Wolfenbüttel, Würzburg 2013.

3 Die bisherigen Arbeiten und weitere Entwicklungen des Forschungsverbunds MWW können auf der verbundeigenen Webseite nachvollzogen werden: <https://www.mww-forschung.de> [zuletzt 16.12.2020].

Die Beteiligung an den international seit Langem geführten Diskussionen in und um die Digital Humanities mit dem speziellen Fokus auf digitaler Sammlungsforschung birgt die Hoffnung, die Mühe möge sich auszahlen und den Status der drei Institutionen als zentrale Orte nicht nur der Bewahrung und Pflege, sondern auch der »Erschließung und Erforschung der deutschen literarischen und intellektuellen Traditionen vom Mittelalter bis zur Gegenwart« bestätigen.⁴

Ein derart groß angelegtes Projekt bringt jedoch auch erhebliche infrastrukturelle Herausforderungen mit sich und ist gut beraten, nicht in dieselben Fallen zu tappen, die bereits zuvor ähnlich aufgestellte Großprojekte der digitalen Forschungsinfrastruktur an ihr Ende brachten. Viel lässt sich diesbezüglich lernen etwa aus dem Scheitern des Projektes »Bamboo« der UC Berkeley und der University of Chicago, das von 2008 bis 2012 von der Andrew W. Mellon Foundation gefördert wurde.⁵ Als einen der Hauptgründe des Scheiterns nennt Dombrowski die fehlende Einigung auf eine konkretisierte Vision: »The project struggled to identify a coherent vision that neatly encapsulated all the work being done in the name of Bamboo, or to clearly describe what future state the work would collectively realize.«⁶

In MWW ist die größte Herausforderung, neben den unterschiedlichen historischen Schwerpunkten der drei beteiligten Institutionen und der Vielzahl der integrierten und integrierbaren geisteswissenschaftlichen Disziplinen, die diversen Traditionen des Sammelns in Archiven, Bibliotheken und Museen miteinander zu verknüpfen – und dies bezieht sich gerade auch auf verschiedene Arten der digitalen Erfassung und Verfügbarmachung von Bestandsdaten sowie ihre digitale Erforschung. »Der Begriff der Digitalität, der in der Wortbildung die Verbindung von digital, Materialität und Realität aufruft, bezeichnet für die Geisteswissenschaften die Verschränkung digitaler und analoger Praktiken«, betonen Gleixner und Steyer; und das ist vor allem in der digitalen Sammlungsforschung der Fall.⁷ Die Objekte in den unterschiedlichen Sammlungen der drei Institutionen beziehen sich nicht selten inhaltlich aufeinander; in den Datenbanken finden sich derlei Relationen bislang jedoch nur in Ausnahmefällen, wodurch die Sammlungen bestenfalls autark bleiben, schlimmstenfalls jedoch fehlinterpretiert werden. Es benötigt Systeme, die sowohl in den unterschiedlichen medialen Beständen der einzelnen Standorte

4 Wissenschaftsrat, Stellungnahme Klassik Stiftung, S. 17.

5 Vgl. Quinn Dombrowski: What Ever Happened to Project Bamboo?, in: *Literary & Linguistic Computing* (2014), doi: 10.1093/lc/fqu026.

6 Dombrowski, *Project Bamboo*, S. 6.

7 Ulrike Gleixner und Timo Steyer: Forschung in Bibliotheken, in: *Zeitschrift für Bibliothekswesen und Bibliographie (ZfBB)* 68,1 (2021), S. 27–37.

suchen und relevante Ergebnisse liefern, als auch institutionenübergreifend semantische Bezüge abbilden können. Der projektförmig geförderte Forschungsverbund MWW geht dieses Großvorhaben modellhaft an und demonstriert gleichsam *bottom-up* in mehreren Fallstudien, wie Sammlungsbestände offen digital erforschbar gemacht und in virtuellen Laboren (re-)präsentiert werden können. Um die Diskussion um digitale Sammlungsforschung sowohl institutionsübergreifend zu gestalten als auch an größere Debatten außerhalb der drei Standorte anzuschließen, erarbeiten drei Forschungsgruppen zudem ein theoretisches (Begriffs-)Fundament zur kritischen Reflexion der Themen Provenienz, Raum und Ökonomie des Sammelns. In den Gruppen wird der Sammlungsbegriff selbst im Spannungsfeld seiner digitalen Herausforderung diskutiert: Wie ist die Beziehung zwischen nicht digitalen Sammlungsobjekten und ihrer digitalen Erfassung in Form von Metadaten oder digitaler Reproduktion? Kann man noch von Sammlungen sprechen, wenn Objekte nicht mehr ortsgebunden sowie permanent verfügbar sind und Teil beliebig vieler Sammlungen gleichzeitig sein können? Entsteht ein neues Verständnis von Sammlungen, wenn sich multimodale Bestände sichtbar aufeinander beziehen können und damit die originäre Gegensätzlichkeit ihrer Objekte im Sinne einer Medienkonvergenz im Digitalen überwindbar erscheint?⁸ Wie kann der Sammlungsbegriff nutzbar gemacht werden, wenn es um Fragen der Bewertung expansiver Mengen von digitalen Forschungsdaten geht? Der Forschungsverbund MWW hat sich zum Ziel gemacht, bis 2024 zumindest einige dieser Fragen zu beantworten; der vorliegende Beitrag möchte das Problemfeld skizzieren und erste digitale und infrastrukturelle Lösungen präsentieren.

