

Infrastrukturierung der Filmforschung

Auf dem Weg zu digitalen Forschungsumgebungen
im Netz?

Malte Hagener / Dietmar Kammerer*

In dem Maße, in dem die digital vernetzte Kommunikation und die Migration von Daten, Diensten und anderen Dingen in modulare, verschaltete und in Echtzeit operierende Online-Netzwerke fortschreiten und sich weiter entwickeln, lässt sich auch eine Historisierung beobachten, die damit den Blick auf unsere Gegenwart verändert. Historisch zu denken bedeutet dabei nicht, der Wikipedia im Anhäufen von Fakten Konkurrenz zu machen, sondern die gestaltende und argumentative Perspektivierung von vergangenen Entwicklungen auf unsere Gegenwart und Zukunft hin. Das Internet als universale Plattform, die (angeblich) alles modulieren kann, und der Computer als (vermeintliches) Universal-Medium sind längst kein Neuland mehr, sondern sind zu alltäglichen Begleitern und selbstverständlichen Werkzeugen geworden, die immer schon einen Teil von wissenschaftlichen Praktiken ausmachen. Wenn also die Verschränkung von digitalen Wissensobjekten und Praktiken der Forschung und Lehre längst untrennbar ist, so kommt es darauf an zu reflektieren, wie das Material unserer wissenschaftlichen Arbeit – die Quellen – uns zur Verfügung steht. Und nicht nur das:

* [Anm.d.Red.:] Die Autoren betreiben seit 2017 das Open Access-Repository der Medienwissenschaft media/rep/ (mediarep.org). Dietmar Kammerer ist zudem Betreiber des Weblogs filmwissenschaft.umsonst.

Es gilt auch zu überlegen, was sich überhaupt als Quelle, als Material, als Datum qualifiziert und wie sich diese Einschätzungen unter den derzeitigen technischen und medialen Bedingungen verändern (Verhoeven 2012).

Mit dieser Argumentation laufen wir natürlich offene Türen ein, denn wenn die Medienwissenschaft so etwas wie einen Refrain hat – einige würden es auch als Platte mit Sprung bezeichnen –, dann ist es der *ceterum censeo*-hafte Hinweis auf die Medialität von medial vermittelten Inhalten und Erfahrungen. Ob nun McLuhanesque («the medium is the message»), Kittlerianisch («das mediale Apriori») oder gar vorwurfsvoll gegen andere («Medienvergessenheit») gewendet, der Hinweis darauf, dass die mediale Form und Beschaffenheit mitzubedenken sei, ja eigentlich zentral steht, wenn wir es mit medial bedingter Formgebung und Sinnerzeugung zu tun bekommen, gehört inzwischen zu den Selbstverständlichkeiten nicht nur in der Medienwissenschaft, sondern längst auch darüber hinaus. In dieser Perspektive lohnt sich ein medienwissenschaftlicher Blick auf spezifisch filmwissenschaftliche Gegenstände, nämlich digitale Archive, Repositorien, Datenbanken und Infrastrukturen mit filmbezogenen Informationen, die zunehmend unser alltägliches Geschäft in der Filmforschung (mit) bestimmen. Dabei soll dieses Phänomen weniger aus der Perspektive der Neuheit beleuchtet werden, vielmehr geht es darum, aus einer historischen Betrachtung die Veränderung der Forschungspraxis zu umreißen, die diese Entwicklungen zum Teil antreibt und zum Teil auch davon angetrieben wird.

Dieser Beitrag möchte zunächst eine historische Situierung und Kontextualisierung vornehmen, ehe ein systematischer Blick auf die gegenwärtige Praxis geworfen wird. Dabei soll recht nahe an konkreten Szenarien und vorhandenen Infrastrukturen argumentiert werden. Denn wir sind der Meinung, dass die forschungsrelevanten Infrastrukturen von der Wissenschaft (mit) entworfen und betrieben werden müssen, damit dieser Prozess wissenschaftsgeleitet erfolgt.

Filmdaten im Internet – ein Phasenmodell

Datenbanken und Software, Portale und Algorithmen sind Werkzeuge, die nicht an und für sich gut oder schlecht sind, sondern es ist von entscheidender Bedeutung, wie die digitalen Forschungstools konkret ausgestaltet sind. Dass an diesen Tools – und damit an der Frage ihrer Beschaffenheit – kein Weg vorbeiführt, wird wohl kaum noch jemand bestreiten wollen. Wir

recherchieren in Datenbanken und Beständen, die sich an anderen Orten befinden als dort, wo wir uns gerade aufhalten. Wir greifen über digitale Werkzeuge auf online verfügbare Bestände zu, wir wenden Methoden und Tools auf große Datenmengen an, was mit analogen Verfahren zwar ebenfalls möglich, aber kaum praktikabel wäre. In diesem Sinne ist die Film- und Medienwissenschaft längst eine digitale Geisteswissenschaft, ohne dass sie sich dafür mit dem Etikett ‚Digital Humanities‘ schmücken (oder verunstalten) lassen muss. Was sich allerdings immer wieder verändert, ist die Frequenz, mit der wir unterschiedliche digitale Tools nutzen, die Menge der aggregierten Datensätze, die uns zur Verfügung steht, der Grad der Akzeptanz für derartige Methoden in der Community der Forschenden sowie das Ausmaß der Vernetzung der Angebote untereinander. Wir sind der Überzeugung, dass man diese Entwicklung am sinnvollsten historisch betrachtet, weil dadurch jeweilige Veränderungen und Transformationen am produktivsten in den Blick geraten. Dazu soll heuristisch eine Periodisierung von vier Phasen vorgenommen werden.

Es gab sie tatsächlich – die Phase, in der das Internet neu war und großen Freiraum bot, die Zeit, in der noch nicht klar zu erkennen war, welchen Weg das junge Medium einschlagen würde. Diese Pionierphase, wie man sie nennen könnte, ähnelt in gewisser Weise dem frühen Kino vor 1907/08, als weder formal und ästhetisch noch institutionell und ökonomisch klar war, in welche Richtung sich das Medium entwickeln würde.¹ In dieser Phase fand man viele Angebote über Mund-zu-Mund-Propaganda, weil es keine zentrale Anlaufstelle gab. Suchmaschinen waren tendenziell Linklisten (Yahoo), und die meisten Angebote stammten eher von Privatpersonen oder kleinen Gruppen besessener Amateure als von institutionellen Anbietern. Diese Situation war bis in die zweite Hälfte der 1990er-Jahre prägend.

Eine zweite Phase sorgte dann für eine erste Konsolidierung, in der sich bestimmte Angebote nicht nur durchsetzten, sondern meist auch eine kommerzielle Gestalt annahmen. Das Paradebeispiel ist hier sicher die *Internet Movie Database* (imdb.com), die von einer Gruppe Cinephiler um Colin Needham als Newsgroup begonnen wurde, sich rasch zu einer wich-

1 Als Beispiele einer archäologischen Situierung von Mediengeschichte können Marvin (1990) sowie grundsätzlich die revisionistische Forschung zum frühen Kino genannt werden, die sich um eine nicht-teleologische Sichtweise des Films vor 1910 bemüht, wie auch die Medienarchäologie, die sich der Rekonstruktion von Schichten der Medientechnologie und -nutzung widmet.

tigen Informationsquelle für Interessierte, auch für die Filmwissenschaft, entwickelte und 1998 in den Besitz von Amazon übergang.

Eine dritte Phase, die nach dem Platzen der sogenannten Dot-Com-Blase mit gehöriger Vorsicht einsetzte, sorgte dann für die Institutionalisierung und Konzentration der Angebote wie auch für die systematische Zusammenführung von zuvor separat bestehenden Beständen. Hier traten dann noch einmal verstärkt institutionelle Player auf den Plan, etwa im Fall der Plattform *filmportal.de*, die sich als «zentrale Internet-Plattform für umfassende, zuverlässige und kostenlose Informationen zu allen deutschen Kinofilmen»² versteht. *Filmportal* wird von einer Reihe wichtiger Stakeholder aus dem Filmbereich (darunter die Beauftragte der Bundesregierung für Kultur und Medien, die Spio und die Friedrich-Wilhelm-Murnau-Stiftung) finanziell getragen und ist am Deutschen Filminstitut (DIF) angesiedelt. *Filmportal*, das Datenbestände mehrerer Mitglieder des Kinematheksverbunds zusammenführt, ging auf eine CD-Rom-Veröffentlichung filmografischer Daten (*DeFi – Deutsche Filmographie*) zurück und ist seit 2005 online.

An diese Phase der Institutionalisierung schließt sich die derzeitige, vierte Phase der Vernetzung an, in der Linked-Open-Data und die Nutzung von bestehenden und neuen Schnittstellen dafür sorgen (könnten), dass die Menge der Daten, auf die zugegriffen werden kann, exponentiell wächst. Projekte wie Wikidata oder auch aktuelle Diskussionen um den Einsatz von Normdaten und Standards sind Ausdruck dieser Entwicklung.

Man darf sich ein derartiges Phasenmodell – Pionierzeit, Konsolidierung, Institutionalisierung, Vernetzung – jedoch nicht als eine starre Abfolge von Zuständen vorstellen, die sich nach Art eines Staffellorens ablösen. Reste und Überbleibsel voriger Phasen finden sich bis heute. So ist etwa die von Herbert Birett über Jahrzehnte zusammengestellte und seit den 1990er-Jahren auch online existierende Datensammlung noch immer im Netz verfügbar. Die – posthum von der Universitätsbibliothek Köln übernommene – Anhäufung von Fakten widersetzt sich heute üblichen Standards von Aufbereitung, Nutzungsfreundlichkeit und Gestaltung.³ Und Catherine Grants Kanal *Audiovisualcy: Videographic Film and Moving Image Studies and Video Essays*,⁴ der maßgeblich zur Verbreitung und Durchsetzung des Video-Essays als wissenschaftliche Form beigetragen hat, entstand

2 So lautet die Selbstbeschreibung auf der Website: <https://www.filmportal.de/seite/die-ziele-von-filmportalde> (letzter Zugriff am 9.11.2019).