Die AG »Digitale Sammlungen« der Allianz der deutschen Wissenschaftsorganisationen beschäftigt sich in ihrem 2020 veröffentlichten Positionspapier mit der letztgenannten Frage. Das Konzept der »digitalen Sammlung« kann laut der AG »zur Priorisierung spezifischer Arten von Forschungsdaten [dienen], die es ›wert‹ sind, in eine Sammlung aufgenommen zu werden.«⁹ »Ob Forschungsdaten zu einer Sammlung zusammengeführt werden, ist das Ergebnis von [disziplinär oder thematisch zu organisierenden] Aushandlungsprozessen zwischen Forschungs-Communities und datenhaltenden Infrastruktureinrichtungen.«¹⁰ Die AG entwirft hier ein datafiziertes

8 Zum Begriff der Medienkonvergenz vgl. Henry Jenkins: *Convergence Culture. Where Old and New Media Collide*. New York 2006.

9 Arbeitsgruppe Digitale Sammlungen: Zur Bedeutung des Konzepts »Digitale Sammlung«, in: Schwerpunktinitiative »Digitale Information« der Allianz der deutschen Wissenschaftsorganisationen (2020), S. 1, doi: 10.2312/allianzoa.040.

10 Ebd., S. 2.

Verständnis von »Sammlung« vor der Folie des traditionellen nicht digitalen Sammlungsbegriffs und stellt damit große Schnittmengen mit dem in den Digital Humanities bereits fest etablierten Korpusbegriff her.¹¹ Es wird zu klären sein, welche Aspekte »digitale Sammlungen« von »Korpora« unterscheiden und wo der Begriff epistemischen Mehrwert bietet. Klar ist jedoch, dass bei der

Erstellung und Erschließung von digitalen Sammlungen bibliothekarische Expertise ebenso von Nöten ist wie Kenntnisse im Feld der Digital Humanities sowie eine fachwissenschaftliche Perspektive auf die Sammlung. Die Forschung muss daher bereits in einem frühen Stadium in der Sammlungserschließung stärker beteiligt werden – aus der *Datafication* resultiert eine notwendigerweise größere Nähe von Forschung und Erschließung.¹²

Der Sammlungsbegriff kann hier laut Gleixner und Steyer ordnungs- und orientierungsstiftend wirken, wenn Bestands- und Forschungsdaten im Digitalen eng aufeinander bezogen ein neues Datenset bilden.

Der vorliegende Beitrag stellt mit Fokus auf die DH-Teilprojekte die bisherigen Schritte vor, die im Forschungsverbund MWW gegangen wurden, um ortsunabhängig offen zugängliche und sich aufeinander beziehende Sammlungsdaten erforschbar zu machen. Zudem sollen Potenziale und Zukunftspläne Erwähnung finden, die für Forschung und auch Lehre in der nachhaltigen Bereitstellung einer offenen digitalen Forschungsinfrastruktur gegeben sind.

2. Das Digitale Labor

Das Digitale Labor des Forschungsverbunds MWW hat einen Schwerpunkt in der Anwendung und Vermittlung von Methoden und Tools der Digital Humanities. Bedarfsorientierung und die Möglichkeit zur Individualisierung je nach Fragestellung sind in diesem Rahmen zwei wesentliche Leitkriterien. Schaut man sich die Forschungslandschaft der Digital Humanities an, gibt es eine Vielzahl von Tools, die häufig projektspezifisch zur Lösung konkreter Probleme entwickelt wurden, jedoch oft nicht nachhaltig zur Verfügung gestellt werden. Möchte man hiervon ausgewählte Tools für die digitale (kol-

¹¹ Vgl. etwa Christof Schöch: Aufbau von Datensammlungen, in: Fotis Jannidis, Hubertus Kohle und Malte Rehbein (Hrsg.): Digital Humanities. Eine Einführung, Stuttgart 2017, S. 223–233.

¹² Gleixner und Steyer, Forschung in Bibliotheken.

laborative) Sammlungsforschung anbieten, muss man sich fragen, welche Tools (a) in welchen Forschungsphasen sinnvoll zum Einsatz kommen können, (b) über einen konkreten Projektkontext hinausgehend weiterentwickelt werden und (c) so niedrigschwellig vermittelt werden können, dass auch Forschende ohne informatische Vorkenntnisse in die Lage versetzt werden können, ihre Fragestellungen mithilfe dieser Tools zu verfolgen.

Die üblichen Phasen von digitalen Forschungsprojekten finden eine Formalisierung in der von DARIAH-EU entwickelten TaDiRAH-Taxonomie.¹³ Diese Taxonomie bietet im Forschungsverbund einen wesentlichen Referenzpunkt bei der Auswahl von zur Verfügung gestellten Tools. Der Anspruch ist, jede einzelne Forschungsphase methodisch mit digitalen Tools unterstützen zu können.