3 <https://www.kinematographie.de/> (letzter Zugriff am 23. Oktober 2019).

4 <https://vimeo.com/groups/audiovisualcy> (letzter Zugriff am 9.11.2019).

als zunächst private Initiative einer einzelnen Person, die auf der Infrastruktur des Internets und eines (kommerziellen) Videoportals aufsetzt.

Mit dem Verweis auf bestehende Plattformen, Infrastrukturen und Dienste – seien diese kommerziell-proprietär, selbstprogrammiert oder im Rahmen von Open-Source-Communities entwickelt – ist eine Struktur angedeutet, welche die Komplexität der gegenwärtigen Situation adäquat beschreibt. Jedes Angebot findet sich in einer komplexen Umwelt, die sich nicht nur in Konkurrenzverhältnissen manifestiert (das ist in Zeiten hochgradiger Spezialisierung eher die Ausnahme), sondern zumeist in komplexen, selten auf einzelne Felder begrenzbaren Wechselwirkungen. So findet sich ein Betreiber von Word-Press-basierten Websites an deutschen Universitäten häufig in der Situation, dass diese aufgrund von Sicherheitsbedenken oder Erwägungen zur Corporate Identity der Hochschule nicht direkt vom Hochschul-Rechenzentrum gehostet werden können. Ein Umzug zu privatwirtschaftlichen Hosting-Unternehmen erzeugt dagegen wiederum andere Probleme – etwa der Finanzierung, der Haftung oder auch des Zusammenspiels mit anderen Komponenten. Meist sind die Situationen so beschaffen, dass sich eine komplexe Interdependenz von unterschiedlichen Faktoren ergibt. Sie sind nicht selten durch Entscheidungen (etwa für eine bestimmte grundlegende Software) bestimmt, die vor längerer Zeit in Unkenntnis der Spätfolgen und Kollateralschäden getroffen wurden. Ein distanziert medienökologischer Blick auf die Kausalitäten, Anpassungen, Resilienzen und Transformationen in der digitalen *longue durée* lohnt sich, wenn man jenseits der tagesaktuellen Innovationsversprechen verstehen will, wie sich wissenschaftlich relevante Infrastrukturen formieren, wie sie gestaltet werden und sich verändern.

Bereitstellung von und Zugang zu Daten: Modelle

Es wäre sicher naiv zu behaupten, dass die Zugänglichkeit von Materialien nicht schon immer – also auch in prädigitalen Zeiten – bestimmt hätte, was auf welche Weise erinnert, erforscht und kanonisiert wird. So sorgte beispielsweise Hans Richter durch Schenkungen dafür, dass maßgebliche Institutionen wie das *Museum of Modern Art* und das *Guggenheim Museum* die Filme Viking Eggelings und auch seine eigenen im Bestand – und damit zur Verfügung – hatten (Hagener/Zimmermann 2009). Dadurch stellte er sicher, dass bei der Gestaltung entsprechender Programme regelmäßig auf diese Filme zurückgegriffen wurde, schon weil sie für die Kuratoren einfach

greifbar waren. Und der Zugang zu bestimmten Quellen und Zeitschriften hatte stets großen Einfluss darauf, was erforscht wurde. Insofern bedeutet die digitale Infrastruktur vor allem eine Intensivierung und Akzeleration bestehender Prozess-Strukturen, keineswegs aber eine radikale Veränderung. Anhand von zwei Beispielen sollen einige Aspekte dieser Transformation aufgezeigt werden.

Ein Vorzeigeprojekt und sicher von Umfang wie Anspruch her ein Vorreiter ist die *Media History Digital Library*,⁵ eine nicht-kommerzielle Initiative zur Digitalisierung und Bereitstellung von historischen Publikationen über Film, Rundfunk und Tonaufzeichnung. Sie ist an der University of Wisconsin, Madison beheimatet, wurde von David Pierce aus der Taufe gehoben und wird seit einigen Jahren mit großem Einsatz von Eric Hoyt betrieben und stetig ausgebaut. Viele Zeitschriften und Bücher aus der Zeit vor 1960 (und auch einige aus der Zeit danach) sind nach US-Recht nicht mehr durch das Copyright geschützt, sodass sie problemlos der Allgemeinheit zugänglich gemacht werden können. Hier stehen derzeit knapp zwei Millionen Seiten gescannt zur Verfügung, die sich komfortabel mit dem *Lantern-Tool* durchsuchen lassen.⁶ Das OCR-Werkzeug, das dank der Texterkennung die Volltextrecherche gestattet, ist noch nicht für alle historischen Schriftsätze optimiert, aber die Breite und Tiefe des Zugriffs vor allem auf das US-amerikanische Material ist beeindruckend. Inzwischen versammelt die *Media History Digital Library* unter dem Stichwort «Global Cinema Collection» auch Quellen aus anderen Ländern. An deutschsprachigen Quellen finden sich derzeit vier Jahrgänge der Zeitschrift *Der Kine-matograph*, Werbematerial der Ufa, einzelne Nummern weniger bekannter Zeitschriften und einige Nachschlagewerke. Die Herkunft der Dokumente erschließt sich zum Teil aus den Stempeln in den gescannten Werken (vor allem stammen sie aus dem *Museum of Modern Art*), die Zusammenstellung erscheint willkürlich und wenig repräsentativ. Dennoch findet sich vergleichbares Material nicht systematisch auf deutsch(sprachig)en Seiten. Hier zeigt sich, dass die Furcht vor Copyright-Verletzung und den möglichen Folgen, die in Deutschland ein konsequentes Digitalisieren, Aufbereiten und Bereitstellen derartiger Quellen bisher verhindert, auch insofern problematisch ist, als nun derartige Sammlungen wie die der MHDL als die am einfachsten zugänglichen Quellenkonvolute zum deutschen Film gelten müssen. Die Nutzung in der Lehre könnte dazu beitragen, dass diese

5 mediahistoryproject.org (letzter Zugriff am 9.11.2019).

6 <http://lantern.mediahist.org/> (letzter Zugriff am 9.11.2019).

überaus nützliche, aber eben auch in der Auswahl idiosynkratische Sammlung als wichtige Grundlage für studentische Projekte und weitere Forschung erhalten muss. Zudem ist bedenkenswert, dass die Forschung zum deutschsprachigen Film zukünftig zunehmend von einer US-amerikanischen Plattform abhängig ist, wenn es nicht gelingt, die wichtigsten Quellen selbst konsequent online zugänglich zu machen und dauerhaft vorzuhalten.

Erfreulicher als hierzulande ist die Situation in der Schweiz, wo in einer Kooperation der Schweizerischen Nationalbibliothek und des Seminars für Filmwissenschaft der Universität Zürich eine Reihe wichtiger Filmzeitschriften digitalisiert wurde und online bereit steht.⁷ Diese fokussieren sich insbesondere auf die Phase von 1913 bis 1944 und sind über das Internet in Volltext durchsuchbar. Weiterführende Links verweisen auf andere Ressourcen, die online verfügbar sind. Hier zeigt sich beispielhaft, dass der Auftrag an die Wissenschaft nicht nur lautet, neue Forschung zu entwickeln und diesen Forschungsstand in der Lehre weiterzugeben, sondern sich auch um die Fortentwicklung der Infrastruktur zu kümmern. Am besten gelingen derartige Projekte in der Zusammenarbeit von großen Infrastruktureinrichtungen, die für die massenhafte und professionelle Digitalisierung wie auch für die Aufbereitung und Bereitstellung dieser Daten den nötigen Apparat aufweisen, und den betroffenen Forschenden, die das Wissen über die Bedarfe und Fallstricke besitzen. Im Fazit werden wir noch ausführlicher auf derartige Konstellationen eingehen.

Aspekte der Nutzung von Infrastrukturen

Die Nutzung von Ressourcen für die Filmwissenschaft wird bestimmt durch praktische, methodische, rechtliche, technische und infrastrukturelle Bedingungen. Es gibt bereits eine Reihe von Vorschlägen, um diese verschiedenen Bedingungen zu bündeln.⁸ Aus unserer Sicht stehen vor allem Fragen des Zugriffs, der Langfristigkeit, der Verwendbarkeit und der Validität der Daten im Vordergrund. Im Folgenden erläutern wir ausgewählte Aspekte der genannten Punkte.

7 <https://www.film.uzh.ch/de/bibliothek/zeitschriften/historical.html> (letzter Zugriff am 9.11.2019).

8 Zu den Mindestanforderungen, die DARIAH-DE an Forschungsdaten stellt (Validität, Reliabilität, Maschinenlesbarkeit, Referenzierbarkeit) sowie zu den FAIR Data Principles vgl. den Beitrag von Sarah-Mai Dang in diesem Heft.

Auffindbarkeit und Zugriff

Damit eine Ressource genutzt werden kann, muss sie zunächst einmal auffindbar sein. Allgemeine Suchmaschinen, allen voran *Google Search*, haben die gewünschten Informationen zwar aller Wahrscheinlichkeit nach im Index, erscheinen aber für gezielte Suchanfragen nach filmwissenschaftlichen Quellen, Datenbanken, Repositorien, Blogs oder ähnlichen Ressourcen wenig geeignet. Akademische Suchmaschinen wie *Google Scholar* oder *Microsoft Academic* sind für Publikationen im PDF-Format optimiert.⁹ *Google Dataset Search* verspricht zwar, auch Forschungsdaten zentral durchsuchbar zu machen, befindet sich aber noch im Beta-Stadium.¹⁰ Aufgenommen werden zudem nur Daten aus institutionell geführten Repositorien. Wie wir schon ausgeführt haben, trifft das auf die Vielzahl der interessanten Initiativen und Aktivitäten nicht zu.