MWW stützt sich bei der Auswahl nicht nur auf einen sich abzeichnenden Kanon von Tools, die in den Digital Humanities zum Einsatz kommen. Durch die Integration in einen nachhaltig zur Verfügung stehenden virtuellen Forschungsraum fungiert der Verbund gleichzeitig kanonbildend – eine Funktion, die bereits bei der Auswahl von Tools kritisch reflektiert werden sollte, zugleich aber auch die Verankerung des Verbunds in sammlungsführenden Häusern begründet, deren eigentliche Kompetenz die auswählende Bewahrung auch digitaler Artefakte (zu denen Tools zählen sollten) und nachhaltige Bereitstellung von Infrastrukturen ist. Eine Konzentration auf häufig zur Anwendung kommende Grundfunktionalitäten wie der Entwicklung medienübergreifender Suchtechnologien, Digitalisierungsstandards etwa für die Volltextdigitalisierung von Drucken und Handschriften, die automatische Erkennung und Kennzeichnung benannter Entitäten als Voraussetzung für die Modellierung von Netzwerken, und nicht zuletzt die Möglichkeit, Daten in unterschiedlichen Visualisierungen zu erforschen und zu präsentieren, bewährt sich hierbei. Viel ist gewonnen, wenn verschiedenste Projekte in der Sammlungsforschung die gleichen Tools nutzen und mit ihrer Hilfe zu Einsichten gelangen können.

Schließlich sollte die Toolauswahl auch vor dem Hintergrund der Vermittlung stattfinden. Wer sind die Nutzenden einer virtuellen Forschungsumgebung? Die Hauptzielgruppe ist hierbei das Gros der geisteswissenschaftlichen Forschungsgemeinschaft, das häufig keine dezidierte Expertise in den Digital Humanities hat,¹⁴ durch sein spezifisches Fachwissen jedoch den wissenschaftlichen Diskurs auch in den DH enorm voranbringen kann. Die digi-

13 Vgl. <https://tadirah.info/index.html> [zuletzt: 24.8.2022].

14 Vgl. Gleixner und Steyer: Forschung in Bibliotheken: »Denn anders als bei analogen Forschungsprozessen kann nicht davon ausgegangen werden, dass das Curriculum

talen Methoden müssen daher in ihrem epistemischen Potenzial theoretisch reflektiert und die Tools niedrigschwellig vermittelt werden. Nur eine solche Form der *embedded Digital Humanities* kann verhindern, dass ein hoch spezialisiertes Angebot für eine kleine Gruppe von Expertennutzer*innen geschaffen wird, von der die Mehrheit der Forschenden ausgeschlossen bleibt. Fragen der niedrigschwelligen Vermittelbarkeit (etwa: Gibt es grafische Userinterfaces? Gibt es Tutorials, Beispielanwendungen, Handbücher, User-support etc.?) spielen daher eine nicht zu vernachlässigende Rolle in der Auswahl der bereitzustellenden Tools. MWW verfolgt somit eine Strategie, die nicht zuletzt das generelle Ziel einer »data literacy« hat, die mit Ridsdale et al. beschrieben werden kann als »the ability to collect, manage, evaluate, and apply data in a critical manner«.¹⁵ Wo dies nicht gegeben ist, ein Tool aber dennoch großes Erkenntnispotenzial für die digitale Sammlungsfor-schung verspricht, muss überlegt werden, in welcher Form das Tool zur Verfügung gestellt, d.h. für eine breitere Nutzer*innengruppe handhabbar gemacht werden kann etwa durch die Erstellung von Webanwendungen, die eine grafische Interaktion mit den jeweiligen Algorithmen ermöglichen, oder die Erarbeitung von Tutorials durch den Verbund selbst. Gleixner und Steyer betonen: »Das angestrebte Selbstbild und die wissenschaftliche Bedeutung von Forschungsbibliotheken [und das gilt gleichermaßen für Archive und Museen; Anm. JH] werden zukünftig maßgeblich von der Verbindung von Daten, Tools und Vermittlungsangeboten beeinflusst sein.«¹⁶

Ergänzend zu den genannten Kriterien spielen die Möglichkeit der digitalen Kollaboration (s. Abschnitt 3) und die offene Zugänglichkeit von Sammlungsdaten in Form offener Schnittstellen und Forschungsergebnissen eine wichtige Rolle in der Ausrichtung des Digitalen Labors. Da »die Kultur des Besitzens einer Kultur des Teilens weichen muss«,¹⁷ sind die wichtigsten Stichworte in diesem Zusammenhang Open Access in den Modellen des goldenen oder

der Geistes- und Kulturwissenschaften derzeit die notwendigen Fertigkeiten für digitalbasierte Forschung vermittelt.«

15 Chantel Ridsdale, James Rothwell, Mike Smit, Hossam Ali-Hassan, Michael Bliemel, Dean Irvine, Daniel Kelley, Stan Matwin und Brad Wuetherick: *Strategies and Best Practices for Data Literacy Education*. Knowledge Synthesis Report (2015), S. 2, doi: 10.13140/RG.2.1.1922.5044.

16 Gleixner und Steyer, *Forschung in Bibliotheken*.

17 Thomas Stäcker: *Die Sammlung ist tot, es lebe die Sammlung! Die Digitale Sammlung als Paradigma moderner Bibliotheksarbeit*, in: *Bibliothek – Forschung und Praxis* 43,2 (2019), S. 304–310, hier S. 304, doi: 10.1515/bfp-2019-2066.

grünen Wegs,¹⁸ die FAIR data principles¹⁹ sowie die Anbindung der digital erschlossenen Bestände an die sogenannte LOD-Cloud (»Linked Open Data«) als Standard des Semantic Web, das seinerseits die Vision eines Netzwerks verlinkter und semantisierten Daten darstellt.²⁰ Digitale Sammlungsforschung im Speziellen wie die digitale Transformation des kulturellen Erbes im Allgemeinen »sollte[n] dort, wo es datenschutz- und urheberrechtlich möglich ist, [...] dem Open-Science-Gedanken verpflichtet sein, sodass [digitale Sammlungen] ungehindert referenziert, aggregiert, selektiert, rearrangiert oder homogenisiert werden können«.²¹ Sammlungsführende Institutionen sollten sich aus der durch digitales Arbeiten entstehenden engeren Verknüpfung von Sammlungs- und Forschungsdaten den Auftrag ableiten, digitale Forschungsprozesse mit ihren Mitteln zu befördern. Tun sie das nicht, machen sie sich als Wissenszentren schlimmstenfalls obsolet, da ihre Bestände in den verschränkten und auf Nutzer*innenverhalten reagierenden Datenstrukturen des Internets mit der Zeit ganz einfach nicht mehr gefunden werden. In ähnlicher Richtung konstatiert Stäcker: »Es schöpft das Potential des Digitalen nicht aus oder muss sogar als Einschränkung von Forschungsmöglichkeiten betrachtet werden, wenn Bibliotheken in Anhänglichkeit an traditionelle Sammlungskonzepte Digitalisate mit Nutzungsbedingungen, die keine uneingeschränkte Nutzung erlauben, ausstatten, wenn sie nicht den massenhaften Download für Big-Data-Analysen gestatten, verminderte Bildqualitäten anbieten und so die Datenqualität verschlechtern oder auch Volltexte oder mit deskriptivem Markup versehen [sic!] Volltext verbergen oder Metadaten nicht über frei verfügbare Schnittstellen zur Verfügung stellen.«²²

3. Ortsunabhängige digitale Zusammenarbeit im Virtuellen Forschungsraum

Wird Ortsunabhängigkeit zum Leitmotiv eines neuen Selbstverständnisses von sammlungsbewahrenden Institutionen, sollten sämtliche damit einher-

18 Vgl. die Open-Access-Strategie des Bundesministeriums für Bildung und Forschung (2016): https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/1/24102_Open_Access_in_Deutschland.pdf?__blob=publicationFile&v=5 [zuletzt 24.8.2022].

19 FAIR steht für die Grundsätze nachhaltigen Datenmanagements: »findable«, »accessible«, »interoperable« und »re-usable«; vgl. https://www.forschungsdaten.org/index.php/FAIR_data_principles [zuletzt 25.11.2020].

20 Vgl. <https://www.w3.org/standards/semanticweb/> [zuletzt 25.11.2020].

21 Arbeitsgruppe Digitale Sammlungen, Digitale Sammlung, S. 3.

22 Stäcker, Die Sammlung ist tot, S. 309.

gehende Vorteile digitalen Forschens einbezogen werden. Mit den Methoden der Digital Humanities können digitalisierte Bestände so angereichert werden, dass ihre Provenienz, ihre internen wie externen Bezugnahmen über Institutionen hinweg, ihre Entstehung usw. dem Digitalisat eingeschrieben werden. Die Digitalisierung selbst – und damit ist mehr als lediglich die Bereitstellung von Scans im Bildformat gemeint – sollte bereits vor dem Hintergrund dieser Möglichkeiten im Rahmen der in den Fachcommunities ausgehandelten Standards geschehen. Standards gibt es spartenspezifisch sowohl in Bezug auf Meta- und Normdatenformate als auch in Bezug auf Dokumentstrukturen (wie etwa der TEI-XML-Standard in der Volltexterfassung). Die massenhafte Volltextdigitalisierung der in Sammlungsinstitutionen liegenden historischen Drucke ist dabei nicht nur ein seitens der Förderer explizit formuliertes Ziel,²³ sondern auch im Sinne der Forschung: Viele Methoden der Digital Humanities zielen darauf ab, semantische Strukturen von Texten automatisiert zu erfassen. »Wenn Bibliotheken früher Bücher mit Katalogisaten ausstatteten, um sie zugänglich zu machen, müssen sie heute Dokumentinhalte durch Volltexte und strukturelle Metadaten zugänglich machen.«²⁴