In vielem erinnert die gegenwärtige Situation an die Anfänge der Suchmaschinen für das Internet. Vor dem Siegeszug der Google-Suche, die als erste Suchmaschine ein konsequent aufgeräumtes Interface mit nicht viel mehr als einem Briefschlitz zur Eingabe der Suchanfrage anbot, waren Suchmaschinen oft als redaktionell geführte Web-Portale angelegt mit der Tendenz, ihr Interface mit unübersichtlich vielen Links zu überfrachten. Auch heute müssen Forschende in der Regel lange Listen von Links konsultieren, wollen Sie sich einen Überblick über filmwissenschaftliche (oder andere) Forschungsressourcen verschaffen. Das von der Universitätsbibliothek Regensburg gepflegte *Datenbank Infosystem* (DBIS) stellt vermutlich die vollständigste und am aufwändigsten gepflegte Sammlung von Verweisen auf fachspezifische Ressourcen dar. Freilich ist auch das DBIS nicht vollständig, vor allem werden Websites von Forschungsprojekten nicht aufgeführt.¹¹ Zudem ist der Zuschnitt des Fachgebiets als Aggregat aus «Medien- und Kommunikationswissenschaften, Publizistik, Film- und Theaterwissenschaft» aus filmwissenschaftlicher Sicht nicht ideal gewählt. Schließlich sind viele Ressourcen, auf die DBIS verweist, lizenzpflichtig

9 Vgl. Google Scholar (<https://scholar.google.com>) sowie Microsoft Academic (<https://academic.microsoft.com>).

10 <https://toolbox.google.com/datasetsearch> (letzter Zugriff am 9.11.2019). Eine Suche nach dem Stichwort «film» ergibt aktuell weltweit 10 Treffer, darunter Zahlen zur «Plastics Packaging Film and Sheet Manufacturing»-Industrie.

11 So fehlt in DBIS etwa die Timeline of Historical Film Colors (filmcolors.org) von Barbara Flückiger.

und nicht allgemein frei zugänglich.¹² Für das Fach passgenauer sind privat betriebene Übersichten, die von Mitgliedern der Fachcommunity gepflegt werden. Aber auch hier dominieren Linklisten, deren Verfügbarkeit, Aktualität, Vollständigkeit und Granularität völlig vom privaten Engagement ihrer Autor*innen abhängen.¹³

Langfristigkeit

Diese Erfahrungen wird jede*r Forschende einmal gemacht haben: Der in einem Text zitierte Link führt nicht auf die versprochene Ressource, stattdessen gibt der Browser zurück: «404 Seite nicht gefunden». Oder unter einer Domain ist mittlerweile ein anderer Anbieter zu finden. Zwei Beispiele: Unter <http://www.medien-buehne-film.de/> ist nicht mehr die Virtuelle Fachbibliothek *medien buehne film* zu finden, sondern eine Website, die für Versicherungen wirbt. Unter <http://www.sfb626.de> (ehemals die Website des Sonderforschungsbereichs 626: «Ästhetische Erfahrung im Zeichen der Entgrenzung der Künste») bekommt man Tipps zum Börsenhandel.¹⁴ Und unter der Web-Adresse www.movii.de des von 2001 bis 2003 vom BMBF geförderten Projekts *movii – moving images & interfaces*, das «eine Reihe von netzbasierten Modulen [für] eine umfassende Propädeutik der Mediengestaltung» anbot,¹⁵ erreicht man aktuell eine obskure Website mit pornografischen Inhalten. Es gibt auch Fälle, in denen eine Ressource zwar noch existiert, aber nicht mehr korrekt angezeigt wird, weil die im Hintergrund arbeitende Datenbank veraltet ist oder weil der Browser das Plugin nicht mehr unterstützt.¹⁶ Oder die vor Jahren gekaufte CD-ROM, die wichtige Daten enthält, ist nicht mehr abspielbar.

12 Vgl. die Einträge unter http://dbis.ur.de/dbliste.php?bib_id=alle&lett=f&gebiete=53. Greift man von einem Universitätsrechner auf DBIS zu, bekommt man jeweils andere Angebote mit den jeweils am Standort erworbenen Lizenzen angezeigt.

13 Vgl. die Listen auf [filmwissenschaft.umsonst](https://filmwissenschaft.umsonst.wordpress.com/) (<https://filmwissenschaftumsonst.wordpress.com/>), oa books (<https://oabooks.de/linkliste/>), Film Studies for Free (<https://filmstudiesforfree.blogspot.com>) oder open access in media studies (<https://oamediaudies.com/discover/datasets/>).

14 Der Bestand der Fachbibliothek ist mittlerweile in adlr.link eingeflossen. Die Websites des SFB werden archiviert unter: <https://www.geschkult.fu-berlin.de/e/sfb626/index.html>.