Die Daten treten mit der Digitalisierung ein in den *Data Life Cycle* und können dort immer weiter angereichert, (idealerweise interoperabel) verarbeitet, analysiert, visualisiert und interpretiert werden.²⁵ Dieser verstärkte Austausch bezieht sich jedoch nicht nur auf die Daten selbst, sondern auch auf die Forschenden, die mit den Daten umgehen: »Nicht allein die digitale Verknüpfung von Beständen unterschiedlicher Bibliotheken und Archive führt zu neuen Forschungsergebnissen, sondern auch die analoge Vernetzung ihrer MitarbeiterInnen ist Teil der Digitalität.«²⁶ Die hier »analog« genannte Vernetzung kann jedoch ebenfalls digital zustande kommen und unterstützt werden.²⁷ Digitale Kollaboration – das gemeinsame Arbeiten an bestimmten Datensätzen mit digitalen Tools – kann durch die neuen infrastrukturellen Möglichkeiten auch in Bezug auf wissenschaftliche Forschungsfragen Anwendung finden. Die Grundvoraussetzung ist dabei, dass ein Tool von sich aus eine Möglichkeit der gemeinsamen Projektarbeit anbietet, oder dass das Tool

23 Vgl. etwa das seitens der Deutschen Forschungsgemeinschaft geförderte Großunternehmen OCR-D: <https://ocr-d.de/de/> [zuletzt 14.12.2020].

24 Stäcker, *Die Sammlung ist tot*, S. 307.

25 Zum *Data Life Cycle* vgl. etwa Jeannette M. Wing: *The Data Life Cycle*, in: *Harvard Data Science Review* 1,1 (2019), doi: 10.1162/99608f92.e26845b4.

26 Gleixner und Steyer, *Forschung in Bibliotheken*.

27 Analog sollte nicht als Gegenteil von digital verstanden werden. Vielmehr stehen die beiden Modi in dynamischen und historisch wandelbaren Austauschprozessen zueinander (vgl. etwa Schröter 2004).

eingebettet ist in eine virtuelle Forschungsumgebung, welche die einzelnen Zugriffe der unterschiedlichen Nutzer*innen orchestriert, indem Arbeitsschritte synchronisiert und potenziell verschiedene Zugriffsrechte geregelt werden.

Im Forschungsverbund MWW wurde vor diesem Hintergrund der sogenannte Virtuelle Forschungsraum (VFR)²⁸ entwickelt, um sowohl die institutionsübergreifende Suche digitalisierter und digitaler Sammlungen, die kollaborative Arbeit mit ausgewählten Daten unter Hinzuziehung forschungsphasenspezifischer Tools, als auch die Dokumentation und Präsentation von Forschungsdaten und -ergebnissen zu ermöglichen. Von Beginn an wurden dabei unterschiedliche Nutzer*innengruppen und Nutzungsszenarien berücksichtigt, um die Arbeit mit den im VFR integrierten Tools möglichst für alle zu ermöglichen.²⁹ Die (in den kommenden Jahren noch zu erweiternde) Auswahl konkreter Tools geschah sowohl vor dem Hintergrund von Nutzer*inneninterviews als auch, wie erwähnt, mit Bezug auf die von DARIAH-EU erarbeitete TaDiRAH-Forschungsphasen-Taxonomie.³⁰ Beachtet werden muss hierbei, dass sowohl Disziplinen untereinander verschiedene Vorgehensweisen und Schwerpunktsetzungen haben, als auch disziplinenintern eine große Diversität von Forschungsszenarien anzutreffen sind: Die traditionellere historisch-hermeneutische Arbeit am Einzelobjekt kann in Form digitaler Annotation stattfinden; gleichzeitig bieten maschinelle Analysemethoden von Text- oder Objektkonglomeraten die ergänzende Möglichkeit, Relationen etwa in Form von Netzwerken abzubilden, d.h. Sammlungsobjekte in ihrem internen wie externen Verweiskontext erfahr- und erforschbar zu machen. Eine Rückkoppelung mit den diversen Bedarfen der unterschiedlichen potenziellen Nutzungsgruppen verhindert dabei, in die gleiche Falle zu gehen, die auch schon für das Scheitern des Projekts »Bamboo« mitverantwortlich war: »Effective communication with scholarly and professional communities was never one of Project Bamboo's greatest strengths.«³¹ Ein auf Grundlage von Bedarfen ausgewähltes Set an Tools nachhaltig verfügbar zu halten, sollte das Bestreben einer jeden sammlungshaltenden Institution sein, so sie ihre Anschlussfähigkeit an Forschungs-

28 Vgl. <https://vfr.mww-forschung.de/> [zuletzt 14.12.2020].

29 Vgl. Swantje Dogunke und Timo Steyer: Virtuell Zusammenwachsen: Konzeption, Aufbau und Intention der digitalen Forschungsinfrastruktur im Forschungsverbund MWW, in: Martin Huber, Sybille Krämer und Claus Pias (Hrsg.): Forschungsinfrastrukturen in den digitalen Geisteswissenschaften: Wie verändern digitale Infrastrukturen die Praxis der Geisteswissenschaften?, Frankfurt a.M. 2019, S. 111-128, hier S. 114f., urn:nbn:de:hebis:30:3-519476.