15 Vgl. dazu die Beschreibung des Projekts auf www.uni-trier.de/index.php?id=16629 (letzter Zugriff am 7.2.2020).

16 Vgl. z. B. die Siegener Kino-Datenbanken unter <http://www.fk615.221b.de/siegen/start/show/index.php> (PHP 7 unterstützt MySQL nicht mehr) oder die Beiträge der Zeitschrift

Das alles zeigt: Digitalisierung ist keine einfache Epochenschwelle, über die man einmal in den digitalen Raum eintritt, um sich dann wie in einem glatten Raum ohne Hindernisse zu bewegen, sondern ein transformativer Prozess, der nie abgeschlossen ist. Domains müssen erhalten und gepflegt werden, Websites müssen auf neue Generationen von Browsern eingestellt werden, Systeme brauchen Updates, Daten müssen migriert und an neue Standards angeglichen werden. Um ein Angebot langfristig zu erhalten, müssen viele Aspekte berücksichtigt werden, mindestens jedoch diese:

Erstens: Angebote müssen stabil erreichbar sein. Für Publikationen hat sich hier das System der PID (Persistent Identifiers) bewährt, etwa in Form von *Digital Object Identifiers* (DOIs), *Uniform Resource Names* (URNs) oder durch ein Handle-System. Weniger bekannt ist, dass auch Domain-Namen, sprich: ULRs persistent gemacht werden können (PURLs).¹⁷

Zweitens: Angebote müssen erhalten bleiben. Die individuell erstellte Website ist von Drittmitteln oder Projektlaufzeiten abhängig, wenn nicht von privatem Engagement und (finanziellen wie zeitlichen) Ressourcen, die begrenzt sind. Hier gilt es, dauerhaftere Lösungen zu finden, etwa indem man als Betreiber/in einer Website oder Datenbank Kooperationen mit Institutionen wie Universitätsbibliotheken oder Repositorien eingeht.

Drittens: Um eine dauerhafte Funktionalität zu gewährleisten, ist es notwendig, Daten technisch zu migrieren, wovon nicht nur die Daten selbst, sondern ebenso die Trägermedien, Software, Hardware, Schnittstellen und weitere Anwendungen betroffen sein können. Dies überfordert oft Individuen, die Angebote bereitstellen, weil sie die infrastrukturellen Kapazitäten nicht zur Hand haben, um entsprechende Arbeiten zu verrichten.

Viertens: Daten müssen dauerhaft gepflegt werden. Keine Datenbank ist ohne Fehler. Korrekturen müssen möglich sein, ebenso wie Ergänzungen, Erweiterungen oder Umformatierungen auf geänderte oder neue Metadaten-Standards. Selbst mit Projektende «abgeschlossene» Datenbanken werden als FAIRe Daten prinzipiell erweiterbar und unabgeschlossen, insofern nach den Prinzipien des Linked Open Data alle Daten als Knoten in einer umfassenden Datenbank («Giant Global Graph») zu verstehen sind. Überhaupt geht Wissenschaft in ihrer modernen Form davon aus, dass Er-

Vectors – Journal of Culture and Technology in a Dynamic Vernacular unter <https://vectors.usc.edu/journal/index.php> (das Flash-Plugin wird von aktuellen Browsern nicht mehr unterstützt).

17 Siehe <http://purl.org/> (letzter Zugriff am 9.11.2019). Der Dienst wird seit 2016 vom Internet Archive angeboten.

kenntnisse grundsätzlich temporär, falsifizierbar, ergänzungsbedürftig und revidierbar sind, sodass es in der Logik der (wissenschaftlichen) Sache liegt, nicht für alle Zeiten festgeschrieben zu sein.

Qualität

Hat man eine Quelle filmwissenschaftlicher Daten gefunden, so stellt sich die Frage nach deren Güte oder Qualität: Wie belastbar (valide, reliabel, objektiv) sind sie in Bezug auf den jeweiligen Kontext ihrer Erhebung oder Fragestellung? Wie vollständig und umfassend wurden die Daten erhoben? Macht die datengebende Instanz die Methode und den Vorgang der Datengewinnung durch Dokumentation transparent und nachvollziehbar? Wie aktuell sind die Daten? Fehlen wichtige Angaben? Wenn ja, wird das offengelegt? Welche Typen von Metadaten – deskriptive, strukturelle, administrative (Zeng/Qin 2017, 18–22) – begleiten den Datensatz?

Vor allem in Hinblick auf die Dokumentation sowie die Auswahl und Anwendung geeigneter Metadaten-Schemata oder auch Normdaten sind Institutionen, die die entsprechende Expertise dafür aufweisen (etwa Bibliotheken), besser aufgestellt als die/der einzelne Wissenschaftler/in. Spezielle Herausforderungen stellen sich zudem, wenn Daten in einem offenen Verfahren durch Crowdsourcing erhoben werden. Durch die Verteilung der Datenerhebung an viele lassen sich Datenmengen zwar leicht skalieren. Die Dokumentation aber wird schwieriger, mit der Folge, dass sich die Qualität der Daten für Dritte nicht mehr ohne weiteres einschätzen lässt. Wikipedia, das vermutlich größte Crowdsourcing-Projekt im Bereich des Wissens, begegnet diesem Problem durch eine Flut von Richtlinien, Dokumentationen und automatisierten Prüfprozessen der in ihr enthaltenen Informationen. Das erhöht freilich die Zugangsschwelle zur ‚freien Enzyklopädie‘ nicht unerheblich.¹⁸

Wiederverwendbarkeit und Anschlussfähigkeit

Daten sind nicht gleich Daten. Der Datengeber – Repository, Archiv, Website – kann dieselben Daten zu unterschiedlichen Bedingungen bereitstellen und so deren Wiederverwendbarkeit und Anschlussfähigkeit regulieren (stärken oder einschränken). Besonders deutlich wird dies am

18 Vgl. die Einträge unter <https://de.wikipedia.org/wiki/Wikipedia:Richtlinien> (letzter Zugriff am 9.11.2019).

Beispiel des Urheberrechts oder der damit verwandten immateriellen Schutzrechte (etwa das Leistungsschutzrecht). Die Filmwissenschaft, die sich in der Regel mit urheberrechtlich geschütztem Material auseinandersetzt, steht hier bekanntlich vor besonderen Herausforderungen. Ein Gutachten zu urheberrechtlichen Aspekten der Verwendung von AV-Materialien in Forschung und Lehre aus dem Jahr 2015, das im Auftrag der Gesellschaft für Medienwissenschaft und des Verbands der Historiker und Historikerinnen Deutschlands von iRights angefertigt wurde, bezeichnet die Sachlage als «juristisches Minenfeld» (Klimpel/König 2015, 5). Da das Urheberrecht als Persönlichkeitsrecht unveräußerlich ist, also automatisch wirksam wird, greift es in jedem Fall. Es reicht also nicht, wenn die/der Urheber/in ein Werk (Texte, Bilder, Forschungsdaten usw.) online stellt, damit Dritte es verwenden dürfen. Fehlt auf einer Website eine genaue Angabe zur Lizenzierung der Daten (etwa eine CC-Lizenz), muss man annehmen, dass das Werk nur mit expliziter Einwilligung der Urheber verwendet werden darf.

Ein anderes Hindernis ist die Praxis mancher Archive, eigentlich gemeinfreie Werke durch Gebühren und Nutzungsvereinbarungen wieder einzusperren (Klimpel/Rack 2017). So kann es passieren, dass Abbildungen nicht in Publikationen verwendet werden, obwohl die Schutzdauer des Urheberrechts (in Deutschland: 70 Jahre nach Tod der/des Urheber/in) für dieses Werk längst abgelaufen ist. Aus Angst vor Rechtsstreitigkeiten wird mitunter von Autor/innen sogar das Zitatrecht nicht in Anspruch genommen. Gerade die Unübersichtlichkeit und Komplexität der rechtlichen Rahmenbedingungen führen zum Teil dazu, dass selbst Nutzungen, die unstrittig sind, aus Furcht vor negativen Folgen unterlassen werden.

Ein weiterer Faktor, der die Anschlussfähigkeit und Wiederverwendbarkeit von Forschungsdaten bestimmt, besteht in Vokabular und Standardisierung der Daten. Ein Datensatz, in dem ein und dasselbe Datum mal als «9. Nov 1989», mal als «1989-11-09» notiert ist, lässt sich nur schwer automatisiert auswerten. Wird als Regisseur lediglich «Frank Müller» angegeben, ist nicht klar, wer gemeint ist: Der Regisseur von HERR SCHULTZ STICHT IN SEE (D 2012) oder von 24H BERLIN – EIN TAG IM LEBEN (D 2009)? Wenn als Filmtitel nur «Psycho» angegeben ist, ist das Original von Alfred Hitchcock oder das Remake von Gus Van Sant gemeint? Zur Disambiguierung von Personennamen und Film- oder anderen Werktiteln bieten sich Identifikatoren an, wie sie beispielsweise von *filmportal.de*, der *Gemeinsamen Normdatei* (GND) der Deutschen Nationalbibliothek oder Wikidata zur Verfügung gestellt werden. Gerade bei der Nutzung und Ent-

wicklung von Normdaten steht die film- und medienwissenschaftliche Forschung erst ganz am Anfang. Wichtig ist, dafür einen offenen und kollaborativen Prozess in Gang zu setzen, der die Fachcommunity einbindet und reflexiv die Probleme und Schwierigkeiten nicht unter den Teppich kehrt, sondern sichtbar macht. Die Gestaltung derartiger Vokabulare, Standards oder Ontologien ist ein komplexes Thema, dem in der Kürze dieses Beitrags nicht ausreichend Rechnung getragen werden kann.