30 Vgl. ebd.

31 Dombrowski, Project Bamboo, S. 7.

diskurse bewahren möchte. Für die Entwickler*innen der Tools bietet die Einbindung neben erhöhter Sichtbarkeit und Anerkennung auch die Chance der Nachhaltigkeit. Für die Forschungscommunity bedeuten nachhaltige Lösungen nicht zuletzt die Reproduzierbarkeit und damit die intersubjektive Nachvollziehbarkeit digital gewonnener Forschungsergebnisse. Dogunke und Steyer betonen: »[D]ie Forschung benötigt kuratierte Tools, um auch ein Vertrauen für ihre nachhaltige Nutzung zu sichern.«³²

Neben der modularen Struktur individualisierter Labore und der Möglichkeit strukturierter Zusammenarbeit durch ein integriertes Zugriffs- und Rechtemanagement bietet die Einbettung von Tools in virtuelle Forschungs-umgebungen im europäischen Raum zudem die Möglichkeit, Instanzen von Tools anzubieten, die sonst außerhalb des europäischen Datenschutzgesetzes arbeiten. Der VFR wird von der Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) gehostet, wodurch die Bedingungen der DSGVO gewährleistet sind.³³ Instanzen von außereuropäischen Tools im VFR bieten somit einerseits die Sicherheit des datenschutzkonformen Arbeitens. Andererseits sind eigene Toolinstanzen nicht mehr auf dem neuesten Stand, sobald das Original-Tool ein Update erfahren hat und dieses im VFR noch nicht nachgerüstet sein sollte. Es ist eine Frage der Vermittlungsarbeit, den Nutzenden diese Vor- und Nachteile vor Augen zu führen, damit in jedem Bedarfsfall eine je individuelle Entscheidung getroffen werden kann.

4. Formen virtueller Labore und Dynamik im Virtuellen Forschungsraum

Im VFR ist es nun möglich, in der sammlungsübergreifenden Suche digitalisierte Sammlungen institutionenübergreifend zu suchen und sich somit eigene forschungsspezifische Sammlungen anzulegen.³⁴ Die eigentliche kol-

32 Dogunke und Steyer, *Virtuell Zusammenwachsen*, S. 121.

33 Zur Europäischen Datenschutz-Grundverordnung (DSGVO) vgl: <https://dsgvo-gesetz.de/> [zuletzt 14.12.2020].

34 Die Suche (<https://vfr.mww-forschung.de/suche> [zuletzt 14.12.2020]) stellt eine Instanz der von DARIAH-DE entwickelten Generic Search dar (<https://search.de.dariah.eu/search/> [zuletzt 14.12.2020] und enthält derzeit 31 Kollektionen. Die Integration weiterer digitalisierter Sammlungen aus Marbach, Weimar und Wolfenbüttel ist geplant. Digitalisiert bedeutet hier jedoch nicht immer das Gleiche und kann sich auf die unterschiedlichen Stufen von digitaler Erfassung, Digitalisierung in Form von Image Scans, oder – in Bezug auf Texte – die Volltextdigitalisierung durch eine Optical Character Recognition (OCR) oder digitale Edition beziehen.

laborative Arbeit findet dann in sogenannten Laboren statt. Diese können unterschiedlichste Ausformungen haben, die auf die jeweiligen Bedarfe der beteiligten Forschenden zurückgeht. Generell können Projekt-, Ausstellungs- und Veranstaltungslabore unterschieden werden. Während in Veranstaltungslaboren vornehmlich Tools zur ortsunabhängigen Kommunikation, Planung und zur Präsentation bzw. Dokumentation der Veranstaltung gefragt sind, gehen Projektlabore in der Regel über diese Möglichkeiten hinaus. Je nach Forschungsphase können hier zusätzlich Tools zur Erfassung und Digitalisierung nicht digitaler Bestände, zur Analyse von Text-, Bild-, Objekt- oder Raumdaten, zur Visualisierung etc. modular hinzugefügt werden. Sollen in einem Projekt unterschiedliche Zugriffs- und Nutzungsrechte vergeben werden, ist dies in der Verwaltung des jeweiligen Labors ebenfalls möglich.³⁵ Die Ausstellungslabore schließlich dienen zur Dokumentation und zeitunabhängigen Rekapitulation von während der ersten Förderphase von MWW an den drei Standorten stattgefundenen Ausstellungen: »Luthermania«, »Du bist Faust« und »Die Familie. Ein Archiv«.³⁶

Der VFR ist damit sowohl Präsentationsplattform für die Außendarstellung und Vermittlung der in den Fallstudien und Forschungsgruppen gewonnenen Forschungsergebnisse als auch ein Content Management System, wie es den meisten Webseiten zugrunde liegt. In jedem Labor können öffentliche und private Seiten unterschieden werden, wobei es ebenso möglich ist, sämtliche Seiten öffentlich oder auch das gesamte Labor privat einzustellen. Da Forschung in der Regel nicht während ihres eigentlichen Prozesses veröffentlicht wird, sondern zunächst vermeintliche Ergebnisse konsolidiert und präsentationsfähig gemacht werden, unterstützt diese doppelte Grundfunktionalität die Forschenden in mehreren Prozessstufen: Digital gewonnene Ergebnisse etwa in Form von angereicherten Metadaten, volltexterschlossenen Drucken, Netzwerken oder anderen Visualisierungen müssen nicht migriert werden, um sie zu veröffentlichen, sondern können dort, wo sie erstellt wurden, bleiben. Arbeits- und Präsentationsraum gehen im Virtuellen Forschungsraum von MWW ineinander über und katalysieren somit die diversen Arbeitsschritte digitaler Sammlungsforschung.