Wiederverwendbarkeit heißt schließlich auch, dass die Daten in Formaten vorliegen, die plattformunabhängig sind, nicht-proprietär und nicht exklusiv an bestimmte Betriebssysteme oder Anwendungen gebunden. Auch erhöhen öffentliche, gut dokumentierte und standardisierte Schnittstellen die Chance, dass Daten von Dritten wiederverwendet werden können. Für die automatisierte Auswertung sollten Daten zudem (auch) in maschinenlesbarem Format vorliegen. Eine Bibliografie lässt sich als Word- oder PDF-Datei von Menschen zwar durchlesen – als BibTex- oder RIS-Datei ist es aber möglich, sie problemlos in ein Literaturverwaltungsprogramm zu importieren und automatisiert auszuwerten.

Fazit

Wenn sich also, wie wir glauben, einige entscheidende Aspekte im Hinblick auf Zugang zu und Umgang mit Quellen und Materialien in der vernetzt-digitalen Welt verändert haben, so darf die Wissenschaft dabei keine passive Zuschauerin bleiben. Wir sind der Überzeugung, dass die forschungsleitenden Infrastrukturen von der Wissenschaft (mit) entworfen und betrieben werden müssen, weil die Prozesse selbst kaum noch von der Forschung zu trennen sind. Tatsächlich sind Trends wie *Data Papers* oder *Open Science* Ausdruck einer ähnlichen Grundhaltung, Wissenschaft als Prozess zu verstehen – und nicht lediglich als Ergebnis von Prozessen, die am Ende nicht mehr transparent nachvollziehbar sind. Die aktive Gestaltung der Infrastruktur durch die Forschung selbst, in enger Zusammenarbeit mit den klassischen infrastrukturellen Einrichtungen (Bibliotheken, Archive, Rechenzentren usw.), kann eine Ausgangsbasis schaffen, auf der neue digitale Methoden und Werkzeuge aufsetzen, die große Datenmengen voraussetzen. Gerade im Bereich der kulturellen Überlieferung und der Geisteswissenschaft haben wir es ja mit Daten zu tun, die oftmals umstritten, ambivalent oder nicht genau bestimmbar sind, sodass die Schaffung von Datenpools mit akzeptabler Qualität ein vordringliches Desiderat ist.

Erst dann lässt sich überhaupt erkunden, ob und welche digitalen Methoden für die Filmwissenschaft produktive Ergebnisse zu versprechen vermögen. Da derzeit zudem mit der *Nationalen Forschungsdateninfrastruktur*, mit zahlreichen Ausschreibungen zu Digitalisierung und Erschließung und mit weiteren Optionen, die bisher in der Filmforschung noch nicht erprobt wurden (z. B. Akademievorhaben), viele Möglichkeiten bestehen, diese aufwändigen Arbeiten projektförmig anzugehen, scheint der Zeitpunkt für eine Neugestaltung der Infrastruktur des Faches sehr günstig. Entscheidend wird dabei sein, dass die einzelnen Projekte, die häufig zunächst unverbunden und solitär gestartet wurden, von vornherein mit Blick auf offene Schnittstellen und einen produktiven Datenaustausch angelegt sind. Ansonsten würden wir nur mit enormem Aufwand weitere Datencontainer schaffen, die nach außen isoliert wären, statt sich sinnvoll im Hinblick auf Austausch und Wiederverwendung zu öffnen.

Literatur

- Hagener, Malte / Zimmermann, Yvonne (2019) Viking Eggeling and the European Avant-garde Cinema. In: *A Cultural History of the Avantgarde in the Nordic Countries. 1925–1950*. Hg. v. Benedikt Hjarson, Andrea Kollnitz, Per Stounbjerg & Tania Ørum. Leiden / Boston: Brill / Rodopi, S. 82–101.
- Klimpel, Paul / König, Eva-Marie (2015) *Urheberrechtliche Aspekte beim Umgang mit audiovisuellen Materialien in Forschung und Lehre*. Rechtsgutachten. irights.info [https://irights.info/2015/10/16/gutachten-zu-audiovisuellen-medien-in-forschung-und-lehre-veroeffentlicht/26258 (letzter Zugriff am 18.02.2020)].
- / Reck, Fabian (2017) Archivgebühren: Was gemeinfrei ist, muss gemeinfrei bleiben. In: irights.info [https://irights.info/artikel/archivgebuehren-was-gemeinfrei-ist-muss-gemeinfrei-bleiben/28815 (letzter Zugriff am 18.02.2020)].
- Marvin, Carolyn (1990) *When Old Technologies Were New. Thinking about Electric Communication in the Late Nineteenth Century*. Oxford: Oxford University Press.
- Verhoeven, Deb (2012) New Cinema History and the Computational Turn. In: *WCCA 2012. Beyond Art, Beyond Humanities, Beyond Technology: A New Creativity. World Congress of Communication and the Arts Confer-*

ence Proceedings. Guimarães, Portugal: COPEC – Science and Education Research Council 2012 [<http://dro.deakin.edu.au/eserv/DU:30044939/verhoeven-newcinema-2012.pdf> (letzter Zugriff am 18.02.2020)].

Zeng, Marcia Lei / Qin, Jian (2016) *Metadata*. 2. Auflage. London: Facet Publishing.