Die Projektlabore demonstrieren beispielhaft Anwendungsfelder der digitalen Sammlungsforschung in ihrer ganzen Breite: Sowohl die thematisch

35 Beispielhafte Anwendungsfälle hierfür wären etwa die Einbindung von studentischen Hilfskräften in die Projektarbeit oder die Nutzung eines VFR-Labors im Rahmen von Lehrveranstaltungen, bei denen die Zugriffsrechte der Studierenden ggf. eingeschränkt werden sollen.

36 Vgl. <https://vfr.mww-forschung.de/ausstellungen> [zuletzt 14.12.2020].

aufgefächerten Forschungsgruppen als auch die sechs Fallstudien von MWW, deren historische Fokusse sich von der Frühen Neuzeit bis in die Gegenwart erstrecken, als auch bereits abgeschlossene Forschungsprojekte besitzen ein virtuelles Labor; mehrere davon bereits im öffentlichen Präsentationsmodus.³⁷ Die Anschlussfähigkeit für weitere Forschungsprojekte ist evident und auch wenn MWW selbst den Fokus dezidiert auf digitale Sammlungsforschung setzt, funktioniert die kollaborative Arbeit im Labor auch für geisteswissenschaftliche Fragestellungen abseits der Sammlungsforschung. Die im Frühjahr 2020 ausgebrochene Covid-19-Pandemie etwa hat ganze Fachgemeinschaften vor die Aufgabe gestellt, die Lehre des Sommersemesters ad hoc vollständig digital abzuhalten. Viele Universitäten haben in dieser Situation eigene eLearning-Lösungen angeboten – eine ortsübergreifende disziplinspezifische Möglichkeit des Erfahrungsaustauschs und des effektiven Erlernens digitaler Lehrformate gab es jedoch nicht. Auch für diese – der Sammlungsforschung ferne – Anforderung hat sich die Struktur des VFR als hilfreich erwiesen, wodurch die Anschlussfähigkeit des Systems eindrücklich unter Beweis gestellt wurde. In einem gemeinsamen Kraftakt haben über 20 Vertreter*innen der nationalen und internationalen Germanistik zusammen mit dem Forschungsverbund das Portal »Digitale Lehre Germanistik« binnen kürzester Zeit ins Leben gerufen – in Form eines Labors im VFR.³⁸ Dem erfolgreichen Launch folgte im Anschluss an das Sommersemester eine ebenso erfolgreiche virtuelle internationale Konferenz zum gleichen Thema, die ebenfalls über den VFR organisiert und koordiniert wurde.³⁹ Im Vergleich zu nicht digitalen Veranstaltungen dieser Art war der Planungsaufwand dabei relativ gering, da die digitale Infrastruktur vorhanden und einsatzbereit war.

37 Die veröffentlichten Projektlabore finden sich unter <https://vfr.mww-forschung.de/projekte> [zuletzt 16.12.2020]. Die sechs in der derzeitigen Förderphase zu bearbeitenden Fallstudien heißen (in chronologischer Reihenfolge): 1. Intellektuelle Netzwerke. Frühneuzeitliche Gelehrtenbibliotheken als Wissens- und Kommunikationsräume, 2. Weltwissen. Das kosmopolitische Sammlungsinteresse des frühneuzeitlichen Adels, 3. Goethe digital, 4. Kunst und Memoria. Der dingliche Nachlass des Weimarer Nietzsche-Archivs, 5. Transatlantischer Bücherverkehr. Migrationswege und Transfer Routen vor und nach 1945 und 6. Archivierung, Erschließung und Erforschung von Born-digitals; vgl. <https://www.mww-forschung.de/fallstudien> [zuletzt 16.12.2020].

38 Das Portal findet sich unter <https://vfr.mww-forschung.de/web/digitale-lehre-germanistik/> [zuletzt 16.12.2020].

39 Vgl. den Tagungsband: Kai Bremer, Thomas Ernst, Andrea Geier, Jan Horstmann, Ariane Larrat, Thorsten Ries und Claudius Sittig (Hrsg.): Während und nach Corona: Digitale Lehre in der Germanistik: Ergebnisse der digitalen Konferenz am 25./26. August 2020. Frankfurt am Main: GinDok (2020). URN: urn:nbn:de:hebis:30:3-592851.

Die Vielzahl diverser Beispielprojekte, die im VFR realisiert wurden und werden, vermag den Mehrwert einer nachhaltigen digitalen Arbeits-, Austausch- und Präsentationsplattform zu verdeutlichen und vermeidet zugleich einen weiteren entscheidenden Fehler, der etwa im Projekt »Bamboo« begangen wurde: »Demonstrator projects could have effectively and concretely shown that Bamboo was producing something useful, or provided an opportunity for feedback at a stage where it could have been incorporated productively.«⁴⁰

5. Fazit und Ausblick – suchen, erforschen, speichern, publizieren

Neben den Laboren für Projekte, Veranstaltungen und Ausstellungen werden im Forschungsverbund weitere zentrale Anwendungen erarbeitet, die verbunden mit der Seite des VFR bzw. dort eingebettet das digitale Angebot funktional ergänzen. Die sammlungsübergreifende Suche ermöglicht das Auffinden digitalisierter Sammlungen an allen drei Verbundinstitutionen durch die Integration in ein gemeinsames, spartenübergreifendes Metadatenformat.⁴¹ Entwickelt wird zudem ein Virtueller Sammlungsraum (VSR), der das Herstellen von semantischen Relationen zwischen medial potenziell heterogenen Sammlungsbeständen aus unterschiedlichen Datenbanken in visueller Weise ermöglichen wird. Schließlich erarbeitet der Verbund standortübergreifende Lösungen für die Langzeitspeicherung sowie das Forschungsdatenmanagement, abgerundet durch eine ebenfalls in den VFR integrierte digitale »Platform for open access Publication« (PoP). Neben der verbundeigenen »Zeitschrift für digitale Geisteswissenschaften« (ZfdG) – für verschriftlichte Artikel und Projektvorstellungen – und dem VFR – für Labore mit Visualisierungen, Blogposts, Auswertungen von Daten etc. – wird PoP die offen zugänglichen Publikationsmöglichkeiten für die digitale Sammlungsforschung durch Möglichkeiten der Publikation von Forschungsdaten oder Zweitpublikation von wissenschaftlichen Beiträgen erweitern und komplettieren. Sammlungs- und Forschungsdaten können hierdurch eng aufeinander bezogen im Sinne einer *enhanced publication* der Öffentlichkeit zur Verfügung gestellt werden.

In einer Zeit der sich überschlagenden technologischen Neuerungen, der sich potenzierenden digitalen Daten und Tools und damit verbundenen wissenschaftlichen Erkenntnissen, sind das Sammeln und Bewahren des kulturellen Erbes und mit ihnen die sammlungstragenden Institutionen vor

⁴⁰ Dombrowski, Project Bamboo, S. 9.

⁴¹ Zum gemeinsamen Metadatenformat vgl. <https://vfr.mww-forschung.de/web/digitales-labor/metadaten> [zuletzt 16.12.2020].

erhebliche Herausforderungen aber auch wichtige Aufgaben gestellt. Was ist bewahrungswürdig und welche Versionen eines sich stets weiterentwickelnden, d.h. verändernden digitalen Werkzeugs, Algorithmus oder Kunstwerks sollen langfristig archiviert werden? Die auf diesen Selektionsprozessen auf-sattelnde digitale Erschließung und Aufbereitung bildet einen zweiten Filter, durch den Forschung auf kulturelles Erbe blickt. Aus diesen großen Herausforderungen lässt sich jedoch gleichzeitig die wichtige Rolle ablesen, die den sammlungstragenden Institutionen als Träger eines wie auch immer gearteten kulturellen Gedächtnisses und damit als Standpfeiler unserer kulturellen Identität, die aber ebenso der reflektierten digitalen Sammlungsforschung zukommt. Einem virtuellen Forschungsraum kommt in ähnlicher Hinsicht die Aufgabe zu, der Forschungscommunity nachhaltig ein möglichst interoperables Set an Tools zur Verfügung zu stellen, das auf einer bedarfsorientierten Auswahl beruht und sich an Parametern wie Benutzer*innenaufkommen, wissenschaftlicher wie auch niedrigschwelliger Nutzbarkeit, Datensicherheit, Nutzer*innenfreundlichkeit und Supportmöglichkeiten orientiert.

In Zukunft werden diese Aufgaben – nicht nur in Bezug auf Fragen der Archivierung von Born-digitals – nur noch umso wichtiger. Dem exponentiellen Datenwachstum werden von Entwicklerseite neue Methoden entgegengestellt, um Forschung weiter zu ermöglichen, aber auch um Prozesse zu automatisieren. Viel Aufmerksamkeit erhält derzeit etwa die Analyse großer Datenbestände mittels neuronaler Netze in Deep-Learning-Algorithmen, die sowohl auf Text- wie auf Bildbestände angewendet werden können. Unterschieden werden hierbei »recurrent neural networks« (für Texte) und »convolutional neural networks«.⁴² Die Kombination für die Analyse multimedialer Bestände und z.B. Text-Bild-Relationen wird eine Herausforderung sein, die insbesondere vonseiten der digitalen Sammlungsforschung formuliert werden dürfte. Die Grundvoraussetzung ist und bleibt ein Bekenntnis zu offenen Sammlungsdaten und Forschungs- wie Vermittlungsinfrastrukturen sowie eine beherrzte Abkehr von überkommenen proprietären Strukturen des Sammelns.

»Wenn zukünftig alle Texte frei zirkulieren, wenn sich alle Bibliotheken in wahrhaft alexandrinischem Umfang gegenseitig kopieren, dann liegt die Qualität der Sammlung nicht mehr im *Content*, sondern in der Fähigkeit, den Content zu prozessieren, d.h. zu ergänzen, aufzubereiten und sinnvoll zu selektieren, vulgo zu ordnen.«⁴³

42 Vgl. Hannes Bajohr: Algorithmic Empathy. On Two Paradigms of Digital Generative Literature and the Need for a Critique of AI Works, in: Basal Media Culture and Cultural Techniques Working Papers 4 (2020), doi: 10.5451/UNIBAS-EP79106.

43 Stäcker, Die Sammlung ist tot, S. 308.